2022

Long-Term Stability of Baseline Impact Scores in High School Football Players across Age Groups

Christopher Thomas Burley

LONG-TERM STABILITY OF BASELINE IMPACT SCORES IN HIGH SCHOOL FOOTBALL PLAYERS ACROSS AGE GROUPS

by

Chris Burley

A Dissertation Presented to the College of Psychology
of Nova Southeastern University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

NOVA SOUTHEASTERN UNIVERSITY

2022

DISSERTATION APPROVAL SHEET

This Dissertation was submitted by Christopher Burley under the direction of the Chairperson of the Dissertation committed listed below. It was submitted to the School of Psychology and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Clinical Psychology at Nova Southeastern University.

Approved:

8/10/2022
Date of Defense

Charles Golden, Ph.D., Chairperson

Lisa Lashley
Lisa Lashley, Psy.D.

Ryan Black, Ph.D.

Robert Seifer
Robert Seifer, Ph.D.

8/15/2022
Date of Final Approval

Charles Golden, Ph.D., Chairperson

Statement of Original Work

I declare the following:

I have read the Code of Student Conduct and Academic Responsibility as described in the *Student Handbook* of Nova Southeastern University. This dissertation represents my original work, except where I have acknowledged the ideas, words, or material of other authors.

Where another author's ideas have been presented in this dissertation, I have acknowledged the author's ideas by citing them in the required style.

Where another author's words have been presented in this dissertation, I have acknowledged the author's words by using appropriate quotation devices and citations in the required style.

I have obtained permission from the author or publisher—in accordance with the required guidelines—to include any copyrighted material (e.g., tables, figures, survey instruments, large portions of text) in this dissertation manuscript.

Chris Burley
Name

June 20th, 2022
Date

Table of Contents

List of Tables

LONG-TERM STABILITY OF BASELINE IMPACT SCORES IN HIGH SCHOOL FOOTBALL PLAYERS ACROSS AGE GROUPS

by

Chris Burley

Nova Southeastern University

ABSTRACT

In recent decades, the increased awareness of the prevalence of concussions has resulted in significant advancements in concussion assessment and treatment, including annual or bi-annual pre-season computerized baseline assessments. However, limited research has focused on the differences in stability of baseline neurocognitive performance among different age groups in adolescence. The main purpose of the present study was to explore the long-term stability of ImPACT baseline assessments across one-year intervals in multiple age groups within a sample of high school football players.

Subjects who completed two baseline assessments between the ages of 14 and 17 who had completed two baselines one year apart were selected from a de-identified archival database. Subjects were separated into three groups based on age at first baseline (i.e., 14, 15, and 16).

Results indicate differences in baseline performance stability between age groups, with younger athletes demonstrating improvements in performance consistent with ongoing cognitive development and older athletes demonstrating more stability in performance. Importantly, the Reaction Time composite score remained stable among all groups. Results also provide evidence for wide ranging Reliable Change Indices across all age groups, which can make it difficult to interpret meaningful change after injury for those whose baseline was completed one year ago.

These findings indicate younger athletes (i.e., 14- and 15-year-olds) should ideally complete baselines prior to each athletic season (e.g., fall, winter, spring), whereas older athletes (e.g., 16-year-olds) should complete baselines annually. The use of more frequent baselines aims to increase clinicians' ability to detect clinically meaningful change following concussion, which in turn would lead to increased accuracy in return-to-play decision making. However, it is also important to recognize potential pitfalls of increased baseline testing (e.g., practice effects, lack of available resources), which are discussed. When recent baselines are not available, these results indicate the Reaction Time composite may serve as an important indicator of departure from baseline given its demonstrated stability across groups. In addition to clinical implications, the present results are consistent with Luria's theory of cognitive development which indicates some cognitive skills continue to develop throughout adolescence along with corresponding cortical areas. Lastly, these results provide some information regarding the cognitive development of high school football players within the context of repeated exposure to sub-concussive impacts, though such conclusions are limited be the absence of a control group.

Keywords: concussion, baseline assessment, football, cognitive development

CHAPTER I: Statement of the Problem

The controversy surrounding concussions in high school athletics, specifically football, has garnered significant media and medical coverage in recent decades. As such, clinicians responsible for treating these athletes have developed an increased interest in the best ways to approach evaluation and treatment of these athletes. Thus far, concussion protocols, increased education and awareness, newly implemented game rules, and better equipment have changed the landscape of the game. However, the increased risk of concussion remains for these athletes.

The negative impacts of repeated head trauma, especially in high school football players, has elicited major concerns from parents, players, coaches, and politicians alike. Specific sequelae include extended recovery periods after concussion, as well as more serious results such as SIS and possibly CTE. High school football players are of particular research interest due to the continuous development of the brain throughout high school. Such development presents a possible risk factor for more serious deficits from repeated concussive or sub-concussive impacts. Current long-term symptoms of repeated head trauma can include memory loss or dysfunction, executive functioning deficits, and attentional and concentration difficulties. Additionally, significant emotional and behavioral dysfunction have been firmly stated in the literature. Reported symptoms of emotional and behavioral dysfunction include anxiety, depression, frustration, impulsivity, and self-harm behaviors (Giza & Hovda, 2014). Neuropsychologists are now more often being utilized in the diagnosis and treatment of concussions, with increased inclusion in interdisciplinary teams assembled to provide optimal treatment for patients (Zillmer, 2003). However, research regarding the long-term test-retest reliability of computerized neuropsychological tests frequently used in the diagnosis and management of concussion is currently underdeveloped.

Clinicians are encouraged to use results of this study to make educated and accurate impressions regarding RTP decision-making using baseline comparisons. Additionally, clinicians should use results of this study to more accurately track typical development of high school football players in certain age groups, further increasing the accuracy of concussion classification and treatment progress.

CHAPTER II: Review of the Literature

Traumatic Brain Injury (TBI)

TBI is rated based on the severity of the injury and the sequelae that follow as a result, ranging from mild to severe. The Centers for Disease Control and Prevention (CDC) has defined traumatic brain injury as "craniocerebral trauma, arising from blunt or penetrating trauma or from acceleration-deceleration forces" that is associated with a number of subsequent negative symptoms including loss of consciousness, skull fracture, intracranial lesions, bleeds, neuropsychological deficits, and even death (CDC, Thurman, Sniezek, Johnson, et al., 1995). More recently, TBI has been broadly defined as "an alteration in brain function, or other evidence of brain pathology, caused by an external force" by the Demographics and Clinical Assessment Working Group of the International and Interagency Initiative toward Common Data Elements for Research on Traumatic Brain Injury and Psychological Health (Menon, Schwab, Wright & Maas, 2010). Epidemiological research indicates that more than 200,000 individuals are hospitalized for nonfatal TBI each year in the United States (Corrigan, Selassie & Orman, 2010). In addition to the large number of individuals hospitalized for TBI each year, millions of individuals live with long-term consequences of TBI, indicating the significant healthcare and societal cost associated with these injuries.

Sequelae of TBI differ vastly depending on the severity of the injury. TBI ratings are determined by popular screening measures including the Glasgow Coma Scale (GCS) and the Abbreviated Injury Scale (AIS), both of which are administered by trained clinicians or other trained individuals (Corrigan, Selassie & Orman, 2010). Outcomes on these measures differentiate TBIs into ranges of severity including mild, moderate, and severe. Mild traumatic brain injury and concussion are often used interchangeably within the scientific community.

Concussion represents the beginning of the TBI severity spectrum, but it accounts for between 80 and 90 percent of all TBIs (Saatman et al., 2008). Epidemiologists have estimated that in the United States alone there are 3.8 million individuals who require concussion care each year, in comparison to the estimated 200,000 moderate-to-severe TBIs that occur annually (Langlois, Rutland-Brown, & Wald, 2006; Corrigan, Selassie & Orman, 2010). As a result, the CDC has estimated that the annual economic impact of concussions is as high as $40.6 billion annually (Miller, DePadilla, & Xu, 2021). These numbers serve to indicate the significant need for appropriate and effective treatment of concussion.

Sport-Related Concussion

Concussion is considered a mild traumatic brain injury that results in any type of short-term neurological or neurocognitive dysfunction (Giza & Hovda, 2001). Kontos and Collins (2018) define concussion as "a complex pathophysiological process affecting the brain, induced by biomechanical forces, which may involve different symptoms, impairment, clinical profiles or subtypes, and recovery trajectories that are influenced by a variety of risk factors." Common signs of concussion, which are observable indicators of concussion, include loss of consciousness, disorientation, confusion, vomiting, imbalance, dizziness, nausea, and vision problems (Kontos & Collins, 2018). Common symptoms of concussion, which are self-reported interpretations of feelings, include headache, difficulty concentrating, fatigue, drowsiness, dizziness, feeling "foggy," feeling slowed down, sensitivity to noise or light, difficulty remembering, nausea, irritability, nervousness, and sleep disruption (Kontos & Collins, 2018).

These signs and symptoms are attributed to the unique pathophysiological process that occurs immediately following concussion. The pathophysiology of concussion represents an important distinction between concussion and moderate and severe TBIs. Whereas structural damage is ubiquitous in moderate and severe TBI, structural damage is seldom observed in concussion. Unlike moderate and severe TBI, neurons do not typically die in concussion. Rather, axons are stretched out, causing a complex metabolic event known as the neurometabolic cascade (Giza & Hovda, 2001).

A healthy brain functions in an efficient manner and does not require much energy to complete tasks. However, after a concussion the brain becomes inefficient at the cellular level. This inefficient cellular function is a result of the neurometabolic cascade that occurs immediately following concussion. The neurometabolic cascade is comprised of two major events: 1) the dysregulated efflux of Potassium (K^+) ions outside of the cell and 2) the dysregulated influx of Calcium (Ca^{2+}) ions into the cell. This abrupt change in polarization of the neuron causes Sodium-Potassium (Na^+-K^+) pumps to activate a large increase in glucose metabolism in an effort to restore balance in the cell. However, the over-activation of the Na^+-K^+ pumps depletes energy stores and results in an energy crisis within the brain. This energy crisis is most prominent within the first few hours following concussion but can last for days to weeks and is the main source of neuropsychological deficits and symptoms following concussion (Kontos & Collins, 2018).

Typical recovery periods for concussion range from one to three weeks depending on age, among other factors. Henry, Elbin, Collins, Marchetti, & Kontos (2016) reported that recovery from concussive symptoms typically ranges from 3-4 weeks in adolescents. In another study, McCrea et al. (2009) concluded that fewer than 3% of subjects evidenced neuropsychological deficits or reported concussive symptoms at one-month post-injury. In order to determine whether or not an athlete has fully recovered from a concussion, neuropsychologists employ standard protocols described in the next section.

Assessment of Concussion

Neuropsychological assessment has become a major component in the current approach to effective assessment and treatment of SRC. Such assessments are incredibly valuable to multidisciplinary teams focused on treating concussion, especially during the return-to-play (RTP) decision making process (Covassin et al., 2009). The use of neuropsychological assessments in the RTP decision making process can sometimes lead to important ethical considerations for clinicians. This responsibility of neuropsychological assessment and clearance for athletes presents a challenge for neuropsychologists, but also presents opportunities for neuropsychologists to demonstrate their value in an interdisciplinary setting focused on treating brain injuries.

Currently, standard practice among neuropsychologists is to compare performance on a neuropsychological battery after a concussion to the individual's performance on the same neuropsychological battery at a time when the individual was not concussed. The result of the initial, non-concussed battery is known as the individual's "baseline" performance. After a concussion, athletes will complete this battery at multiple timepoints in order to gauge their recovery, as measured by their return to baseline scores. This practice of measuring a return to baseline after concussion, known as serial assessment, is regarded by many as the best practice in the evaluation of recovery in concussion management (Covassin et al., 2009).

The American Medical Society for Sports Medicine maintains that the use of standardized assessments gives concussion management a necessary structure. These assessments not only help gauge neuropsychological return to baseline, but they also help to track the myriad of symptoms that may be present in concussed individuals. Despite the theoretical utility of baseline assessments, they are only as useful as the athlete's effort during baseline assessment. When using data that is not representative of true baseline abilities, changes detected following concussion, or lack thereof, are unreliable and cannot be used to accurately assess differences in functioning after injury (Lichtenstein, Moser & Schatz, 2014). There are numerous explanations for inaccurate baselines, the most common being sandbagging, or intentionally performing below ability level. Other common reasons for inaccurate baselines include the poor test environment, poor test approach, poor sleep the night before, misunderstanding of test instructions, or strenuous physical exertion prior to testing (Iverson & Schatz, 2015). In cases for which baselines are not accurate representations of baseline cognitive ability, the sensitivity of the assessment is significantly decreased, thus limiting the effectiveness of serial assessment following concussion.

Despite the difficulties associated with baseline testing, repeated baseline assessments are necessary due to the continuous neurodevelopment of high school athletes. After understanding the difficulties associated with baseline assessment, it is next important to understand the assessments and protocols used to diagnose and monitor sport-related concussion.

ImPACT

ImPACT is a commonly used computerized neuropsychological assessment in the evaluation of SRC. ImPACT is a well-validated measure with a wealth of psychometric research to support its validity and reliability when used appropriately (Allen & Gfeller, 2011; Maerlender et al., 2010). Through the use of alternate forms, ImPACT is able to minimize practice effects, which is especially important in the serial assessment method used to evaluate recovery from SRC. Additionally, ImPACT is used with athletes at all developmental levels, ranging from youth athletes to professionals. Despite the difficulties posed by serial assessment of individuals across a wide variety of developmental levels, ImPACT is able to maintain diagnostic utility in

SRC. Research has found that ImPACT has a sensitivity of 91.4% and a specificity of 69.1%, providing evidence for the value data acquired through the use of ImPACT (Schatz & Sandel, 2013).

The most essential aspect of defining meaningful neuropsychological changes in serial assessment such as ImPACT is determining the reliability of the instrument. Establishing reliability allows clinicians to accurately attribute observed changes in scores to effects of concussion rather than other extraneous variables. Among neuropsychological tests, time between administrations is the greatest threat to test-retest reliability. However, the majority of ImPACT research has demonstrated stable test-retest reliability across a variety of test-retest intervals including seven days (Iverson, Lovell & Collins, 2003), one month (Schatz & Ferris, 2013), one year (Bruce, Echemendia, Meeuwisse, Comper & Sisco, 2014), and 2 years (Schatz, 2010). Despite the strong research literature to support ImPACT's reliability, other studies have found contradictory results indicating lower estimates of test-retest reliability for ImPACT's composite scores as well as its symptom scale (Broglio, Ferrara, Macciocchi, Baumgartner & Elliot, 2007; Cole et al., 2013; Resch et al., 2013).

