A PSYCHOMETRIC EVALUATION OF THE IMMEDIATE POST-CONCUSSION ASSESSMENT AND COGNITIVE TEST (ImPACT) FOR SPORT CONCUSSION

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Clinical Psychology

Massey University, Albany

New Zealand

Shannon Leigh Martin

2014

Abstract

The increasing awareness of concussion in sport and its effect on cognitive functioning has prompted the development of neuropsychological assessments specific to sport concussion. ImPACT is one of the more popular assessment batteries that purports to measure five areas of cognitive functioning, despite a scarcity of empirical support. The current study assessed ImPACT's factor structure to determine whether its items are accurately measuring the five cognitive domains it claims to measure. Three exploratory factor analyses using a male adolescent sample were computed before the final model, consisting of eight items and two factors, representing Reaction Time and Memory, was reached. The structure was inconsistent with the current ImPACT scoring structure. This model was then successfully validated among a new sample, while a competing model found in the literature was not successfully validated. This model was then assessed for its longitudinal stability over a three year period in addition to its cross-country validity between South African and New Zealand samples. The former was supported, indicating individuals' memory and reaction time as measured by ImPACT, is relatively stable over time and that ImPACT is not subject to practice effects after a one-year interval. It is of note that cross-country invariance was not supported, therefore emphasising the importance of having population-specific norms. Overall, the present study found that ImPACT, at this stage, has several limitations. It is recommended that, while ImPACT has the potential to be a useful tool, modifications need to be made to increase its efficacy.

Acknowledgements

Table of Contents

List of Tables

List of Figures

Introduction

Participation in sports is an important part of many cultures. Adolescents make up the majority of the sport-playing population, with 92% of New Zealand youth involved in club, school, and recreational sports and physical activity (Sport New Zealand, 2001). While there are many benefits that arise from sports participation there is also the risk of injury. The most common sporting injuries include sprains and strains (46.7%), followed by haematomas (23.9%), lacerations (8.8), fractures (5.7%), concussions (also known as mild traumatic brain injury; 4.5%), and dislocations (3.7%) (Bird, Waller, Marshall, Alsop, Chalmers & Gerrard, 1998). Concussion differs from other sporting injuries in that its symptoms are often subtle and ambiguous, making it difficult to identify. However it can be one of the most detrimental injuries, particularly if an athlete returns to play before fully recovering. Therefore it is important to have appropriate methods to assess sport concussion and track an individual's recovery. Assessment of concussion can involve a range of methods, including self-report of symptoms, postural stability testing, and/or neuropsychological assessment. Neuropsychological assessment has been deemed the cornerstone of concussion management by the international concussion in sport consensus group (McCrory et al., 2009). However there is controversy about the utility and application of cognitive testing, particularly the modern computerised neuropsychological batteries used to assess concussion. Several studies have concluded there is not enough empirical evidence to support their current widespread use throughout the United States and other developed countries (Mayers & Redick, 2012a, 2012b).

The present study sought to contribute to the current debate by evaluating the psychometric properties of a commonly-used sport concussion neuropsychological battery, the Immediate Post-concussion Assessment and Cognitive Test (ImPACT; ImPACTonline, 2013). ImPACT was developed in the United States and consists of six test modules which claim to assess five cognitive domains; Verbal Memory, Visual Memory, Reaction Time, Visual Processing Speed, and Impulse Control. While the developers claim that ImPACT is the most well-validated and empirically researched computerised concussion management tool, review of the literature indicates that investigation of its psychometric properties is far from extensive. This is an important observation as the utility of

neuropsychological assessment, such as ImPACT, relies upon demonstrable evidence that the measure accurately and reliably detects changes in cognitive function. In this way, clinicians can interpret measures and provide subsequent recommendations with confidence.

The following chapter describes the current literature pertaining to the definition, symptoms, and potential long-term consequences of concussion. The second chapter reviews the literature on the various methods of concussion assessment, both traditional and contemporary methods. This is followed by the third chapter, which presents a description of the Cattell-Horn-Carroll theory of cognitive functions and how it can be used to inform the current study. This chapter also outlines the cognitive areas believed to be affected by concussion and reviews the associated literature. The fourth chapter describes empirical research pertaining to the psychometric properties of the ImPACT battery, in addition to explaining the concepts of reliability and validity within a context of evaluating a measure. The rationale and objectives for the current study are outlined in Chapter 5 within the context of the reviewed literature presented in previous chapters. The methods of both the collection and analysis of data are described in Chapter 6, followed by the results of the analysis in Chapter 7. Lastly, the results are discussed in Chapter 8, alongside the study's limitations, implications and conclusions.