An abundance of this contradictory ImPACT research has been completed within the high school population, suggesting that test-retest reliability may not be as strong as in older populations. One particular study, conducted by Elbin et al. in 2011, found the following intraclass correlation coefficients (ICCs) for a one-year test-retest interval: Verbal Memory (0.62), Visual Memory (0.70), Visual Motor Speed (0.85), Reaction Time (0.76), and Symptom Scale (0.57). Slick criteria (2006) indicate the following classifications for reliability values: ≥0.90 = very high, 0.80-0.89 = high, 0.70-0.79 = adequate, 0.60-0.69 = marginal, and <0.60 = low. Using the Slick criteria, results from the 2011 Elbin et al. study indicate questionable reliability in some ImPACT domains. Past research has also criticized the use of correlations and recommended the use of reliable change indices (RCIs; Iverson, Lovell & Collins, 2003) and regression based-measures (RBM; Barr, 2002). Researchers in the 2011 Elbin et al. study also calculated test-retest reliability through RBM and found that 91% of baselines fell within an 80% confidence interval (CI). Moreover, researchers found that only 1.0% of Visual Memory scores and only 0.2% of Reaction Time scores were outside a 95% CI. In a study examining one-, two-, and three-year test-retest reliability, investigators found ICCs ranging from .36 to .90, and found little change between intervals (Brett, Smyk, Solomon, Baughman, & Schatz, 2016). That same study also found that when using RCIs and RBM, very few (0-6%) cases fell outside of a 95% confidence interval. Other studies have also found reasonable reliability using RCIs, with a limited number of cases scoring outside of 80% and 95% CIs (Iverson, 2001; Jacobson & Truax, 1991). Despite the promising results of these studies, other researchers have produced contradictory findings. A study conducted by Tsushima et al. in 2016 examined the reliability of ImPACT over a two-year test-retest interval in a sample of high school athletes. Tsushima et al.'s study yielded the following ICCs: Verbal Memory (0.21), Visual Memory (0.49), Visual Motor Speed (0.72), Reaction Time (0.46), and Symptom Scale (0.40). These values are much lower than the values produced by the Elbin et al. study, suggesting poor reliability across the majority of ImPACT domains.

The development of psychometrically sound computerized neuropsychological tests is a significant issue in the evaluation of SRC. Specifically, questionable test-retest reliability has led researchers to caution clinicians regarding the use of computerized neuropsychological data as a comparative criterion in RTP decision-making (Alsalaheen et al., 2015). Researchers in the field have recommended multiple solutions for this issue, including the use of normative comparisons rather than individual baselines (Echemendia et al., 2012). However, others have argued that the use of normative comparisons may inaccurately classify concussed individuals who are outside of the average range (Schatz & Robertshaw, 2014). Others have suggested the use of multivariate base rates to identify concussion, proposing the criterion of two or more composite scores in the impaired range (Iverson & Schatz, 2015).

Due to the conflicting results present in the current literature, further investigation into the test-retest reliability of ImPACT is necessary, specifically in high school athletes. However, collision sport athletes, such as high school football players, present a unique population within high school athletes. Recent research has suggested that sub-concussive blows to the head may accumulate to create a cumulative deficit in football players. This proposed mechanism of dysfunction is especially important to consider in high school football players, as they are at a critical point in neurodevelopment. As such, neurodevelopment and sport-related concussed are discussed next. In order to better understand the biological and risk factors that may confound the utility of baseline to post-injury comparisons in high school football players, adolescent neurodevelopment is discussed below.

SRC and Adolescent Neurodevelopment

High school athletes are often at a major turning point in neurodevelopment, with the rise of new hormones throughout puberty and the rapid development of the brain throughout adolescence. According to the Luria's theory of brain organization and function, the sequence of development of cognitive function is dependent upon the physical and functional changes that occur in the natural maturation of different cortical areas (Obrzut & Hynd, 1986). The theory maintains that the development throughout each stage is consistent with qualitative organizational changes in adaptive and intellectual skills. Researchers have suggested that primary, secondary, and tertiary regions become functional by 12-years-old, thus largely solidifying general intellectual ability (Luria, 1980). However, cognitive development continues throughout adolescence and into adulthood.

A major area of neurodevelopment in high school athletes is the growth of the executive function system. This growth is driven by proliferations of neurons in the prefrontal cortex and limbic system during adolescence, and these new neural connections increase the ability of the individual to utilize higher-order functions such as abstract reasoning and planning behaviors. This higher-order control system, known as executive functioning, regulates the lower-order brain functions such as short-term memory, sensory perceptions, language, and motor skills. Most importantly, with the development of the prefrontal cortex, individuals learn to organize future/goal-oriented thinking and behaviors (Pokhrel et al., 2013). Differences in development of these skills is common in adolescents as individuals mature at varying rates (Spear, 2000). Because of the crucial neuropsychological development occurring throughout adolescence and into adulthood, research regarding concussion in these athletes is essential.

Research has shown that this period of development leaves high-school athletes at a higher risk for concussion, as well as prolonged recovery periods when compared to college athletes (Gessel, Fields, Collins, Dick, & Cmstock, 2007; Field, Collins, Lovell, & Maroon, 2003). There have been several suggested theories to explain increased vulnerability of younger brains to concussion. Those theories include less extensive myelination between neurons, a larger head-to-body size ratio, and thinner cranial bones, all of which result in less protection for the nervous system. Additionally, differences in levels of experience and fitness, as well as differences in equipment have been suggested as explanations for increased vulnerability in adolescents (Theye & Mueller, 2004). However, despite the underlying cause of increased vulnerability, neuropsychologists must be cognizant of increased vulnerability, prolonged recovery periods, and continuous neurodevelopment when working with high school athletes. This continuous development plays a crucial role in the recovery after concussion as well as the appropriate protocols for baseline assessment. Specifically, the unique neurodevelopment occurring in high school athletes indicates the need for research regarding changes that may occur across various age groups. After understanding the role of development in neuropsychological dysfunction and reported symptoms following concussion, it is also important to understand the risk factors associated with repeated brain injury in contact sports such as football.

Risk Factors Associated with Repeated Sub-concussive Injury

As previously mentioned, the signs and symptoms resulting from concussion are typically resolved within four weeks in high school athletes. RTP protocols have been implemented to minimize the risk of prolonged neuropsychological and somatic deficits resulting from SRC. However, impacts with sub-concussive force may occur during participation in sport that do not produce overt neurological symptoms. These “sub-concussive” hits are subtle brain injuries which are associated with tenuous neuropsychological deficits (Gysland et al., 2012). The majority of sub-concussive signs and symptoms are often momentary and go undetected following the impact. Such impacts are commonly referred to as “getting your bell rung.” Because the symptoms are short-lived, players, coaches, parents, and medical staff often overlook mild sport-related brain injury, as the sub-concussive impacts do not produce immediate or sustained signs of concussion. Of all sports, football is the likely to produce the highest volume of such sub-concussive hits. Due to the high volume of sub-concussive hits and the low likelihood that these hits are identified, high school football players may put themselves at greater risk for long-term neuropsychological dysfunction with repeated exposure to undiagnosed concussion and sub-concussive hits (McCrory et al., 2017).

The immediate effects of undiagnosed concussion or sub-concussive impacts are observed through extended recovery periods as well as more serious injury including Second Impact Syndrome (SIS). As described previously, the injured brain is most vulnerable to subsequent injuries during the energy crisis immediately following mild brain injury, especially in adolescent populations (Giza & Hovda, 2014). Research suggests that subsequent injury in this vulnerable period can have significant negative effects on an individual's recovery from concussion (Elbin et al., 2016). Concussion recovery is often likened to recovery from a sprained ankle; just as running on a sprained ankle too soon after injury may adversely impact recovery, so does "running" on a concussed brain (Kontos & Collins, 2018). Typical recovery periods for concussion range from one to three weeks depending on age, among other factors. Henry, Elbin, Collins, Marchetti, & Kontos (2016) reported that recovery from concussive symptoms typically ranges from 3-4 weeks in adolescents. In another study, McCrea et al. (2009) concluded that fewer than 3% of subjects evidenced neuropsychological deficits or reported concussive symptoms at one-month post-injury. However, both animal studies involving mice and rats (Giza, Griesbach, & Hovda, 2005; Griesbach, Hovda, Molteni, Wu, & Gomez- Pinilla, 2004) and research involving humans (Brown et al., 2014; Majerske et al., 2008) have shown that taxing the brain during the vulnerable period immediately following a concussion can lead to extended recovery periods. In extreme circumstances, subsequent injury during the vulnerable period has led to SIS, which is a rapid cerebral edema triggered by a secondary impact during the vulnerable period. SIS is a severe injury resulting in death or severe irreversible neurological impairment. Fortunately, SIS appears to be uncommon, with only 17 confirmed cases (McLendon et al., 2016).

In addition to extended recovery periods, research has suggested an association between concussion, sub-concussive impacts, and long-term neuropsychological dysfunction in athletes participating in high impact sports such as boxing and football. The disorder, Chronic traumatic encephalopathy (CTE), is a neurological condition characterized by confusion, slowing of speech, tremors, Parkinsonian symptoms, and overall mental deterioration (Saffary, Chin, & Cantu, 2012). A growing body of research exists in the area of CTE and its prevalence, etiology, and core symptoms, but there is still much research to be done prior to reaching conclusions regarding these areas. CTE has previously been referred to as "dementia pugilistica" when researched in boxers. Dementia pugilistica was thought to be a neurological disorder specific to boxers who had experienced repeated brain injury. Recently becoming a hot topic in sports, CTE has been confirmed not only former boxers but also retired football players from the National Football League (NFL) as well as professional soccer and rugby players (Omalu et al., 2005), and individuals with no history of brain injury or participation in contact sports (Iverson et al, 2019). Although CTE neuropathology has been discovered in individuals with no history of participation in contact sports, it is believed to be the result of repeated concussive or sub-concussive injuries over time. Given the believed association between repeated brain injury and CTE, it is important to also consider the possible effects of sub-concussive injury on the test-retest reliability of ImPACT scores.

Purpose

The purpose of the current study was to explore the long-term stability of ImPACT baseline assessments across one-year intervals in multiple age groups within a high school football player sample. In regard to psychometric research completed with ImPACT, test-retest reliability has been controversial. Whereas some research has shown adequate stability across intervals up to two years, other studies suggest that ImPACT has highly questionable stability across one-year intervals, specifically in high school athletes. As such, further research should provide clarification of test-retest reliability.

Currently, existing studies on psychometric properties of ImPACT focus on samples with a range of ages (Bruce, Echemendia, Meeuwisse, Comper & Sisco, 2014; Allen & Gfeller, 2011; Maerlender et al., 2010). Although this information is helpful for general interpretation of such ImPACT and the reliable changes to be expected across test-retest intervals, it does not provide age-based RCIs. In order to more accurately assess reliable change in high school athletes, age-based reliable change should be determined due to the development occurring in high school athletes.

Furthermore, research regarding long-term outcomes of repeated brain trauma from football have been well-established (Giza & Hovda, 2001; Omalu et al., 2005; McKee et al., 2009), but no research has considered the possible effects of repeated brain trauma in the reliability of ImPACT for high school athletes. High school football players sustaining concussion are currently being evaluated for RTP using the previously established general RCIs (Iverson, Lovell, & Collins, 2003). Although the overall RCIs of ImPACT have been established, there has been no formal analysis regarding the differences in RCIs for various groups. The present study provides the beginning groundwork regarding these gaps in psychometric information for this commonly utilized neuropsychological test in this specific population. Results of this study provide clinicians and researchers with valuable information for accurate assessment in the RTP decision-making process within this particular population.

Hypotheses

Hypothesis One. It was hypothesized that there would be significant differences between baseline one and baseline two ImPACT composite scores and the Symptom Scale total scores for each of the three specified age groups (i.e. 14-, 15-, and 16-years-old at initial baseline, respectively).

Justification One. Currently, research suggests controversial test-retest agreement of ImPACT in high school athletes in general. Previous studies have demonstrated significant differences between baselines (Elbin et al., 2011; Brett et al., 2016) whereas others have demonstrated negligible differences (Tsushima, 2015; Barr, 2002; Iverson, Lovell & Collins, 2003). However, no research has examined changes within age groups, suggesting a gap in the literature regarding differences at specific timepoints in development. In order to verify the existence of differences between baseline ImPACT scores in high school football players in specific age groups, each age group was examined independently.

Hypothesis Two. It was hypothesized that there would be significant differences in difference scores between age groups.

Justification Two. High school athletes are often at a major turning point in neurodevelopment, with the rise of new hormones throughout puberty and the rapid development of the brain throughout adolescence. According to the Luria's theory of brain organization and function, the sequence of development of cognitive function is dependent upon the physical and functional changes that occur in the natural maturation of different cortical areas (Obrzut & Hynd, 1986). The theory maintains that the development throughout each stage is consistent with qualitative organizational changes in adaptive and intellectual skills. Although largely established by 12-years-old, cognitive development continues throughout adolescence and into adulthood. This continuous development suggests that test-retest reliability of high school athletes is confounded by the significant neuropsychological development that occurs over the course of a year in adolescents. As such, high school athletes could be more likely to produce less reliable scores at younger ages as they are more likely to further develop cognitively.

Hypothesis Three. It was hypothesized that the Reliable Change Indexes (RCIs) would be different between age groups, with RCIs decreasing as age increases.

Justification Three. Currently, ImPACT uses a single RCI value for each composite score to determine significant departure from baseline across all age groups (Iverson, Lovell, & Collins, 2003). These test scores can be impacted by a multitude of factors, including practice effects, regression to the mean, varying rates of neurodevelopment, and other forms of measurement error. Because of these factors impacting test scores, proper interpretation of these scores requires an understanding of the probable range of measurement error regarding change scores. However, there are no published age-based RCIs to be utilized in ImPACT, suggesting a lapse in the current literature regarding possible differences in reliability of change scores across age groups. ImPACT is frequently used in order to make RTP decisions, requiring clinicians to evaluate whether or not departure from baseline scores represents a clinically meaningful change. In order to assess clinically meaningful changes, clinicians currently use the published RCIs across all age groups. In order to determine whether clinically significant changes vary across age groups, it is necessary to evaluate whether RCIs for ImPACT composite scores vary in non-concussed players across baselines in different age groups.

CHAPTER III: Methods

Subjects

Subjects were selected from a de-identified archival database of high-school ImPACT baseline scores in south Florida (N=660). Subjects included in the study were between the ages of 14 and 17 and completed two valid baseline assessments. Subjects with reported concussion history, history of Attention-Deficit/Hyperactivity Disorder, or history of Learning Disability were excluded. Subjects had a mean age of 14.87 (SD=0.78) at initial baseline, and average days between initial and follow-up baselines was 337.8 (SD=68.0; Mdn= 350.0). Notably, racial/ethnic identity was not recorded in the data obtained for this study. Subjects were separated into three groups based on age at first baseline with the following sample sizes: Age 14 at initial baseline: n=249; Age 15 at initial baseline: n=246; Age 16 at initial baseline: n=165. Ninety-eight percent of the Subjects were male. Subjects were included if they completed two valid baselines as determined by ImPACT's embedded performance validity measures between 2013 and 2019.

Measures

ImPACT is a computerized neuropsychological test used to assess several neuropsychological domains as well as concussion symptoms. The test is composed of six individual subtests that yield composite scores for Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control, as well as a Post-Concussion Symptom Scale. Each ImPACT subtest is displayed in a predetermined order to the athlete with the ability to pause between subtests.

Post-Concussion Symptom Scale (PCSS). The PCSS is the symptom inventory completed by examinees prior to beginning ImPACT. The PCSS is completed at baseline as well as at post-injury. The inventory is comprised of 22 common concussion symptoms, rated from 0 (none) to 6 (severe) experienced within 24 hours of testing. A total symptom score is derived from the sum of responses for the 22 items. RTP protocol typically requires that athletes are not cleared to return to play until they are asymptomatic (McCrory et al., 2013). Thus, the PCSS is utilized in the management of concussion by assessing and documenting clinical symptomology in athletes at baseline to determine pre-existing symptoms.

Subtests. The first subtest, Word Memory, evaluates attention and verbal recognition memory. This subtest presents athletes with twelve target words for 750 milliseconds each. Next, they are given a second presentation of the same word list. After the second presentation, the athlete is presented with a randomized list of 12 target and 12 non-target words that have been chosen from the same semantic category. Athletes are asked to discriminate whether the word was a target word by indicating “yes” or “no” on the computer screen. A delay condition (approximately 15 minutes) is administered using the same method (ImPACT, 2018).

The second subtest, Design Memory, evaluates attention and visual recognition memory. This subtest uses a similar process of to that of Word Discrimination. However, it uses target designs rather than target words. In this subtest, the non-target designs are target designs that have been rotated. A delay condition (approximately 15 minutes) is administered for this subtest as well (ImPACT, 2018). For both Word Discrimination and Design Memory, percent scores are provided for correct “yes” and “no” responses.

The third subtest, X’s and O’s, measures working memory and visual processing/motor speed. This subtest requires the athlete practices a distracter task then requires examinees to complete a memory task. The distracter task asks the examinee to press the “Q” key if a blue square appears on the screen, and to press the “P” key if a red circle is presented. During the memory task, the examinee is presented with a random assortment of X's and O's for 1.5 seconds and asked to remember which three X's or O's were highlighted in yellow in their respective locations on the screen. After each presentation of X's and O's, the examinee completes the distracter task. The memory task then reappears and the examinee is instructed to select the X's or O's that were previously highlighted in their location on the screen. Each athlete completes four trials of this subtest, resulting in three scores: correct identification, reaction time during the distracter task, and errors during the distracter task (ImPACT, 2018).

The fourth subtest, Symbol Match, evaluates visual processing speed, learning, and memory. The examinee is presented with a 2x9 grid with 9 common symbols (i.e., square, circle, triangle, etc.) that are paired with a number from 1 to 9 underneath. Below this grid, a symbol is presented and the athlete is asked to select the corresponding number on the grid as quickly and accurately as possible. Correct performance is reinforced through green illumination of the number and incorrect performances produce a red illumination. After completing 27 trials, the symbols disappear from the top grid and the examinee is asked to recall the correct pairing by clicking the corresponding number in the grid to the symbol that appears below. Average reaction time scores and correct memory recognition scores are provided (ImPACT, 2018).

The fifth subtest, Color Match, evaluates choice reaction time and impulse control. The test uses a response inhibition task adapted from Stroop. The subtest begins with a task to ensure the athlete can discriminate between colors red, blue, and green colors. Next, color words (i.e., RED, BLUE, GREEN) are displayed in the same color ink or in a different color ink. The examinee is instructed to click the box only when the word is displayed in the same color ink as quickly as possible. Both reaction time and error scores are provided (ImPACT, 2018).

The final subtest, Three Letters, evaluates working memory and visual-motor speed. Three randomized consonant letters are displayed on the screen and are immediately followed by a distractor task. In the distractor task, examinees use the computer mouse to click in backward order on a 5x5 grid that contains numbers from 1 to 25. The distractor task allows 18 seconds and instructs examinees to work as quickly as possible. Following the distractor task, the examinee is presented with a memory task in which they are asked to recall the three consonants by typing them on the keyboard. The subtest consists of 5 trials and scores include correctly identified letters and the average number of correctly clicked numbers during the distractor task (ImPACT, 2018).

Composite Scores. The results of the six subtests on ImPACT are combined and five composite scores are calculated: Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. The Verbal Memory Composite represents attention, verbal learning, and verbal memory. Scores range from 1 to 100, with higher scores indicating better performance. It is comprised of the average of the following scores: Word Memory percent correct, Symbol Match total correct hidden symbols matched, Three Letters percent total letters correct (ImPACT, 2018).

The Visual Memory Composite represents visual attention, scanning, learning, and memory. Scores range from 1 to 100, with higher scores indicating better performance. It is comprised of an average of the following scores: Design Memory total percent correct and X's and O's total percent correct (ImPACT, 2018).

The Visual Motor Speed Composite represents visual processing, learning and memory, and visual-motor response speed. Scores range from 1 to 100, with higher scores indicating better performance. It is comprised of an average of the following scores: X's and O's total number correct on the distractor task divided by 4 (number correct/4) and Three Letters average numbers counted correctly multiplied by 3 (number correct x 3).

The Reaction Time Composite represents average response speed measured in seconds. Scores range from 0 to 1, with lower scores indicating better performance. It is comprised of an average of the following scores: X's and O's distractor average correct reaction time (RT), Symbol Match average correct RT divided by 3 (avg. correct/3), and Color Match average correct RT (ImPACT, 2018).

The Impulse Control Composite provides a measure of errors and serves as a validity indicator. Scores range from 0 to 132 with lower scores indicating better performance. Each point represents an individual error. It is comprised of the following scores: X's and O's distractor task total incorrect and Color Match total commissions (ImPACT, 2018).

In addition to composite scores, percentile ranks are produced for all composite scales except Impulse Control to assist the clinician's interpretation of the athlete's scores compared within the normative sample. These percentiles are particularly valuable in comparing baseline testing to determine differences at multiple timepoints for each athlete.

The foundational study regarding ImPACT's psychometric properties was conducted by Iverson, Lovell, & Collins in 2003, which examined the psychometric properties of ImPACT. They found no test-retest practice effects after a two-week interval. There were also no significant differences in two-week test-retest for any of the original test composite scores (Verbal Memory: $t(55) = -0.17$, $p < .87$, Visual Memory: $t(55) = 0.85$, $p < .40$, Reaction Time: $t(55) = 0.97$, $p < .34$ and Total Symptom Score: $t(55) = -0.54$, $p < .60$) except for Processing Speed, which is now known as Visual Motor Speed, ($t(55) = -3.26$, $p < .003$) in which 68% of athletes were faster at re-test. More recent findings suggest stronger evidence for ImPACT's reliability (Nakayama, et al., 2014). ICCs were calculated for baseline to day 45, day 45 to day 50, baseline to day 50, and overall. Results indicated all ICCs exceeded the Slick criteria (2006) threshold value of 0.60 (Verbal Memory: 0.76, 0.69, 0.65, and 0.78; Visual Memory: 0.72, 0.66, 0.60, and 0.74; Visual Motor Speed: 0.87, 0.88, 0.85, and 0.91; Reaction Time: 0.67, 0.81, 0.71, and 0.80).

Statistical Analyses

Hypothesis One. It was hypothesized that there would be significant differences between baseline one and baseline two ImPACT composite scores and the Symptom Scale total scores for each of the three specified age groups (i.e. 14-, 15-, and 16-years-old at initial baseline, respectively).

Statistical Analysis: For hypothesis one, three repeated measures t-tests were conducted on data obtained from athletes in each age group who completed two baseline assessments to compare initial and follow-up baseline scores among each age group. ImPACT composite scores (Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control) and the PCSS total score were treated as dependent variables. In order to correct for the potential inflation of Type I Error, an alpha level of 0.01 (.05 divided by 5 dependent variables) was employed.

Hypothesis Two. It was hypothesized that mean difference scores would vary by age group.

Statistical Analysis: For hypothesis two, change scores were compared using a one-way analysis of variance (ANOVA). The independent variable was the age group whereas the dependent variables were the change scores for ImPACT composite scores for one analysis, and

PCSS total score for the second analysis. In order to correct for Type I Error, an alpha level of 0.01 was employed.

Hypothesis Three. It was hypothesized that the Reliable Change Indexes (RCIs) would be different between age groups, with RCIs decreasing as age increases.

Statistical Analysis: For hypothesis three, RCIs were calculated for the three age groups. RCIs were calculated using the methodology outlined by Iverson, Lovell, and Collins (2003). This method requires the calculation of the standard errors of measurement (*SEM*) for each baseline assessment (i.e., SEM_1 and SEM_2), standard error of difference (S_{diff}), and confidence intervals (CIs) associated with S_{diff}. The formula used to calculate RCIs corrects for potential practice effects (Chelune et al., 1993; Iverson & Green, 2001; Iverson, Lovell, & Collins, 2003). In congruence with previous calculations of ImPACT's recommended RCIs, the current study calculated RCIs with both 80% and 90% CIs for each age group. In order to determine meaningful differences in RCIs, CIs were evaluated for overlap; overlapping values were not considered meaningfully different.

Institutional Review Board Requirements

Institutional Review Board (IRB) approval was acquired at Nova Southeastern University prior to the beginning of this study. To remain in accordance with the IRB and the American Psychological Association (APA), all data were de-identified.

CHAPTER IV: Results

Preliminary Analyses

Prior to conducting analyses, the dataset was examined for outliers. To determine the degree to which the distributions of dependent variable deviate from normality, cut off scores between approximately -2 and 2 were employed for skewness and kurtosis (George & Mallery, 2010). Table 1 presents the means (*M*s), standard deviations (*SD*s), skewness, kurtosis, and Cronbach's Alpha (α) for baseline one and baseline two scores in the overall sample. Notably, all baseline one and baseline two variables utilized in the analysis were normally distributed. Internal consistency was poor for all cognitive measures, though the PCSS demonstrated good internal consistency.

Table 1. *Descriptive Statistics: Overall Sample (N = 660)*

Composite	*M*	*SD*	Skewness	Kurtosis	α[a]
Verbal Memory 1	81.63	10.58	-.37	-.18	.38
Verbal Memory 2	83.28	11.02	-.59	-.14	.33
Visual Memory 1	70.42	13.53	-.24	-.33	.40
Visual Memory 2	73.87	13.03	-.29	-.23	.43
Visual Motor Speed 1	32.64	6.48	-.27	.73	.24
Visual Motor Speed 2	35.30	6.87	-.17	.51	.21
Reaction Time 1	.67	0.10	.78	.88	.31
Reaction Time 2	.66	0.10	1.12	1.53	.29
Impulse Control 1	5.93	3.99	1.14	1.25	.09
Impulse Control 2	6.19	4.1	.94	.59	.11
PCSS Total Score 1	2.04	3.15	1.64	1.89	.81
PCSS Total Score 2	2.12	3.38	1.76	2.49	.85

[a]The Verbal Memory and Reaction Time Composites are each comprised of the average of three scores, whereas Visual Memory, Visual Motor Speed, and Impulse Control Composites are each comprised of the average of two scores. The PCSS Total Score is comprised of 22 total items.

In addition to analysis of sample distribution for the overall sample, the same analyses were also employed to determine the distribution of scores in different age groups. Table 2 presents the *M*s, standard deviations *SDs*, skewness, kurtosis, and α for baseline one and baseline two scores for subjects who completed baseline one at age 14. All variables utilized in the analysis were normally distributed. Similar to the overall sample, internal consistency was poor for all cognitive measures and good for the PCSS.

Table 2. *Descriptive Statistics: Age 14 (N = 249)*

Composite	*M*	*SD*	Skewness	Kurtosis	α
Verbal Memory 1	80.37	10.80	-.48	.08	.43
Verbal Memory 2	82.61	10.57	-.49	-.12	.33
Visual Memory 1	69.80	13.22	-.21	-.38	.25
Visual Memory 2	73.96	13.75	-.40	-.35	.50
Visual Motor Speed 1	32.40	6.19	-.11	.93	.24
Visual Motor Speed 2	35.10	6.03	.46	-.10	.18
Reaction Time 1	.67	.09	.77	.90	.35
Reaction Time 2	.65	0.10	1.14	1.50	.47
Impulse Control 1	6.34	3.97	.99	.79	.11
Impulse Control 2	6.74	4.23	.75	.17	.13
PCSS Total Score 1	1.51	2.55	1.70	1.85	.81
PCSS Total Score 2	1.73	2.79	1.69	1.94	.86

Table 3 presents the *M*s, *SDs*, skewness, and kurtosis for baseline one and baseline two scores for subjects who completed baseline one at age 15. All variables utilized in the analysis were normally distributed. Internal consistency was again poor for cognitive measures and good for the PCSS.

Table 3. *Descriptive Statistics: Age 15 (N = 246)*

Composite	*M*	*SD*	Skewness	Kurtosis	α
Verbal Memory 1	81.77	9.92	-.10	-.66	.29
Verbal Memory 2	83.23	11.06	-.53	-.46	.28
Visual Memory 1	69.74	13.70	-.20	-.44	.43
Visual Memory 2	73.79	11.79	-.12	-.21	.30
Visual Motor Speed 1	32.21	6.73	-.38	.63	.26
Visual Motor Speed 2	34.85	7.41	-.29	.38	.20
Reaction Time 1	.67	.10	.76	1.11	.20
Reaction Time 2	.66	0.10	.95	1.18	.28
Impulse Control 1	5.55	3.95	1.33	1.71	.07
Impulse Control 2	5.86	4.03	1.04	.70	.11
PCSS Total Score 1	2.30	3.38	1.54	1.43	.83
PCSS Total Score 2	2.52	3.77	1.66	2.12	.81

Table 4 presents the *M*s, *SDs*, skewness, and kurtosis for baseline one and baseline two scores for subjects who completed baseline one at age 16. All variables utilized in the analysis were normally distributed. Consistent with previous groups, internal consistency was poor for cognitive measures and good for the PCSS.

Table 4. *Descriptive Statistics: Age 16 (N = 165)*

Composite	*M*	*SD*	Skewness	Kurtosis	α
Verbal Memory 1	83.30	10.99	-.56	-.09	.40
Verbal Memory 2	84.37	11.61	-.85	.43	.40
Visual Memory 1	72.39	13.63	-.37	.02	.56
Visual Memory 2	73.84	13.72	-.30	-.18	.49
Visual Motor Speed 1	33.67	6.47	-.31	.73	.21
Visual Motor Speed 2	36.27	7.16	-.54	1.13	.26
Reaction Time 1	.66	.11	.85	.60	.42
Reaction Time 2	.67	.12	1.20	1.52	.18
Impulse Control 1	5.90	4.03	1.17	1.67	.07
Impulse Control 2	5.85	3.94	1.11	1.54	.06
PCSS Total Score 1	2.43	3.51	1.48	1.25	.74
PCSS Total Score 2	2.10	3.51	1.71	1.92	.87

Results of preliminary analyses indicate all variables utilized in the study were normally distributed within the overall sample and within each age group. As such, all variables were included in future analyses.

Data Analysis

Analyses were conducted using IBM SPSS Statistics V.28 statistical software.

Hypothesis One

It was hypothesized that there would be significant differences between baseline one and baseline two ImPACT composite scores and the PCSS total scores for each of the three specified age groups (i.e. 14-, 15-, and 16-years-old at baseline one, respectively).

Repeated measures t-tests were conducted to compare baseline one and baseline two scores within subjects for each age group. ImPACT composite scores (Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control) and the PCSS total score were considered dependent variables. In order to correct for Type I Error, an alpha level of 0.01 was employed. Table 5 presents the results of repeated measures t-tests for subjects who completed baseline one at age 14.