Chapter 1: Concussion in Sport

In recent years diagnosis, treatment, and management of sports concussion has gained widespread attention in the area of neuropsychology and sports medicine (Aubry et al., 2002; Green & Jordan, 1998; White et al., 2013). The strong growth of interest was prompted by media coverage of high profile athletes sustaining a concussion that resulted in either death or forced retirement (Lovell & Burke, 2000; Lovell & Pardini, 2010; Lovell, 1999). This highlighted the frequency of concussions among athletes at all levels. Earlier studies estimated the annual prevalence of sport concussion in the US to be 300,000 athletes (Thurman, Branche, & Sniezek, 1998). More recent estimates however, suggest 1.6 to 3.8 million sports concussions occur each year among US athletes (Langlois, Rutland-Brown, & Wald, 2006). The large variation in estimates reflects several developments. Firstly, previous estimates only included concussions that involved loss of consciousness (LOC; Langlois et al., 2006; Thurman et al., 1998). However, it is now known that LOC is not a necessary condition of concussion (Hinton-Bayre, Geffen, & Friis, 2004; McCrory et al., 2009). Additionally, the rise in estimate reflects current consensus that many athletes fail to report concussive symptoms due to fear of being sidelined, absence of concussion management guidelines, or simply that symptoms are not recognised due to either lack of knowledge or as a result of their often rapid onset and spontaneous recovery (McCrea, Hammeke, Olsen, Leo, & Guskiewicz, 2004; Puga, 2011). As awareness of concussion increases there is likely to be a parallel increase in the frequency of reported concussions.

Sports concussion incidence is highest among high school athletes, with it being the second leading cause of concussion following motor vehicle accidents among 15 to 24 year olds in the US (Marar, McIlvain, Fields, & Comstock, 2012). Gessel, Fields, Collins, Dick, and Comstock (2007) reported that sport-concussion accounted for 8.9% of all high school athletic injuries, with its highest prevalence in the US reported in football, followed by lacrosse and soccer. US epidemiology studies acknowledge that concussion incidence among rugby players is high, but empirical data is lacking (Halstead & Walter, 2010; Marshall & Spencer, 2001). Notwithstanding, a US study found that, among high school club rugby players, 16.1% of males and 14.3% of female rugby players sustained a concussion

(Collins, Micheli, Yard, & Comstock, 2008). This finding is comparable to an Australian study that found 13% – 17% of male athletes sustained one concussion per season (Hinton-Bayre et al., 2004).

1.1 Concussion Defined

Despite decades of research and numerous publications on concussion in sport, a universally accepted definition has yet to be developed (Lovell & Pardini, 2010). Sports concussion was first defined in 1966 by the Committee on Head Injury Nomenclature of Neurological Surgeons (Congress of Neurosurgeons, 1968). They defined concussion as "a clinical syndrome characterized by the immediate and transient post-traumatic impairment of neural function such as alteration of consciousness, disturbance of vision or equilibrium, etc., due to brain stem dysfunction" (Congress of Neurosurgeons, 1968, p. 392). This definition was limited as it did not account for concussions that resulted in persistent impairments, nor did it reflect the complexity of the disorder. As new research emerged, such as that indicating that neither brain stem dysfunction nor LOC were universal in concussion presentation, modifications in the definition were observed. For instance, in 1997 the American Academy of Neurology (AAN) defined concussion as "a trauma induced alteration in mental status that may or may not include a loss of consciousness" (American Academy of Neurology, 1997, p. 582).

Although these definitions were initially endorsed by influential parties, such as the American Medical Association, their ambiguity and simplicity soon became apparent. There was an obvious need for a more comprehensive understanding of sports concussion. An international sports concussion panel, consisting of the International Ice Hockey Federation, the Federation Internationale de Football Association Medical Assessment and Research Centre, and the International Olympic Committee Medical Commission, responded to this need by organising four international sports conferences to date (Aubry et al., 2002; McCrory et al., 2005, 2009, 2013). The first conference was in Vienna (Aubrey et al., 2002) in 2001, the next in Prague in 2004 (McCrory et al., 2005), and the third and fourth conferences were held in Zurich, during 2008 and 2012, respectively (McCrory et al., 2009, 2013). The Vienna conference provided detailed guidelines for the

diagnosis and management of concussion within the sporting arena. Subsequent conferences have made adjustments consistent with the evolving literature.