Table 5. *Repeated Measures t-test Results: Age 14 at Baseline One (N = 249)*

	Paired Differences			Significance	Effect Size
Composite	*M*	*SD*	*t*(248)	Two-Tailed *p*	Cohen's *d*
Verbal Memory	**-2.24**	**11.83**	**-2.99**	**.003**	**.19**
Visual Memory	**-4.17**	**14.72**	**-4.47**	**<.001**	**.28**
Visual Motor Speed	**-2.70**	**5.22**	**-8.18**	**<.001**	**.52**
Reaction Time	.01	.11	2.03	.044	.13
Impulse Control	-.40	4.27	-1.48	.139	.09
PCSS	-.22	2.65	-1.29	.198	.08

The null hypothesis of equal performance for baseline one and baseline two was rejected for the Verbal Memory Composite ($t(248) = -2.99$, $p = .003$), the Visual Memory Composite ($t(248) = -4.47$, $p < .001$), and the Visual Motor Speed Composite ($t(248) = -8.18$, $p < .001$) for subjects who completed baseline one at 14-years-old. The Visual Motor Speed Composite demonstrated a medium effect (Cohen's $d = .52$), whereas the Visual Memory Composite (Cohen's $d = .28$) and the Verbal Memory Composite (Cohen's *d* = .19) demonstrated small effects. The Reaction Time Composite, Impulse Control Composite, and PCSS Total Score did not have significant differences between baseline one and baseline two for subjects who completed baseline one at 14-years old. The hypothesis was only partially supported in this group.

Table 6 presents the results of repeated measures t-tests for subjects who completed baseline one at age 15.

Table 6. *Repeated Measures t-test Results: Age 15 at Baseline One (N = 246)*

	Paired Differences			Significance	Effect Size
Composite	*M*	*SD*	*t*(245)	Two-Tailed *p*	Cohen's *d*
Verbal Memory	-1.46	11.31	-2.02	.044	.13
Visual Memory	**-4.05**	**13.20**	**-4.82**	**<.001**	**.31**
Visual Motor Speed	**-2.65**	**5.26**	**-7.89**	**<.001**	**.50**
Reaction Time	.01	.11	1.89	.059	.12
Impulse Control	-.31	4.17	-1.18	.240	.08
PCSS Total Score	-.22	3.76	-.92	.360	.06

The null hypothesis of equal performance for baseline one and baseline two was rejected for the Visual Memory Composite ($t(245) = -4.82$, $p < .001$) and the Visual Motor Speed Composite ($t(245) = -7.89$, $p < .001$) for subjects who completed baseline one at 15-years-old. The Visual Motor Speed Composite demonstrated a medium effect (Cohen's *d* = .50), whereas the Visual Memory Composite (Cohen's *d* = .31) demonstrated a small effect. The Verbal Memory Composite, Reaction Time Composite, Impulse Control Composite, and PCSS Total Score did not have significant differences between baseline one and baseline two for subjects who completed baseline one at 15-years old. The hypothesis was only partially supported in this group.

Table 7 presents the results of repeated measures t-tests for subjects who completed baseline one at age 16.

Table 7. *Repeated Measures t-test Results: Age 16 at Baseline One (N = 165)*

	Paired Differences			Significance	Effect Size
Composite	*M*	*SD*	*t*(245)	Two-Tailed *p*	Cohen's *d*
Verbal Memory	-1.07	12.10	-1.13	.259	.09
Visual Memory	-1.44	13.27	-1.40	.169	.11
Visual Motor Speed	**-2.60**	**6.28**	**-5.32**	**<.001**	**.41**
Reaction Time	-.01	.13	-.97	.333	.08
Impulse Control	.05	4.13	.17	.866	.01
PCSS Total Score	.33	3.85	1.09	.276	.08

The null hypothesis of equal performance for baseline one and baseline two was rejected for the Visual Motor Speed Composite ($t(164) = -5.32$, $p < .001$) for subjects who completed baseline one at 16-years-old. The Visual Motor Speed Composite demonstrated a small to medium effect (Cohen's $d = .41$). The Verbal Memory Composite, Visual Memory Composite, Reaction Time Composite, Impulse Control Composite, and PCSS Total Score did not have significant differences between baseline one and baseline two for subjects who completed baseline one at 16-years old. The hypothesis was only partially supported in this group. Overall, only 6 of the 18 comparisons supported the original hypothesis.

Hypothesis Two

It was hypothesized that there would be significant differences in change scores between age groups. In order to test this hypothesis, change scores were compared using a one-way analysis of variance (ANOVA). The independent variable was age at baseline one and the dependent variables were the change scores for ImPACT composite and PCSS scores. In order to correct for Type I Error, an alpha level of 0.01 was employed. Table 8 presents the results of the ANOVA between age groups.

Table 8. *Test of Between-Groups Differences in ImPACT Composite Change Scores (N = 660)*

Composite	F	df	Significance	Cohen's *f*
Verbal Memory	.55	2, 657	.57	.04
Visual Memory	2.32	2, 657	.10	.08
Visual Motor Speed	.02	2, 657	.98	.01
Reaction Time	2.58	2, 657	.08	.09
Impulse Control	.62	2, 657	.54	.04
PCSS	1.59	2, 657	.21	.07

The Verbal Memory Composite, Reaction Time Composite, Visual Motor Speed Composite, Impulse Control Composite, and PCSS Total Score did not have significant differences in change scores between age groups. The effect sizes for group differences were negligible for all comparisons. The hypothesis of significant differences in change scores across age groups was not supported in any comparison.

Hypothesis Three

It was hypothesized that the reliable change indexes (RCIs) would be different between age groups, with RCIs decreasing as age increases. In order to test this hypothesis, RCIs for ImPACT composite and PCSS total scores were calculated with both 80% and 90% confidence intervals (CIs) for each age group. In order to evaluate meaningful differences between RCIs and reliability of difference scores, CIs were evaluated for overlap; overlapping values were not considered meaningfully different. Table 9 presents the Pearson test-retest correlation coefficients, standard errors of measurement (*SEMs*), standard errors of difference (S_{diffs}), and RCI confidence intervals for ImPACT composites and PCSS total scores for each age group.

Table 9. *Descriptive Statistics, Pearson r, SEMs, S_{diffs}, and Reliable Change Confidence Intervals*

Composite	Age	*M (SD)*		*r*	*SEM_1*	*SEM_2*	S_{diff}	RCI Confidence Intervals	
		Baseline 1	*Baseline 2*					0.80	0.90
Verbal Memory	14	80.37 (10.80)	82.61 (10.57)	.39	8.44	8.26	11.80	±15.11	±19.36
	15	81.77 (9.92)	83.23 (11.06)	.42	7.55	8.42	11.31	±14.48	±18.56
	16	83.30 (10.99)	84.37 (11.61)	.43	8.30	8.77	12.07	±15.45	±19.79
Visual Memory	14	69.80 (13.22)	73.96 (13.75)	.41	10.15	10.56	14.65	±18.75	±24.03
	15	69.74 (13.70)	73.79 (11.79)	.47	9.97	8.58	13.16	±16.84	±21.58
	16	72.39 (13.63)	73.84 (13.72)	.53	9.34	9.41	13.26	±16.97	±21.74
Visual Motor Speed	14	32.40 (6.19)	35.10 (6.03)	.64	3.71	3.62	5.18	±6.64	±8.50
	15	32.21 (6.73)	34.85 (7.41)	.73	3.50	3.85	5.20	±6.66	±8.53
	16	33.67 (6.47)	36.27 (7.16)	.58	4.19	4.64	6.25	±8.01	±10.26
Reaction Time	14	.67 (.09)	.65 (.10)	.33	.07	0.08	0.11	±0.14	±0.18
	15	.67 (.10)	.66 (.10)	.37	.08	.08	0.11	±0.14	±0.18
	16	.66 (.11)	.67 (.12)	.35	0.09	0.10	0.13	±0.17	±0.22
Impulse Control	14	6.34 (3.97)	6.74 (4.23)	.46	2.92	3.11	4.26	±5.46	±6.99
	15	5.55 (3.95)	5.86 (4.03)	.45	2.93	2.99	4.18	±5.36	±6.86
	16	5.90 (4.03)	5.85 (3.94)	.46	2.96	2.90	4.14	±5.30	±6.79
PCSS	14	1.51 (2.55)	1.73 (2.79)	.51	1.79	1.95	2.65	±3.39	±4.34
	15	2.30 (3.38)	2.52 (3.77)	.45	2.51	2.80	3.76	±4.81	±6.16
	16	2.30 (3.38)	2.52 (3.77)	.45	2.51	2.80	3.76	±4.81	±6.16

Note. SEM: standard error of measurement; S_{diff}: standard error of difference

The 80% and 90% confidence intervals for evaluating statistically significant change have varying degrees of overlap across age groups for all composite scores and PCSS total score. Although alpha values were not calculated in this analysis, the significant overlap of RCI values indicates negligible differences across age groups. As such, the hypothesis of meaningful differences in RCIs across age groups was not supported for any composite scores or the PCSS total score.

CHAPTER V: Discussion

The purpose of the current study was to determine the long-term stability of baseline ImPACT scores among high school football players completing baselines at different ages. Results of this study offer unique information regarding normal stability and variability of baseline ImPACT performances between age groups of high school football players who have not sustained a traumatic brain injury. Clinicians are encouraged to use results of this study to inform practice for baseline assessments as well as interpretation of expected departure from baseline performance in return to play decisions. Additionally, clinicians should use results of this study to more accurately track typical development of high school football players in certain age groups, further increasing the accuracy of concussion classification, treatment progress, and return to play decisions.

Hypothesis One

In hypothesis one, it was hypothesized that there would be significant differences between baseline one and baseline two ImPACT composite scores and the PCSS total scores for each of the three specified age groups (i.e. 14-, 15-, and 16-years-old at baseline one, respectively). The hypothesis was partially supported, as 6 of the 18 comparisons supported the original hypothesis.

Age 14 at Baseline One

For subjects who completed baseline one at age 14, the statistically significant and clinically meaningful improvements in processing speed and verbal and visual memory, as well as stability in reaction time, have a multitude of important clinical, research, and theoretical implications.

Clinically, the improvements in multiple domains demonstrated by those who completed baseline one at age 14 provide important information regarding the interpretation of change following concussion. Given these improvements, it is essential for clinicians to carefully consider the time since baseline for this age group. For example, if a baseline was obtained a year or more prior to injury, it is likely that those scores no longer represent the athlete's baseline cognitive abilities. If clinicians consider a year-old baseline an accurate representation of cognitive skills, it is possible that athletes may be cleared to return to play prior to returning to their true cognitive baseline. These potential errors in clinical decision making can have a widespread and severe impact on athlete health and recovery. Such clinical errors are likely to result in a lowered threshold for future concussions and significantly increased recovery time for subsequent injuries.

Given the importance of return to play decision making and the need for accurate baseline information, it is also essential to consider implications for baseline policies. As discussed in the introduction, it is common practice for adolescent athletes to complete baselines annually. Based on the improvements demonstrated by those in the 14-year-old group and the importance of accurate clinical decisions, adolescent baseline policies should be reconsidered. Results of this study indicate baselines should be conducted prior to each athletic season for 14-year-olds, rather than annually. Because of the continuous, rapid neurodevelopment occurring at 14-years-old, more frequent baselines would offer more accurate estimates of cognitive abilities. Obtaining baselines at the outset of each athletic season would ensure that baselines are obtained within a recent timeframe for athletes sustaining concussions during the season.

In addition to the improvements noted above, it is also important to note the stability of the Reaction Time composite over the course of a year. Although many cognitive skills significantly change over a 1-year interval, the stability of reaction time indicates that this measurement can be an important indicator when tracking concussion recovery. Because the Reaction Time composite is not expected to significantly change outside of sustaining a concussion, clinicians can use reaction time as a reliable indicator of cognitive change, even when using remote (i.e., up to ~1 year) baseline data. These results indicate that although seasonal baselines are recommended, annual baselines will still provide useful clinical information. As such, annual baselines are recommended as a secondary option for clinicians if seasonal baselines are not feasible.

Beyond implications for clinical management of concussion, results also have important theoretical implications for neuropsychological changes among 14-year-old high school football players. Athletes who completed baseline one at 14-years-old demonstrated significant improvements in Visual Memory, Verbal Memory, and Visual Motor Speed over the course of a year. These results indicate that these cognitive functions, including memory and processing speed, continue to develop up to at least age 15. In contrast, results indicate development of reaction time plateaus by age 15 or earlier.

Age 15 at Baseline One

For subjects who completed baseline one at age 15, the statistically significant and clinically meaningful improvements in visual memory and processing speed, as well as stability in verbal memory and reaction time, have important clinical, research, and theoretical implications.

This combination of stability and improvements in multiple domains demonstrated by subjects who completed baseline one at age 15 provides useful clinical information in terms of interpreting change following concussion. Because of the continued development in Visual

Memory and Visual Motor Speed for those who completed baseline one at age 15, it remains important for clinicians to carefully consider time since baseline for this age group. Despite the stability for Verbal Memory and Reaction Time for those who completed baseline one at age 15, the noted improvements in other domains indicate changes in cognitive skills that can have important implications when considering return to play decisions. As previously noted, errors in clinical decision making can have a widespread and severe impact on athlete health and recovery (e.g., lowered concussion threshold and increased recovery time).

Given the importance of baseline cognitive skills in clinical decision making and the continued improvements in cognitive abilities for those completing baseline one at age 15, baseline policies for this age group should also be reconsidered. Consistent with the 14-year-old group, results indicate baselines should be conducted prior to each athletic season for 15-year-olds. Because of the continued neurodevelopment occurring through 15-years-old, more frequent baselines would offer more accurate estimates of cognitive abilities, subsequently reducing the likelihood of error in return to play decision making.

In addition to the improvements in cognitive domains for the 15-year-old group, it is important to also recognize the stabilization of the Verbal Memory composite by the second baseline (i.e., age 16), as well as continued stability of the Reaction Time composite. Despite the changes in other cognitive domains, these areas appear to stabilize by age 16 for those who completed their initial baseline at age 15. The stability of these composite scores indicates that these measurements can be an important indicator when tracking concussion recovery as they are less likely to significantly change when using old (i.e., up to ~1 year) baselines. Consistent with recommended baseline policies for the14-year-old group, these results indicate that although seasonal baselines are recommended, annual baselines will still provide useful clinical information. Both Verbal Memory and Reaction Time composite scores can be used as reliable measures to detect change within a year of baseline completion for athletes completing baseline at age 15. As such, annual baselines are recommended as a secondary option for clinicians if seasonal baselines are not feasible.

Beyond clinical implications in concussion management, results also provide important information regarding neuropsychological changes among 15-year-old high school football players. Subjects who completed baseline one at 15-years-old demonstrated significant improvements in Visual Memory and Visual Motor Speed over the course of a year. These results indicate these cognitive functions, including visual memory and processing speed, continue to develop up to at least age 16. In contrast, results indicate that verbal memory skills plateau by age 16, and reaction time does not continue to develop beyond age 15. These results highlight the variability in development of different cognitive skills throughout adolescence.

Age 16 at Baseline One

For those who completed baseline one at age 16, the statistically significant and clinically meaningful improvement in processing speed, as well the as stabilization across all other measured domains have a multitude of important clinical implications.

The combination of stability and improvement in processing speed demonstrated by subjects who completed baseline one at age 16 provides critical clinical information in terms of interpreting change following concussion. Because of the continued development in Visual Motor Speed for those who completed baseline one at age 16, it remains important for clinicians to carefully consider time since baseline for this age group. Despite the continued improvement in Visual Motor Speed from baseline one to baseline two, the noted stability in all other measured domains indicates that cognitive skills have largely plateaued by baseline two in this group (i.e., age 17). The general stability of baseline scores has important implications for baseline assessment practices.

As previously noted, accurate cognitive baseline information is essential in making informed clinical decisions and maintaining player safety upon returning to play. Currently, it is common practice for adolescent athletes to complete baselines annually. Given the general stability in measured cognitive abilities for those completing baseline one at age 16, results indicate annual baselines are appropriate for this group. Because of the overall stability of cognitive performance at 17-years-old, annual baselines would provide accurate estimates of cognitive abilities and subsequently reduce the likelihood of error in return to play decision making. However, it remains important for clinicians to consider the continued improvement in Visual Motor Speed when evaluating return to baseline for this group.

Beyond implications for clinical management of concussion, results also provide important information regarding neuropsychological changes among 16-year-old high school football players. Athletes who completed baseline one at 16-years-old demonstrated significant improvements in Visual Motor Speed over the course of a year, indicating that processing speed continues to develop up to at least age 17. In contrast, results indicate that visual and verbal memory skills plateau by age 16, and reaction time does not continue to develop beyond age 16. These results highlight the general plateau of many cognitive skills by age 17 with continued development of visual motor processing speed up to at least age 17.

Overall

In sum, these results (e.g., variability at one year retest for ages 14 and 15) indicate that the standard of practice for baseline assessment should be at the outset of every athletic season for those aged 14 and 15. Due to stability in some cognitive domains for these age groups, annual baselines may be implemented when resources are limited. Given the increase in stability of scores for 16-year-olds, annual baseline testing is recommended as the continued standard of practice for this age group, with clinicians acknowledging that Visual Motor Speed continues to improve beyond age 16. The use of more frequent baselines for younger athletes aims to account for the differences in development observed between age groups within this study.

Although the recommended standard of practice for obtaining baseline assessments is either seasonally (i.e., age 14 and 15) or annually (i.e., age 16), it is important to consider barriers to frequent baseline assessments. Resources required for baseline protocols, such as financial support, personnel, and testing space, can all represent barriers to baseline assessment. With both the present study's results and the resources required for baseline testing in mind, it is important to note that if resources are limited, baseline assessments can be foregone for adolescent athletes at any age due to the continuous changes in cognitive abilities across multiple domains, as well as alternative methods to determine pre-injury functioning (e.g., standardized academic testing, GPA). This is not meant to deter athletes and organizations from completing baselines if provided with the appropriate resources, as it is optimal to complete baseline assessments as recommended above. Although baseline assessments are best utilized when obtained according to the recommended standard of practice, it is essential to recognize that concussions can be treated without the use of a baseline assessment.

Given the clinical implications of these results, there are consequent research needs in order to establish scientifically sound practices. Because of the recommended increase in baseline assessment frequency, it is critical to evaluate the psychometric properties of ImPACT within closer baseline re-test intervals. Such research is essential in establishing important psychometric indicators including test-retest coefficients, RCIs, and influence of practice effects in order to adequately evaluate change. Beyond basic psychometric properties, it is particularly important for clinicians to be aware of the potential influence of practice effects during serial assessments.

In addition to psychometric properties, it is also crucial to evaluate differences among demographic variables (e.g., gender, race, and exposure to repeated head impacts) and their role in cognitive development. Such differences were not evaluated in the current study, though they are well-known to play an important role in neuropsychological performance in adult populations. It is important to consider the potential importance of these factors can have in cognitive functioning throughout adolescent development. With more data regarding potential differences in ImPACT performance among these groups, it will allow for more accurate clinical assessment and management of concussion. Moreover, such data would have relevant theoretical implications regarding potential differences in adolescent cognitive development among various demographic groups.

Beyond clinical and research implications, it is also important to consider the theoretical implications for neuropsychological development in adolescents. Of important note, this pattern of adolescent cognitive development (i.e., stability in some areas and continued development in others) is generally consistent with Luria’s theory of brain organization and function. Luria’s theory posits that the sequence of development of cognitive function is dependent upon the physical and functional changes that occur in the natural maturation of different cortical areas. The theory maintains that the development throughout each stage is consistent with qualitative organizational changes in adaptive and intellectual skills. Although largely established by 12-years-old, cognitive development continues throughout adolescence and into adulthood.

A trajectory of cognitive development similar to that proposed by Luria was observed among the study's age groups across measured neuropsychological domains. In the context of this study, the Reaction Time composite did not significantly change among any of the age groups. This pattern indicates reaction time is a foundational cognitive skill that is fully developed prior to age 14. In contrast, other cognitive domains such as visual memory, verbal memory, and visual motor speed varied in their development trajectory among all three age groups. This pattern of neuropsychological performance is likely reflective of the continued maturation of respective cortical areas throughout adolescence.

This continued development in cognitive skills is especially important to consider in high school football players due to repeated exposure to sub-concussive impacts throughout a season. In recent years, there has been an increasing awareness of the neuropsychological impacts of both concussion and sub-concussive impacts, with concern regarding the impact of such injuries on neuropsychological development. Though subjects with a diagnosed concussion were excluded from the present study, high school football players are likely to undergo many sub-concussive impacts throughout the course of a season. The observed improvements indicate high school football players are continuing to develop cognitive skills despite presumably experiencing sub-concussive impacts. Although these results highlight continued development, it is important to note that there is no control group for comparison to determine normal rate of development among adolescents.

In considering trajectory of neuropsychological development, it is important to also highlight the ImPACT scores that did not significantly change in any group. The Reaction Time Composite, Impulse Control Composite, and PCSS Total Score remained stable across all age groups. For the Reaction Time Composite, these results demonstrate negligible change in reaction time throughout adolescence. Subjects who may be susceptible to changes in reaction time, such as those with ADHD and/or learning disabilities, were excluded from the study. For the Impulse Control Composite, it is important to recall that this serves as a measure of errors and is used as a validity indicator. Because invalid cases were not included in the current study, it is logical for the Impulse Control Composite score to remain stable. The stability of PCSS scores across age groups indicates consistent symptom reporting in a valid sample.

Hypothesis Two

In hypothesis two, it was theorized that there would be significant differences in change scores between age groups due to varying rates of neurodevelopment between age groups. This hypothesis was not supported in any comparison. These results have important theoretical implications which consequently have wide-ranging impacts on clinical and research aspects of the assessment and treatment of concussion.

In terms of theoretical implications, these results provide important information regarding the conceptualization of neurodevelopment among adolescents. Broadly, the absence of significant differences in change scores indicates that the rate of neurodevelopment is generally consistent among high school football players aged 14 to 17. Similar to the results in hypothesis one, this pattern of results is again congruent with Luria's theory of brain organization and function.

As described previously, Luria's theory indicates that the sequence of development of cognitive function is dependent upon the natural maturation of different cortical areas, which in turn results in changes in adaptive and intellectual skills. Importantly, these skills are thought to be largely established by 12-years-old, though cognitive development continues throughout adolescence and into adulthood. The current results indicate that although some of the cognitive domains measured by ImPACT are not fully developed by 12-years-old, the rate at which these which these cognitive skills develop beyond 12-years-old is similar for athletes aged 14 to 17.

The similar rate of development for athletes aged 14 to 17 demonstrates that the magnitude of annual improvement in cognitive skills among these age groups is generally consistent among age groups after age 12. The lack of significant differences in rate of neurodevelopment is consistent with the theory of cognitive development proposed by Luria, as Luria's theory suggests that the majority of development in cognitive skills occurs by age 12. According to Luria's theory of cognitive development, we would be more likely to observe significant differences in changes scores between age groups prior 12, as that is when the most rapid neurodevelopmental changes occur. Because intelligence is largely established by age 12, it is logical that there are no significant differences in annual improvements among age groups from age 14 to 17.

Given the theoretical implications of these results, there are important clinical considerations for the assessment and treatment of concussion. Most importantly, these results provide support for the use of annual baseline assessments when indicated (i.e., for athletes aged 16 or older, or for when resources are limited for younger athletes). Because there are no significant differences in change scores between age groups, there is limited evidence to support changes to previously described baseline assessment recommendations.

Had results indicated significant differences in change scores, there would have been further evidence to support more frequent baselines for athletes whose change scores are expected to be significantly different from others. For example, had 14-year-old athletes demonstrated greater change scores than 15- or 16-year-old athletes, it would provide further evidence for more frequent baselines at younger ages. However, without evidence for increased variability between groups, there is no additional evidence to support previously mentioned baselines policy recommendations. Conversely, it is important to note that the current results (i.e., no evidence for between group variability in change scores) do not indicate that baselines should be conducted less often, as hypothesis one results demonstrate statistically significant and clinically meaningful variability in multiple domains of baseline performance between groups. The absence of statistically significant variability in change scores alone does not indicate congruent neurodevelopmental trajectories among age groups.

In addition to clinical implications related to baseline frequency, these results provide valuable evidence for interpretation of change following concussion. Given continued development of some cognitive skills throughout adolescence, the absence of significantly different change scores across age groups provides important information about the expected departure from baseline scores. Specifically, these results highlight the need for clinicians to consider time since baseline and potential improvements in cognitive skills when estimating premorbid functioning in concussed athletes.

The importance of accurate cognitive baseline information in return to play decision making is well-established. When considering the accuracy of a previous baseline in post-injury assessment and treatment, it is essential to also consider the time since baseline and resultant estimated improvements in cognitive abilities that have occurred within the context of adolescent neurodevelopment. The current results provide important information regarding estimated improvements in baseline performance across multiple age groups in adolescence. Importantly, these results indicate that although some cognitive areas are less stable at earlier ages, the year-to-year improvements in Verbal Memory, Visual Memory, Visual Motor Speed, and Reaction Time do not significantly differ between age groups. As a result, it is important to consider the theoretical underpinnings of the hypothesis for different rates of cognitive development as well as how such biases may impact clinical decision making.

It is important to note that these results are not consistent with the hypothesis, which indicated change scores would differ between groups. The proposed theoretical mechanism for significant differences in change scores between groups was rooted in the ongoing physical development and hormonal changes occurring throughout ages 14 to 17, as well as continuing neurodevelopment into adulthood. As noted above, intellectual abilities are thought to be established by 12 years old with some cognitive development continuing into adulthood. This continued cognitive development into adulthood is dependent upon the physical and functional changes that occur in the natural maturation of different cortical areas. Because of the logical theoretical underpinnings of the hypothesis, it is important for clinicians to remain aware of potential biases their estimates of change since baseline when determining premorbid functioning.

The results of the current study underscore the need for clinicians to utilize an evidence-based approach in the assessment and treatment of concussion. As assessment and treatment relate to results of hypothesis two, it is important for clinicians to not assume that younger athletes will have greater change scores than older athletes when utilizing distant neurocognitive baseline results. Such erroneous conclusions can lead to poor outcomes for athletes whose baseline abilities are not accurately represented by remote (e.g., completed ≥1 year ago) baseline scores. Clinicians underestimating the improvement in baseline cognitive abilities for athletes who completed baselines ≥1 year ago are susceptible to Type II errors as players who have not fully returned to their true baseline may be cleared for play due to inaccurate estimates of baseline abilities. These errors are likely to disproportionately impact older adolescent athletes who may be perceived to have lower rates of cognitive development than their younger peers.

In addition to the clinical implications of the absence of differences in cognitive performance, it is also important to recognize the importance of the absence in differences between change scores for reported symptoms on the Post-Concussion Symptom Scale (PCSS). These results indicate that adolescents are not reporting increased concussion symptoms at baseline at different age groups. Though the current study did not examine the effect of years played on PCSS total score, the lack of significant differences in change scores among football players in different age groups suggests that baseline symptoms of chronic pain and depression do not increase as a result of years played in high school football athletes.

In consideration of these clinical implications, future research should continue to focus on change scores across various age groups. Such research should seek to replicate the current findings in larger, more heterogenous samples. Specifically, research should use samples including a more diverse array of sport participation as well as more diversity in gender, race, and education. With more heterogenous samples, researchers can identify potential covariates that may influence change scores on ImPACT throughout adolescence.

In addition to research on the impact of aforementioned demographic variables on change scores throughout adolescent development, research should focus on the elucidating the relationship of age of first exposure to tackle football and subsequent rate of cognitive development. Currently, many initiatives are working to outlaw tackle football prior to age 14 in the absence of minimal empirical support. Given Luria’s theory of cognitive development, foundational cognitive functions should be established by 12 years old, but the potential influence of repetitive sub-concussive impacts during adolescence has yet to be empirically determined.

Hypothesis Three

In hypothesis three, it was hypothesized that the RCIs would be different between age groups, with RCIs decreasing as age increases. Given the overlapping RCI values in all comparisons, the hypothesis of meaningful differences in RCIs across age groups was not supported for any composite scores or the PCSS total score. In addition, RCIs did not decrease as age increased. These results have important implications for clinical, theoretical, and research aspects of the assessment and treatment of concussion.

Overall, these results indicate negligible differences in range of measurement error at one year intervals, regardless of the athlete's age. The hypothesis of differing RCIs across age group was based on the potential impact of varying rates of neurodevelopment on expected measurement error in baseline assessments and post-injury return to baseline. There are no published age-based RCIs to be utilized in ImPACT, suggesting a potential lapse in the current literature regarding possible differences in expected RCIs across age groups. However, current results indicate there are no meaningful differences in reliable change estimates among adolescent football players. In light of these results, the use of uniform RCIs for adolescents and remains the recommended practice in determining return to baseline following injury.

Although the results do not support the hypothesis of differences in RCIs between age groups, it is important to recognize the wide range for RCIs across all composite scores and age groups. Notably, the number of scores used to comprise each composite score is relatively low (e.g., two to three items). In addition, the interitem correlations between scores used to calculate composite scores were low across all measures of cognitive functioning. As a result of the small number of items used to calculate composite scores, as well as their low interitem correlations, internal consistency was poor across all composite scores. Given the poor internal consistency of composite scores, there is increased likelihood of measurement error leading to the wide-ranging RCIs observed in this study. It is important to recognize that although internal consistency is a measure of reliability, it is influenced by the length of the test. ImPACT's abbreviated length and low number of scores comprising composites likely leads to decreased internal consistency.

Of important note, the RCIs calculated in this study are significantly broader than those specified by ImPACT, as their standard RCIs were derived from test-retest intervals ranging from 3 to 11 days. Because we cannot predict when concussions will occur, it is common practice for high school athletes to complete a single pre-season baseline either each year or every two years. Given the wide ranging RCIs across composite scores (often greater than one standard deviation in baseline performance), it is worth considering the clinical utility of baselines that were completed at or beyond one year prior to injury, especially in adolescents. Due to the lack of specificity in the RCIs calculated in the current sample, it would be difficult to accurately detect change following injury when using year old baselines. The non-specific nature of the RCIs obtained in hypothesis three again raise important questions about how to best approach baseline assessments.

In order to determine the best practice for frequency and intervals between baselines, it is important to consider the results of this study as a whole. Results of the first hypothesis indicate that although some scores remain stable, important cognitive skills such as verbal memory, visual memory, and visual motor speed can meaningfully change from year to year in healthy individuals who have not sustained a concussion. Results also indicate that as athletes develop, their performance on measures of verbal memory and visual memory become more stable from year to year. Results of hypothesis three (e.g., broad year to year RCIs across all age groups and composite scores), combined with variability at one year retest for ages 14 and 15 observed in hypothesis one, indicate that baseline testing should ideally occur at the outset of every athletic season for those aged 14 and 15. In slight contrast, the increase in stability of scores for 16-year-olds observed in hypothesis one indicates annual baseline testing is appropriate despite wide ranging RCIs. The use of more frequent baselines for younger athletes aims to account for the differences in development observed between age groups within this study, as well as reduce the potential measurement error by producing narrower RCIs.