The most recent definition of concussion put forward during the 2012 conference defines concussion as "a complex pathophysiological process affecting the brain, induced by biomechanical forces" (McCrory et al., 2013, p. 1). The authors further described four potential features of concussion:

1. "Concussion may be caused by a direct blow to the head, face, neck, or elsewhere on the body with an "impulsive" force transmitted to the head" (McCrory et al., 2013, p. 1).
2. "Concussion typically results in the rapid onset of short lived impairment of neurological function that resolves spontaneously. However, in some cases, symptoms and signs may evolve over a number of minutes to hours" (McCrory et al., 2013, p. 1).
3. "Concussion may result in neuropathological changes, but the acute clinical symptoms largely reflect a functional disturbance rather than a structural injury and, as such, no abnormality is seen on standard structural neuroimaging studies" (McCrory et al., 2013, p. 2).
4. "Concussion results in a graded set of clinical symptoms that may or may not involve loss of consciousness. Resolution of the clinical and cognitive symptoms typically follows a sequential course. However, it is important to note that in some cases symptoms may be prolonged" (McCrory et al., p. 2).

This definition reflects current knowledge regarding concussion. Firstly, it emphasises the metabolic rather than structural dysfunction underlying concussive symptoms. Secondly, it highlights the complexity and the differing symptom presentations of the disorder. However, the author acknowledges this is an ongoing process as there is still plenty that remains unknown regarding concussion (McCrory et al., 2009).

1.2 Pathophysiology and Biokinetics of Concussion

Concussions are induced primarily from acceleration, deceleration, or rotational forces inflicted on the brain caused by an impact to the body, primarily the head,

neck, or face (Barth, Freeman, Broshek, & Varney, 2001; Denny-Brown & Russell, 1940). Impact upon the head causing rotational acceleration by contact with another player or surface, as opposed to linear acceleration from direct impact from a ball or stick, is more likely to result in a concussion (Barth et al., 2001). Studies using helmet-based sensors with football players indicate an average magnitude of 95g is required, but it is not guaranteed to cause a concussion (Brolinson et al., 2006; McCaffrey, Mihalik, Crowell, Shields, & Guskiewicz, 2007). That is equivalent to driving a car into a wall at 48km per hour (Barr & McCrea, 2011).

Contrary to earlier beliefs that concussive symptoms were attributable to structural changes, it is now believed they are a result of metabolic changes (Giza & Hovda, 2001). This is consistent with the failure of Computed Tomography and Magnetic Resonance Imagining scans to identify any visible pathology following concussion (Bazarian, Blyth, & Cimpello, 2006). The metabolic dysfunction is not fully understood; however, based on rodent models, it is hypothesised that forceful impact on the brain triggers a cascade of biochemical reactions (Giza & Hovda, 2001; Katayama, Becker, Tamura, & Hovda, 2009). Immediately following injury an unregulated release of the excitatory amino acid, glutamate, occurs in conjunction with a large efflux of potassium ($K+$), causing depolarisation of the neuronal cell and an influx of calcium ($Ca^{2}+$). The sodium-potassium pump begins to work excessively in an effort to restore the membrane potential. As a result more energy is required. The pump uses energy in the form of adenosine triphosphate (ATP) which is produced from free energy released during glycosis; the process by which glucose is converted into pyruvate. Thus there is a large increase in glucose metabolism known as hypermetabolism (Giza & Hovda, 2001). This occurs in a setting of decreased blood flow. The disparity between glucose (i.e., energy) demand and supply is thought to underlie post-concussive vulnerability. Following this initial reaction a period of hypometabolism occurs. During hypometabolism a persistent increase of calcium occurs that worsens the energy crisis and can lead to cell death. Furthermore, the hyperglycosis leads to an increase in lactate production. Accumulation of lactic acid causes neuronal dysfunction as it alters permeability of the blood-brain-barrier causing fluid build up (i.e. edema) and membrane damage. In summary,

following concussion a cascade of metabolic changes are triggered and while the brain is attempting to restore balance it is in a vulnerable state.

1.3 Signs and Symptoms

There are a wide range of concussive symptoms; however, they generally fall into one of four categories; physical, cognitive, emotional, or sleep-related (Halstead & Walter, 2010). Examples of each are presented in Table 1. Presenting symptoms vary greatly between individuals, depending on the biomechanical forces involved, the location of the affected brain areas, and any history of concussion (Lovell & Pardini, 2010). Athletes may present with only one symptom or a constellation of several symptoms. These symptoms may appear immediately or minutes to hours later. Thus, following a suspected concussion it is essential that all common symptoms are rigorously assessed many times in the post-concussion period (McCrory et al., 2009).

Common on-field cognitive symptoms include confusion, disorientation, and amnesia (Fazio, Lovell, Pardini, & Collins, 2007). Both retrograde amnesia (memory loss of events prior to the incident) and anterograde amnesia (loss of memory of subsequent events) are possible concussive symptoms and are proposed to be a good clinical indicator of concussive severity (Cantu, 2001). Many individuals believe loss of consciousness (LOC) is the hallmark symptom of concussion, yet it occurs in only 10% of sport concussions (Bailes, 2009). LOC has been hypothesised to be indicative of severe concussion and long recovery periods (Cantu, 2001; Halstead & Walter, 2010). However, a study investigating the relationship between LOC and post-injury neuropsychological performance found no difference in test scores between concussed athletes with LOC, and those without LOC (Lovell, Iverson, Collins, McKeag, & Maroon, 1999). Therefore the validity of LOC as an indicator of concussive severity is questionable.