In addition to implications for baseline frequency, these results also raise important questions about interpretation of ImPACT scores in return to play decision making following concussion. For baseline assessments completed a year prior to injury, current results indicate the range for normal measurement error is quite broad with no meaningful differentiation between age groups. Given such a broad range of measurement error in adolescent baselines from year to year, it is recommended that clinicians use ImPACT's specified RCIs, rather than the RCIs obtained in hypothesis three, when evaluating athletes post-injury. Because of the non-specific RCIs obtained with annual baselines, clinicians are best equipped with specific RCIs that reflect expected measurement error within shorter intervals that are typical of repeated assessment following injury. However, clinicians should use ImPACT's published RCIs with the expectation that select composite scores should have improved in the time since baseline for normally developing adolescents.

Because these results indicate the need to use a combination of ImPACT's published RCIs and expected improvements since baseline, it is essential to again consider the clinical importance of the stable baseline composite scores across age groups. As noted previously, the

Reaction Time composite score is the one cognitive domain measured by ImPACT that does not significantly change year-to-year for any adolescent age group in this study. In order to simplify clinical interpretation and limit errors in estimated improvements in other composite scores, the Reaction Time composite can be relied on as a stable measure of cognitive ability following injury. Using measures expected to remain stable from remote baselines will help to limit the likelihood of erroneously returning athletes to play prior to return to baseline cognitive functioning.

In addition to the clinical implications of these results, there are also important underlying theoretical implications. The most salient theoretical implication relates to the utility of RCIs when using computerized neurocognitive measures such as ImPACT.

Jacobson and Truax (1991) stated that the RCI "tells us whether change reflects more than the fluctuations of an imprecise measuring instrument." In other words, the RCI quantifies our ability to determine that two observed scores reflect different true scores, given the possibility that the assessment might have produced different scores if we had obtained more than one measurement at each timepoint. With this in mind, the RCI essentially serves as a substitute for more data. Given the importance of determining differences in baseline and post-injury cognitive performance and the lack of optimal data, RCIs provide a way to estimate meaningful change in true scores without access to optimal data.

Though RCIs have been a useful tool when evaluating concussion recovery on cognitive assessments, the RCIs obtained in this study provide strikingly low sensitivity. This low sensitivity is likely due to the changes in scores observed in performance on annual baselines across age groups, as well as the poor internal consistency across all cognitive composite scores. Further, there are many extraneous variables that are unaccounted for by RCIs such as age, education, regression to the mean, and other environmental or examinee state factors. As such, there appears to be limited utility of RCIs calculated with annual baselines when seeking to determine meaningful differences in cognitive performance.

Given the low sensitivity of the RCIs calculated in this study, it is worth considering an alternative to RCIs to evaluate meaningful change. Alternative measures of meaningful change, such as regression based measures, will likely serve as stronger indicators of clinically significant changes. These alternative measures would likely prove especially valuable in cases in which there are many confounding variables (e.g., performed extremely well at baseline one). However, such methods are not easily accessible to everyday clinicians, potentially making their use unrealistic in a busy sports medicine clinic. The use of different data to calculate more useful indicators of change could be a worthwhile consideration for clinicians and test developers moving forward.

Despite the low sensitivities of the RCIs calculated in this study, it is important to recognize that the RCIs presented in this study are based on annual baseline assessments whereas RCIs used for return to play assessments are based on 3-11 day intervals. Given the theoretical underpinnings of the increased neurodevelopment during adolescence, the RCIs obtained through annual baseline assessments are expected to have a wider range and consequent lower sensitivity. Based on underlying theory discussed previously, the increased changes in performance would result in broader estimates of measurement error through RCIs. As such, these results are consistent with the theoretical ideas underlying the changes observed in this study.

These valuable clinical and theoretical implications also present important implications for research moving forward. First, the question of whether or not stratified RCIs are necessary for different age groups in post-injury assessment remains important. Current results indicate there is little support for age-based RCIs based on annual baselines; however, the results do not indicate whether or not there is a need age-based RCIs based on recent assessments in the post-injury phase. As noted above, the low sensitivity of annual baseline RCIs across age groups limits the utility as an estimate of measurement error. However, more research is needed in order to determine the need for RCIs to be stratified by age when assessments are completed at smaller intervals. Should adolescents begin to complete baselines at the outset of each season as recommended in this study, research should focus on the RCIs associated with baseline assessments at those timepoints.

Beyond the implication for more research regarding the varying range of RCIs at different timepoints and ages, there is also a need for research regarding other measures of change that may be more appropriate, such as Minimal Important Difference. Metrics such as Minimal Important Difference may provide more accurate estimates of change that is outside of the expected range. To date, there is no known research considering reliability estimates other than the RCI for ImPACT. Other measures of meaningful change following concussion may allow providers to make more accurate decisions when helping athletes return to play safely.

General Discussion

The primary purpose of the current study was to explore the stability of baseline ImPACT scores at a one year interval across different age groups among high school football players. Existing literature regarding stability in baseline scores is mixed; though some studies indicated respectable reliability over a one year test-retest interval (Elbin et al., 2011; Brett et al., 2016), other researchers have found lower estimates of reliability in high school athletes over a two-year test-retest period when using RCIs (Tsushima et al., 2015). This study revealed significant changes in various composite scores across one year intervals in high school football players, as well as differing levels of stability between age groups. These results highlight the need for consideration of age at baseline in determining baseline policies.

In terms of stability of baseline composite scores over time, the present study found that some composite scores are stable over time, whereas other plateau at various points in adolescence. Previous research found significant differences in average Visual Memory, Visual Motor Speed, and Reaction Time composite scores at various test-retest intervals (i.e., 1 or 2 years), whereas Verbal Memory and PCSS composite scores remained stable (Elbin et al., 2011; Tsushima et al., 2015, Brett et al., 2016). Results of the present study indicate various rates of improvement between groups, as well as stability of Reaction Time in all groups. For the 14-year-old group, significant improvements in verbal memory, visual memory, and visual motor speed suggest that these areas are continuing to develop up to age 15. Based on results in the 15-year-old group, verbal memory appears to plateau by age 16, whereas visual memory and visual motor speed continue to develop up to age 16. In the 16-year-old group, visual memory appears to plateau by age 17 and visual motor speed continues developing up to age 17. Further, Reaction Time is stable in all groups. Results of the current study further clarify the stability of these composite scores at different ages to allow for more accurate interpretation of baseline scores in return to play decision making.

In addition to further clarifying changes in baseline scores among different age groups, the current study also adds important information about cognitive baseline scores that remain stable across age groups. In contrast to previous studies which indicated improvement in reaction time for adolescents, results of the present study indicate Reaction Time remains stable across athletes completing baseline one at 14-, 15-, and 16-years-old. As noted previously, the Reaction

Time composite score is the one cognitive domain measured by ImPACT that does not significantly change year-to-year for any adolescent age group in this study. Given the demonstrated stability across age groups, the Reaction Time composite can be relied on as a stable measure of cognitive ability following injury. Using measures expected to remain stable from remote baselines will help to limit the likelihood of erroneously returning athletes to play prior to return to baseline cognitive functioning.

Another goal of the current study was to identify the degree to which changes in baseline scores differed among each age group. The second hypothesis, which theorized that there would be significant differences in change scores between age groups, was not supported in any comparison. Because previous literature has not examined differences in baseline performance among different age groups, there is no prior research examining differences in change scores. As such, this hypothesis sought to fill the current gap in the literature. Results indicate that although there are some improvements occurring across age groups (i.e., Visual Memory and Visual Motor Speed), the degree of these improvements does not significantly differ between age groups. These results indicate that the year to year changes in verbal memory, visual memory, visual motor speed, and reaction time, as well as changes in error rate and symptom reporting, are not meaningfully different for high school football players from ages 14 to 17.

The third and final goal of the current study was to further examine the reliability of baseline ImPACT scores and the utility of RCIs for different age groups. Similar to the second hypothesis, these analyses sought to evaluate the potential variability in RCIs among age groups given the gap in the literature. Overall, RCIs for the three age groups in the current study are not considered meaningfully different, and RCIs did not decrease as age increased. However, in comparison to previous research, the RCIs calculated in the present study are different (i.e., no overlap) than those previously published for Verbal Memory, Reaction Time, and PCSS among 1-year intervals with high school athletes (Elbin et al., 2011; Brett et al., 2016).

When compared to previous studies, the current results indicate larger RCIs for Reaction Time and Verbal Memory composites in all age groups, as well as smaller RCIs for PCSS scores in all age groups. These discrepancies with previous research indicate that performance on Verbal Memory and Reaction Time composites was more variable in the present sample, whereas reports of baseline concussion symptom reporting was more consistent across baselines. Notably, these discrepancies with previous research are consistent across all age groups in the present study. The consistency of these discrepancies provides further evidence for the lack of variability in RCIs across age groups in the current study.

Overall, results of this study provide evidence regarding the stability of baseline performance for high school football players above and beyond those previously published. Specifically, this study examined high school football players based on age at initial baseline in order to determine potential differences and identify consequent clinical considerations. Further, this study examined a sample consisting of only high school football players. This is a group that previous research has indicated as potentially at risk in the absence of a diagnosed concussion due to sub-concussive impacts throughout the course of the season (McCrory et al., 2017). Given this potential risk for neurocognitive dysfunction, the present study adds valuable information to the current literature.

Given the current results and the information provided in prior research, it is important to consider the theoretical implications of this study as a whole. This study is the first to evaluate baseline ImPACT performance stratified by age at first baseline. The theory underlying the hypotheses of the present study was largely dependent upon Luria's theory of cognitive development. As previously noted, the theory maintains that intellectual skills are largely established by 12-years-old, though cognitive development of various skills (e.g., executive functioning) continues throughout adolescence and into adulthood based on cortical maturation. This continuous development suggests that test-retest reliability of high school athletes is confounded by the significant neuropsychological development that occurs over the course of a year in adolescents. As such, high school athletes were considered more likely to produce less reliable scores at younger ages as they are more likely to further develop cognitively.

In light of the results presented in this study, Luria's theory appears to be substantiated in the cognitive development of adolescent high school football players. Consistent with Luria's theory, the 14-year-old group demonstrated improvements in three composite scores at baseline two, whereas the 15-year-old group demonstrated improvements in two composite scores, and the 16-year-old group demonstrated improvements in only one composite score. Moreover, all groups demonstrated stable performance in the Reaction Time composite, indicating this skill is well established by age 14.

When considering the relationship between cognitive development and performance on repeat assessments, it is essential to consider the potential impact of practice effects. Practice effects can be considered a source of improvements on cognitive testing as exposure to testing can result in learned strategies and memory for answers. Although ImPACT uses multiple versions to limit practice effects, it is an important consideration when extrapolating underlying mechanisms of improvements across tests. Previous research has indicated that repeated exposure across a one-month interval resulted in improved Visual Motor Speed performance in a sample of college students, though there were no other significant improvements (Schatz & Ferris, 2013). Authors posited that improvements in Visual Motor speed were due to improvements in the physical mechanics used to complete the test. These results highlight the influence of practice effects on Visual Motor Speed as well as the lack of impact of practice effects on other composite scores. In the present study, Visual Motor Speed was the only composite score that improved in all three groups; previous research indicates these improvements may be due to practice effects rather than true cognitive development. However, it is important to note that the study demonstrating practice effects was conducted over a one-month test-retest interval, whereas the current study examined baselines completed nearly a year apart. Further, given the lack of practice effects in the other composite scores, it is likely that practice effects did not significantly influence the results of the other composite scores in this study.

The results of previous research on practice effects further establishes the likelihood that the results of the current study are due to cognitive development that occurs consistent with the Luria's theory. In addition to the lack of evidence for practice effects in prior ImPACT studies indicates that the improvements observed in this study are likely due to true improvements in cognitive skills rather than practice effects. In addition, the practice effects observed on the Visual Motor Speed composite indicate that the cognitive skills measured by ImPACT are fully developed by age 17, as the only improvement for this group was on the Visual Motor Speed composite.

Although there are differences in the observed scores that are likely consistent with ongoing cognitive development in this sample, it is important to recognize there were no significant differences in change scores or RCIs. This pattern of results indicates that the magnitude of development happening in this group is not large enough to indicate statistical significance between groups. Further, it also indicates that the estimates of measurement error do not significantly differ despite variable levels of improvement in performance across these groups. Despite the lack of statistical significance for estimates of measurement error at one year intervals, the improvement in various composite scores indicates the presence of clinically meaningful development across groups.

Given the statistically significant and clinically meaningful changes observed in this study, there are important clinical implications based on the results. Specifically, these results highlight the need for re-considered baseline policies. In addition, these results underscore the potential utility of specific cognitive measures in "expired" baselines.

In order to determine the best practice for frequency and intervals between baselines, it is important to consider the results of this study as a whole. Results of the first hypothesis indicate that although some scores remain stable, important cognitive skills such as verbal memory, visual memory, and visual motor speed can meaningfully change from year to year in healthy individuals who have not sustained a concussion. Results also indicate that as athletes develop, their performance on measures of verbal memory and visual memory become more stable from year to year. These results (e.g., variability at one year retest for ages 14 and 15), combined with the broad year to year RCIs, indicate that baseline testing should ideally occur at the outset of every athletic season for those aged 14 and 15. Given the increase in stability of scores for 16-year-olds, annual baseline testing is appropriate. The use of more frequent baselines for younger athletes aims to account for the differences in development observed between age groups within this study, as well as reduce the potential measurement error by producing narrower RCIs.

Within the context of increased baseline assessment frequency, it is also important to consider the potential influence of practice effects for athletes completing multiple baselines within a year. As noted above, research indicates that practice effects on ImPACT are limited, even at a one-month interval (Schatz & Ferris, 2013). Moreover, practice effects were most prominent for physical aspects of test approach associated with the Visual Motor Speed Composite, whereas there were no practice effects noted for other cognitive domains. Among those who would complete more frequent baselines, the Visual and Verbal Memory composites demonstrated continued cognitive development. In order to limit the influence of practice effects, it is recommended that 14- and 15-year-old athletes do not complete a baseline assessment if they are not participating in a sport in a given season.

Despite the recommendation for more frequent baseline assessments at ages 14 and 15, it is important to note the utility of the cognitive composite score that did not demonstrate significant change in any group: Reaction Time. Given the lack of significant change in Reaction Time composite scores across groups, this measure may serve as a reliable indicator of change following concussion. With this in mind, if a recent baseline (e.g., within less than 1 year) is not available, remote baseline Reaction Time composite scores may serve as a valuable clinical indicator of the status of cognitive recovery.

Beyond these data, it is also critical to consider the resources required for baseline protocols, such as financial support, personnel, and testing space. With both the present study's results and the resources required for baseline testing in mind, authors of this study propose that when resources are limited, baseline assessments are not necessary for athletes below the age of 16. Given the limited change in baseline scores after age 16, the authors propose that athletes complete baselines once every 2 years (e.g., at beginning of sophomore and senior years) if resources prevent more frequent baseline assessments. These recommendations are not meant to deter athletes and organizations from completing more frequent baselines if provided with the appropriate resources, as it is optimal to complete baseline assessments each participating season for younger athletes and annually for those aged 16 or older. Despite the value of a baseline assessment, it is essential to recognize that concussions can be treated without the use of a baseline assessment when necessary.

Limitations

Despite the strengths of the current study, there are also important limitations. These limitations include multiple factors that may confound results (e.g., undetected sandbagging, test administration settings, influence of practice effects, time between baselines) as well as factors that limit the generalizability of the current study (e.g., only high school athletes, predominantly male sample, restricted to football players, potential impact of unmeasured sub-concussive hits). Together, these factors limit the confidence with which current results can be interpreted and generalized to other populations.

Prior to interpreting any neuropsychological test data, it is essential to first consider the validity of the results. In the present study, cases were excluded if invalid performance was detected by the embedded performance validity indicators (ImPACT, 2018). Despite efforts to detect invalid performance, prior research has indicted that 30 to 35% of athletes instructed to sandbag (i.e., provide sub-optimal effort) were undetected by ImPACT's embedded performance validity indicators (Schatz & Glatts, 2013; Higgins, Denny, & Maerlender, 2017). Given the potential for undetected sandbagging, it is possible that the results of the current study may be underestimates of performance in some cases. However, it is important to note that no prior studies have evaluated the potential ability for athletes to sandbag more than one baseline assessment undetected. In this study, athletes were only included if they completed two valid baseline assessments, which may have further reduced the likelihood that sub-optimal performance went undetected.

Beyond the consideration of the measured validity indicators, there are other environmental factors that have been shown to influence performance on baseline ImPACT assessments. One key environmental testing factor that has been well-studied is the administration setting (i.e., individually or in a group). In the present study, the uniformity of test administration is unknown. Athletes completed the assessments through their respective athletic programs, and it is unclear if athletes completed baselines individually or in groups. In addition, the consistency of administration format between baselines is also unknown. This is an important limitation as previous research has found that athletes often perform worse in groups (Moser, Schatz, Neidzwski & Ott, 2011), and the effects are more pronounced in younger athletes (Moser, Schatz & Lichtenstein, 2015). Based on existing literature, the lack of information regarding baseline administration format poses a limitation to the interpretation of current results.

In addition to measured validity indicators and administration format, other aspects of the athlete's state at the time of testing can also influence performance on baseline assessments. Factors such as sleep, exercise, and mood can have negative influences on baseline performance. A study completed by McClure et al. (2013) found that the baseline ImPACT scores were significantly lower in reaction time, verbal memory, and visual memory for high school and college students reporting fewer than 7 hours of sleep prior to baseline assessment. In addition, Pawlukiewicz, Yengo-Kahn, and Solomon (2017) found that reported strenuous exercise prior to assessment also resulted in significantly lower scores. Lastly, Covassin et al. (2012) found that symptoms of severe depression at baseline resulted in lower visual memory scores. The present study did not account for sleep, exercise, or mood functioning at baseline assessments; as such, the extent to which these factors influence performance on baseline assessments in this study is unclear.

In the event that baseline performance represents an underestimate of true cognitive ability due to undetected sandbagging, group administration formats, and/or other aforementioned variables, there are multiple outcomes regarding interpretation of current data. If performance on both baselines in the current study represents underestimates of cognitive abilities, the magnitude of change across baseline assessments may be smaller than the true amount of change. Conversely, if an athlete were to sandbag baseline one and provide best effort on baseline two, the amount of change observed would represent an overestimate of true change. Similarly, varying the administration settings between individual and group environments across baselines could alter observed scores in a parallel fashion. Notably, these variations between optimal and suboptimal performance could also result in larger RCIs among groups given the variability in change between athletes. Given the potential influence of these factors on observed changes in baseline scores, it is important to be cognizant of these confounds when interpreting results of the present study.

In addition to the influence of potentially undetected sandbagging and administration setting, the potential of practice effects in repeated neuropsychological assessment also represents a limitation in this study. Across all neuropsychological assessments, practice effects should be considered as a potential source of improvements on cognitive testing via exposure to testing and learned strategies and memory for answers. Importantly, prior research has found that practice effects evaluated across various neuropsychological measures are generally attenuated through the use of alternate forms and increased intervals between assessments (Calamia, Markon, & Tranel, 2012). Further, previous research has found that practice effects on ImPACT are limited even at one-month test-retest intervals in a sample of college students (Schatz & Ferris, 2013). In the present study, baselines were completed approximately one year apart for all subjects, and alternate forms were used at baseline two. It is also important to note that subjects in the current study did not have documented exposure to ImPACT prior to baseline one. Given the limited evidence for significant practice effects at one month, it is considered unlikely that practice effects account for all changes in observed scores across this study. Further, the RCIs calculated in this study used a formula presented by Iverson et al. (2003) to correct for practice effects when they are present. Nonetheless, it is important to consider the potential impact of practice effects when interpreting present results.

Another important limitation of the present study is the lack of standardized time between baseline assessments. Though all subjects completed baseline two approximately a year after baseline one, the time between assessments averaged about 338 days with a standard deviation of 68 days across all subjects. The lack of standard timeframes for baseline follow-up across subjects limits the extent to which we can interpret the average changes in scores. In addition, the variation in time between assessments likely reduced the test-retest reliability coefficients, which subsequently increases the range of potential measurement error estimated through RCIs.

Beyond the confounding factors that limit confidence in the interpretation of results, there are also multiple study factors that limit the generalizability of current findings. Demographic factors such as the age, gender, race, and sport limit the extent to which present results can be extended to other populations.

Because this study aimed to evaluate potential differences in long-term stability in ImPACT scores across high school athletes, these results should not be generalized beyond high school athletes aged 14 to 17. Importantly, though these results indicate variable improvements in cognitive functioning as adolescents develop, these results should not be extended to those below age 14. Though development is occurring below age 14, the present study's sample does not allow for interpretation of potential change below 14-years-old.

In addition to the age of the current sample, the gender of the sample also limits the generalizability of results. As is expected when examining data from high school football players, the sample was overwhelmingly male. Because the sample largely consisted of males, the results should not be extended to female athletes. Given the differences in male and female development timetables, the results observed in this study may not be applicable to female athletes.

Further, the racial and ethnic backgrounds of the present sample limit the ability to generalize to athletes of other backgrounds. Previous research has shown differences in baseline reaction time and visual motor speed among Black and White high school athletes (Wallace et al., 2018) as well as among athletes with various countries of origin (Burley et al., 2018). The present study did not examine differences among different racial and ethnic identities, thus limiting the ability to interpret differences due to race and/or ethnicity.

Beyond limitations of age, gender, and race, the present study examined only football players. As a result, these results have limited generalizability to adolescent athletes who do not play football. Despite this limitation, it is important to note that high school athletes often play more than one sport throughout the year. The data obtained for this study did not clarify multisport participation, thus limiting the ability to examine subsequent effects of participation in sports other than football. Because athletes are often involved in more than one sport, these results should be considered for athletes who participate football as one of their sports, though football may not be their only sport.

The primary reason the current study focused exclusively on high school football was to determine the influence of potential sub-concussive injuries on the long-term stability of ImPACT scores. Though this approach was a strength of the study, the exclusive examination of a population that may demonstrate effects of repeated brain injury confines the generalizability to other football players. Of note, the current study did not quantify the estimated number of sub-concussive impacts, further limiting the ability to draw conclusions about football participation and cognitive change. Further, the study lacked a control group, which significantly limits the ability to determine potential differences between football players and other groups not exposed to repeated sub-concussive injury.

In addition to the aforementioned factors that limit the extent the results can be generalized to other populations, the study's scope was also limited relative to previous research. To elaborate, prior research regarding long-term stability of ImPACT performance evaluated change with regression based methods (RBMs) in addition to RCIs (e.g., Elbin et al., 2011; Brett et al., 2016; Tsushima et al., 2015). Authors have consistently found that RBMs provide more sensitive estimates of meaningful change, whereas the estimates of change provided by RCIs have been found to produce wide ranging estimates with limited sensitivity over long test-retest intervals (i.e., 1 year). Because ImPACT relies on RCIs to estimate change, the current study examined RCIs for potential discrepancies between age groups. Nonetheless, the study's clinical, theoretical, and research implications are limited by the absence of estimates of reliable change based on RBMs.

Future Research

Future research should seek to explore the potential impacts of current clinical and psychometric implications, as well as corroborate results of this study and further explore the potential effects of the limitations described above. The primary focus of future research should investigate the impact of clinical recommendations resulting from this study (e.g., effect of increased baseline frequency) as well as the psychometric implications (e.g., clinical utility of RCIs). In addition, the absence of existing literature examining the variability of baseline ImPACT performances between age groups indicates a need to reproduce current results. Further, research should aim to increase confidence of interpretation as well as generalizability by further examining the limitations established above.

The first area of future research should examine the potential effects of the proposed changes in baseline policies for adolescent athletes. Given the variability of observed improvements in different age ranges as well as the remarkably wide estimates of reliable change at one year intervals, increased frequency of baselines was recommended for younger athletes. With increased baseline frequency, there is potential that improvements due to practice effects could inadvertently reduce the ability of ImPACT to detect change following concussion, thus decreasing athlete safety. It is essential that future research investigate the effects of increased baseline frequencies on baseline assessment performance. Of note, in order to reduce the likelihood of practice effects, research should use alternative forms at each assessment. Studies should also evaluate outcomes at various test-retest intervals to determine reliability and practice effects for athletes who are completing baselines at different intervals (i.e., 3-sport athletes every 4 months, 2-sport athletes every 6 months, 1-sport athletes once a year). Such research is essential in determining the psychometric and clinical implications of conducting more frequent baselines for adolescent athletes.

Similar to the need for more research regarding potential effects of baseline utility when increasing baseline frequency, it is critical to establish new estimates of predicted change within the specified time intervals. It is likely that these estimates of meaningful change will be smaller due to the presumed decrease in difference scores at shorter intervals.

In addition, future research should also focus on the potential utility of alternative psychometric tools to determine meaningful change. The present study demonstrates that RCIs are uniformly wide ranging for multiple adolescent age ranges at one-year intervals. Given the large range provided when using RCIs, researchers have indicated these measures demonstrate low sensitivity in the detection of meaningful change (Temkin et al., 1999; Barr, 2002). Consequently, it is important determine the utility of RCIs in establishing meaningful departure from baseline assessment. With this in mind, future research should determine the usefulness of RBMs to estimate meaningful change at follow-up assessments with adolescents in different age groups. At present, previous research has only studied RBMs in high school athletes as a whole. It is possible that RBMs accounting for age may produce different ranges of meaningful change across age groups.

Beyond exploring the clinical and psychometric implications of the current study, further research should aim to corroborate present results. In order to increase confidence in the interpretation of current results as well as the generalizability, it is important to reproduce results across different samples. When conducting future studies it will be critical to account for the limitations of the present research in order to rule out potential alternative explanations for observed improvements across age groups with wide ranging RCIs.

First, research should focus on limiting the likelihood that athletes are able to remain undetected when sandbagging baseline performance. Such research should consider implementing adjusted score thresholds for validity indicators beyond those specified by ImPACT, as well as additional standalone performance validity tests. Previous research by

Manderino and Gunstad (2017) has demonstrated increased sensitivity and maintained specificity with the use of adjusted score thresholds for current ImPACT validity indices as well as novel ImPACT validity indices in a college sample. By implementing such measures, future research can increase sensitivity to poor effort to maximize confidence in results.

In addition to increasing the detection of suboptimal effort, research should examine the ability of athletes to repeatedly sandbag baseline assessments. Though there is existent literature regarding the ability to remain undetected when sandbagging a single baseline, there is limited evidence that athletes would be able to successfully sandbag multiple assessments without detection.

Another factor that may result in poorer performance on baseline assessment is administration in large groups (i.e., >5 athletes). The present study did not track administration across subjects, which can have an important influence on long-term stability of baseline scores. As such, future research should make a point to track the administration setting for cases included in research samples.

Other important variables that can impact performance on baseline assessments include sleep, exercise, mood, and medications. The present study did not control for these variables in order to maintain a sample representative of athletes commonly presenting for baseline assessments. However, future research should incorporate these variables in order to help determine the impact of these variables on current results.

Beyond examinee traits and administration setting, it is also important to consider the lack of standardized time between baselines in this study. Baselines were completed about one year apart for all subjects; however, the variation within time between baselines limits the confidence with which results can be interpreted. As a result, future studies should implement more standardized intervals when possible.

Increasing the generalizability of present results is also an important aim for future research. Due to the limited variability in age, gender, and sport played in the current study, it remains uncertain whether these results extend to younger athletes, females, other racial identities, and other sports. Future studies can easily extend the generalizability of this research by using a more heterogenous sample.

Despite the pitfalls of a sample composed exclusively of football players, this study provides important information about the continuation of cognitive development in high school football players. This research is especially important given the potential accumulation of sub-concussive hits sustained by high school football players throughout a season. However, the study did not obtain estimates of sub-concussive impacts, which significantly limits the confidence with which such conclusions can be drawn. In light of this weakness, future studies should attempt to measure these impacts, whether is through self-report or through quantitative data from new technology used to quantify the number and severity of impacts. Such research will provide meaningful data regarding the impact of repetitive sub-concussive hits in this population.

It is also important to note the lack of a control group in the current sample. Without a control group, the influence of suspected sub-concussive injuries on cognitive development cannot be determined. As a result, the use of a randomly selected and demographically matched control group of non-football athletes is recommended in future studies.

Consistent with recommendations for further research on sub-concussive impacts and their role in cognitive development, it is also recommended that future research investigate the role of an athlete's age of first exposure to football. Current literature suggests that participation in football prior to age 12 increases the odds of reported neuropsychiatric symptoms and executive dysfunction later in life (Alosco et al., 2017). Future research should focus on not only elucidating the impact of age of first exposure on cognitive development, but also on long-term outcomes following participation in tackle football prior to age 12. Current data suggest correlations with CTE diagnoses and younger age of first exposure, though there have been no longitudinal cross sectional studies published to date. Such studies would provide important information about age of first exposure beyond the correlations presently available.

In sum, future research should provide follow-up regarding updated psychometric information for ImPACT's test-retest stability at shorter intervals given the clinical implications and recommendations of this study. Further, research should aim to corroborate results and improve upon the limitations present in this study in order to increase generalizability of results. Lastly, further research should evaluate the potential influence of repeated sub-concussive impacts in high school football players as well as the impact of age of first exposure to tackle football on neurocognitive development.

References

Allen, B., & Gfeller, J. (2011). The Immediate Post-Concussion Assessment and Cognitive Testing battery and traditional neuropsychological measures: a construct and concurrent validity study. *Brain Injury*, 25(2), 179-191. https://doi.org/10.3109/02699052.2010.541897

Alosco, M., Kasimis, A., Stamm, J., Chua, A., Baugh, C., Daneshvar, D., Robbins, C., Mariani, M., Hayden, J., Conneely, S., Au, R., Torres, A., McClean, M., McKee, A., Cantu, R., Mez, J., Nowinski, C., Martin, B., Chaisson, C., Tripodis, Y., & Stern, R. A. (2017). Age of first exposure to American football and long-term neuropsychiatric and cognitive outcomes. *Translational Psychiatry*, 7(9) https://doi.org/10.1038/tp.2017.197

Alsalaheen, B. A., Haines, J., Yorke, A., Stockdale, K., & P. Broglio, S. (2015). Reliability and concurrent validity of instrumented balance error scoring system using a portable force plate system. *The Physician and Sports Medicine*, 43(3), 221-226. https://doi.org/10.1080/00913847.2015.1040717

Barr, W. (2002). Neuropsychological testing for assessment of treatment effects: Methodologic issues. *CNS Spectrums*, 7, 300-304. https://doi.org/10.1017/S1092852900017715

Brett, B., Smyk, N., Solomon, G., Baughman, B., & Schatz, P. (2016). Long-term stability and reliability of baseline cognitive assessments in high School athletes using ImPACT at 1-, 2-, and 3-year test-retest intervals. *Archives of Clinical Neuropsychology,* 31(8), 904-914. https://doi.org/10.1093/arclin/acw055

Broglio, S., Ferrara, M., Macciocchi, S., Baumgartner, T., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training*, 42(4), 509-514.

Brown, N., Mannix, R., O'Brien, M., Gostine, D., Collins, M., & Meehan, W. (2014). Effect of cognitive activity level on duration of post-concussion symptoms. *Pediatrics*, 133(2), e299-304. https://doi.org/10.1542/peds.2013-2125

Bruce, J., Echemendia, R., Meeuwisse, W., Comper, P., & Sisco, A. (2014). 1 year test-retest reliability of ImPACT in professional ice hockey players. *The Clinical Neuropsychologist*, 28(1), 14-25. https://doi.org/10.1080/13854046.2013.866272

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, 26(4), 543-570. https://doi.org/10.1080/13854046.2012.680913

Chelune, G., Naugle, R., Lüders, H., Sedlak, J., & Awad, I. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7(1), 41.

Cole, W., Arrieux, J., Schwab, K., Ivins, B., Qashu, F., & Lewis, S. (2013). Test-retest reliability of four computerized neurocognitive assessment tools in an active duty military population. *Archives of Clinical Neuropsychology*, 28(7), 732-742. https://doi.org/10.1093/arclin/act040

Corrigan, J., Selassie, A., & Orman, J. (2010). The epidemiology of traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 25(2), 72-80. https://doi.org/10.1097/HTR.0b013e3181ccc8b4

Covassin, T., Elbin, R., Larson, E., & Kontos, A. (2012). Sex and age differences in depression and baseline sport-related concussion neurocognitive performance and symptoms. *Clinical Journal of Sports Medicine*, 22(2), 98-104. https://doi.org/10.1097/JSM.0b013e31823403d2

Covassin, T., Elbin, R., Stiller-Ostrowski, J., & Kontos, A. (2009). Immediate post-concussion assessment and cognitive testing (ImPACT) practices of sports medicine professionals. *Journal of Athletic Training*, 44(6), 639-644. https://doi.org/10.4085/1062-6050-44.6.639

Echemendia, R., Iverson, G., McCrea, M., Macciocchi, S., Gioia, G., Putukian, M., & Comper, P. (2013). Advances in neuropsychological assessment of sport-related concussion. *British Journal of Sports Medicine*, 47, 294-298. https://doi.org/10.1136/bjsports-2013-092186

Echemendia, R., Iverson, G., McCrea, M., Broshek, D., Gioia, G., Sautter, S., Macciocchi, S., & Barr, W. (2011). Role of neuropsychologists in the evaluation and management of sport-related concussion: An inter-organization position statement. *Archives of Clinical Neuropsychology*, 27(1), 119-122. https://doi.org/10.1093/arclin/acr077

Elbin, R., Schatz, P., & Covassin, T. (2011). One-year test-retest reliability of the online version of ImPACT in high school athletes. *The American Journal of Sports Medicine*, 39(11), 2319-2324. https://doi.org/10.1177/0363546511417173

Elbin, R., Sufrinko, A., Schatz, P., French, J., Henry, L., Burkhart, S., Collins, M., & Kontos, A. (2016). Removal from play after concussion and recovery time. *Pediatrics,* 138(3). https://doi.org/10.1542/peds.2016-0910

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

Gessel, L., Fields, S., Collins, C., Dick, R., & Comstock, R. (2007). Concussions among United States high school and collegiate athletes. *Journal of Athletic Training*, 42(4), 495-503.

Giza, C., Griesbach, G., & Hovda, D. (2005). Experience-dependent behavioral plasticity is disturbed following traumatic injury to the immature brain. *Behavioral Brain Research*, 157(1), 11-22. https://doi.org/10.1016/j.bbr.2004.06.003

Giza, C., & Hovda, D. (2001). The neurometabolic cascade of concussion. *Journal of Athletic Training*, 36(3), 228-235.

Giza, C., & Hovda, D. (2014). The new neurometabolic cascade of concussion. *Neurosurgery*, 75(4), S24-33. https://doi.org/10.1227/neu.0000000000000505

George, D. and Mallery, M. (2010) SPSS for windows step by step: A simple guide and reference, 17.0 Update, 10th Edition, Pearson, Boston.

Griesbach, G., Hovda, D., Molteni, R., Wu, A., & Gomez-Pinilla, F. (2004). Voluntary exercise following traumatic brain injury: brain-derived neurotrophic factor upregulation and recovery of function. *Neuroscience*, 125(1), 129-139. https://doi.org/10.1016/j.neuroscience.2004.01.030

Gysland, S., Mihalik, J., Register-Mihalik, J., Trulock, S., Shields, E., & Guskiewicz, K. (2012). The relationship between sub-concussive impacts and concussion history on clinical measures of neurologic function in collegiate football players. *Annals of Biomedical Engineering*, 40, 14-22. https://doi.org/10.1007/s10439-011-0421-3

Henry, L., Elbin, R., Collins, M., Marchetti, G., & Kontos, A. (2016). Examining Recovery trajectories after sport-related concussion with a multimodal clinical assessment approach. *Neurosurgery*, 78(2), 232-241. https://doi.org/10.1227/neu.0000000000001041

Higgins, K., Denney, R., & Maerlender, A. (2016). Sandbagging on the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) in a high school athlete population. *Archives of Clinical Neuropsychology*, 32(3), 259-266. https://doi.org/10.1093/arclin/acw108

ImPACT Applications (2018). ImPACT Administration and interpretation Manual.

Iverson, G. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology*, 16(2), 183-191.

Iverson, G., Gardner, A., Shultz, S. , Solomon, G., McCrory, P., Zafonte, R., Perry, G., Hazrati, L., Keene, C., & Castellani, R. (2019). Chronic traumatic encephalopathy neuropathology might not be inexorably progressive or unique to repetitive neurotrauma. *Brain: A Journal of Neurology*, 142(12), 3672-3693. https://doi.org/10.1093/brain/awz286

Iverson, G., & Green, P. (2001). Measuring improvement or decline on the WAIS–R in inpatient psychiatry. *Psychological Reports,* 89(2), 457-462.

Iverson, G., Lovell, M., & Collins, M. (2003). Interpreting change on ImPACT following sport concussion. *The Clinical Neuropsychologist*, 17(4), 460-467. https://doi.org/10.1076/clin.17.4.460.27934

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

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19. https://doi.org/10.1037/0022-006X.59.1.12

Kontos, A., & Collins, M. (2018). Concussion: A clinical profile approach to assessment and treatment. *American Psychological Association*. https://books.google.com/books?id=IFKCtgEACAAJ

Langlois, J., Rutland-Brown, W., & Wald, M. (2006). The epidemiology and impact of traumatic brain injury: a brief overview. *Journal of Head Trauma Rehabilitation*, 21(5), 375-378. https://doi.org/10.1097/00001199-200609000-00001

Lichtenstein, J., Moser, R., & Schatz, P. (2014). Age and test setting affect the prevalence of invalid baseline scores on neurocognitive tests. *American Journal of Sports Medicine*, 42(2), 479-484. https://doi.org/10.1177/0363546513509225

Luria, A. (1980). Higher cortical functions in man. *Springer*. https://doi.org/https://doi.org/10.1007/978-1-4615-8579-4

Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., & McAllister, T. (2010). Examination of the construct validity of ImPACT™ computerized test, traditional, and experimental neuropsychological measures. *The Clinical Neuropsychologist,* 24(8), 1309-1325. https://doi.org/10.1080/13854046.2010.516072

Majerske, C., Mihalik, J., Ren, D., Collins, M., Reddy, C., Lovell, M., & Wagner, A. (2008). Concussion in sports: Postconcussive activity levels, symptoms, and neurocognitive performance. *Journal of Athletic Training*, 43, 265-274. https://doi.org/10.4085/1062-6050-43.3.265

Manderino, L., Zachman, A., & Gunstad, J. (2018). Novel ImPACT validity indices in collegiate student-athletes with and without histories of ADHD or academic difficulties. *The Clinical Neuropsychologist*, 33, 1-12. https://doi.org/10.1080/13854046.2018.1539191

McClure, D. J., Zuckerman, S., Kutscher, S., Gregory, A., & Solomon, G. (2014). Baseline neurocognitive testing in sports-related concussions: the importance of a prior night's sleep. *American Journal of Sports Medicine*, 42(2), 472-478. https://doi.org/10.1177/0363546513510389

McCrea, M., Iverson, G., McAllister, T., Hammeke, T., Powell, M., Barr, W., & Kelly, J. (2009). An integrated review of recovery after mild traumatic brain injury (MTBI): implications for clinical management. *The Clinical Neuropsychologist*, 23(8), 1368-1390. https://doi.org/10.1080/13854040903074652

McCrory, P., Meeuwisse, W., Dvořák, J., Echemendia, R., Engebretsen, L., Feddermann-Demont, N., McCrea, M., Makdissi, M., Patricios, J., Schneider, K., & Sills, A. (2017). 5th International conference on concussion in sport (Berlin). *British Journal of Sports Medicine*, 51(11), 837. https://doi.org/10.1136/bjsports-2017-097878

McKee, A., Cantu, R., Nowinski, C., Hedley-Whyte, E., Gavett, B., Budson, A., Santini, V., Lee, H., Kubilus, C., & Stern, R. (2009). Chronic traumatic encephalopathy in athletes: Progressive tauopathy after repetitive head injury. Journal of *Neuropathology &*

Experimental Neurology, 68(7), 709-735. https://doi.org/10.1097/NEN.0b013e3181a9d503

McLendon, L., Kralik, S., Grayson, P., & Golomb, M. (2016). The controversial second impact syndrome: A review of the literature. *Pediatric Neurology*, 62, 9-17. https://doi.org/10.1016/j.pediatrneurol.2016.03.009

Menon, D., Schwab, K., Wright, D., & Maas, A. (2010). Position statement: definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 91(11), 1637-1640. https://doi.org/10.1016/j.apmr.2010.05.017

Miller, G., DePadilla, L., & Xu, L. (2021). Costs of nonfatal traumatic brain injury in the United States, 2016. *Medical Care*, 59(5), 451-455. https://doi.org/10.1097/mlr.0000000000001511

Moser, R., Schatz, P., & Lichtenstein, J. (2015). The importance of proper administration and interpretation of neuropsychological baseline and postconcussion computerized testing. *Applied Neuropsychology: Child*, 4(1), 41-48. https://doi.org/10.1080/21622965.2013.791825

Moser, R., Schatz, P., Neidzwski, K., & Ott, S. (2011). Group versus individual administration affects baseline neurocognitive test performance. *American Journal of Sports Medicine*, 39(11), 2325-2330. https://doi.org/10.1177/0363546511417114

Obrzut, J., & Hynd, G. (1986). Child neuropsychology, Vol. 2: Clinical practice. Academic Press.

Omalu, B., DeKosky, S., Minster, R., Kamboh, M., Hamilton, R., & Wecht, C. (2005). Chronic traumatic encephalopathy in a National Football League player. *Neurosurgery*, 57(1), 128-134; discussion 128-134. https://doi.org/10.1227/01.neu.0000163407.92769.ed

Pawlukiewicz, A., Yengo-Kahn, A. M., & Solomon, G. (2017). The Effect of Pretest Exercise on Baseline Computerized Neurocognitive Test Scores. *Orthopaedic Journal of Sports Medicine*, 5(10). https://doi.org/10.1177/2325967117734496

Pokhrel, P., Herzog, T., Black, D., Zaman, A., Riggs, N., & Sussman, S. (2013). Adolescent neurocognitive development, self-regulation, and school-based drug use prevention. *Prevention Science: The Official Journal of the Society for Prevention Research*, 14. https://doi.org/10.1007/s11121-012-0345-7

Resch, J., Driscoll, A., McCaffrey, N., Brown, C., Ferrara, M., Macciocchi, S., Baumgartner, T., & Walpert, K. (2013). ImPact test-retest reliability: reliably unreliable? *Journal of Athletic Training,* 48(4), 506-511. https://doi.org/10.4085/1062-6050-48.3.09

Saatman, K., Duhaime, A., Bullock, R., Maas, A., Valadka, A., & Manley, G. (2008). Classification of traumatic brain injury for targeted therapies. *Journal of Neurotrauma*, 25(7), 719-738. https://doi.org/10.1089/neu.2008.0586

Saffary, R., Chin, L., & Cantu, R. (2011). Sports medicine: Concussions in sports. *American Journal of Lifestyle Medicine*, 6(2), 133-140. https://doi.org/10.1177/1559827611411649

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *The American Journal of Sports Medicine*, 38(1), 47-53. https://doi.org/10.1177/0363546509343805

Schatz, P., & Ferris, C. (2013). One-month test-retest reliability of the ImPACT test battery. *Archives of Clinical Neuropsychology*, 28(5), 499-504. https://doi.org/10.1093/arclin/act034

Schatz, P., & Glatts, C. (2013). "Sandbagging" baseline test performance on ImPACT, without detection, is more difficult than it appears. *Archives of Clinical Neuropsychology*, 28(3), 236-244. https://doi.org/10.1093/arclin/act009

Schatz, P., & Robertshaw, S. (2014). Comparing post-concussive neurocognitive test data to normative data presents risks for under-classifying "above average" athletes. *Archives of Clinical Neuropsychology*, 29(7), 625-632. https://doi.org/10.1093/arclin/acu041

Schatz, P., & Sandel, N. (2013). Sensitivity and specificity of the online version of ImPACT in high S]school and collegiate athletes. *The American Journal of Sports Medicine*, 41(2), 321-326. https://doi.org/10.1177/0363546512466038

Spear, L. (2000). Neurobehavioral changes in adolescence. *Current Directions in Psychological Science*, 9(4), 111-114. https://doi.org/10.1111/1467-8721.00072

Temkin, N., Heaton, R., Grant, I., & Dikmen, S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychology Society*, 5(4), 357-369. https://doi.org/10.1017/s1355617799544068

Theye, F., & Mueller, K. (2004). "Heads up": concussions in high school sports. *Clinical Medicine Research*, 2(3), 165-171. https://doi.org/10.3121/cmr.2.3.165

Thurman, D., Alverson, C., Dunn, K., Guerrero, J., & Sniezek, J. (1999). Traumatic brain injury in the United States: A public health perspective. *The Journal of Head Trauma Rehabilitation*, 14(6), 602-615.

Tsushima, W., Siu, A., Pearce, A., Zhang, G., & Oshiro, R. (2015). Two-year test–retest reliability of ImPACT in high school athletes. *Archives of Clinical Neuropsychology*, 31(1), 105-111. https://doi.org/10.1093/arclin/acv066

Zillmer, E. (2003). Sports medicine and neuropsychology: The neuropsychologist's role in the assessment and management of sports-related concussions. *Applied Neuropsychology* (1st ed.). Psychology Press. doi.org/10.4324/9781315782270