Although immediate motor symptoms are rare, tonic posturing (stiffening) and convulsive movement, although benign, have been known to occur (McCrory et al., 2009). Common physical symptoms include nausea, balance and visual problems, and sensitivity to light and noise (Duff, 2009). Other possible symptoms include changes in emotional reactivity and sleep patterns. The most

common concussive symptom is headache, reported in 70% of cases (Collins et al., 2003). Headaches may develop immediately or in the following several minutes to hours. They are described as a sensation of pressure in the skull either localised or dispersed. Post-concussive headaches are indicative of additional post-concussive difficulties, and their onset is often correlated with memory problems and delayed reaction response (Collins et al., 2003).

Although these symptoms are often brief and spontaneously resolve, some athletes may experience ongoing cognitive difficulties (McCrory et al., 2009). Although symptoms typically resolve spontaneously within 14 days (Halstead & Walter, 2010), studies have found that in some cases the cognitive consequences of concussion can persist up to three months post-concussion (McCrea et al., 2013) or even two years post-concussion (Rees & Bellon, 2007).

Table 1

Common Concussive Signs and Symptoms

Physical	Cognitive	Emotional	Sleep
Headache	Feeling mentally "foggy"	Irritability	Drowsiness
Nausea/vomiting	Difficulty concentrating	Nervousness	Decreased sleep
Balance problems	Difficulty remembering	Increased	Difficulty falling
Visual problems	Forgetful of recent	emotionality	asleep
Fatigue	information		
Sensitivity to light	Confusion		
Sensitivity to noise	Slow response to questions		

1.3.1 Cumulative Effects.

Concussion has been referred to as a cumulative disorder in that the effects sustained by each concussion are additive (Guskiewicz et al., 2003; Iverson et al., 2004). For instance, those who have had a previous concussion are more likely to sustain another concussion in addition to demonstrating longer recovery periods

compared to athletes with no history of concussion (Guskiewcz et al., 2003). Those who suffer from multiple concussions are at risk of developing subtle chronic cognitive deficits (Moser, Schatz, & Jordan, 2005; Wall et al., 2006). For example, Moser and colleagues (2005) found that high school athletes with a history of two or more concussions produced similar scores on neuropsychological tests to those without a history of concussion but who had been concussed within the previous week. They also found that those with two concussions had significantly lower grade-point averages than students without a history of concussion. Additionally, Iverson and colleagues (2004) found that high school athletes with three or more concussions demonstrated more concussive symptoms and performed worse on a memory task than athletes with no concussive history during pre-season. Collins and colleagues (2002) compared on-field concussion presentation between athletes with no history of concussion and athletes who had sustained three or more concussions previously. They found that athletes with a history of concussion were 9.3 times more likely than the non-concussive history group to demonstrate three to four markers of concussive severity, including LOC, anterograde or retrograde amnesia, and confusion.

These studies suggest that not only are the effects of concussion cumulative, but that they have the potential to be long-lasting. Guskiewicz and colleagues (2005) considered the relationship between previous concussion and the likelihood of developing mild cognitive impairment (MCI) and Alzheimer's disease in a group of retired football players with a mean age of 53.8 years. They found an association between recurrent concussion, clinically diagnosed MCI and self-reported memory impairments. Those with three or more concussions were five times more likely to receive a MCI diagnosis and three times more likely to report significant memory impairments compared with retirees with no concussive history. Concussive history was not associated with the presence of Alzheimer's disease, however, it appeared that football retirees had earlier onset of Alzheimer's disease than the general male population.

1.3.2. Long-Term Effects.

Long-term effects of concussion appear predominantly in cognitive and emotional domains (Halstead & Walter, 2010). Residual cognitive symptoms include

memory problems, slowed thinking and reaction time, impaired attention and judgement, and difficulty problem solving. Emotional problems that are known to persist include irritability, restlessness, depression, anxiety, and personality changes (Duff, 2009). The fourth volume of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV TR; American Psychiatric Association, 2000) defines Post-Concussion Syndrome as being a minimum of three months duration of three or more of the following symptoms; fatigue, disordered sleep, headache, vertigo/dizziness, irritability or aggressiveness, anxiety or depression, personality changes, and/or apathy. This diagnosis, however, fails to recognise cognitive symptoms, despite neuropsychological testing demonstrating that the most common difficulties are often found in memory and attention (Moser et al., 2007). While there remains much more to learn about the long-term effects of concussion, preliminary data support the notion that concussion, even at the non-elite level, has the potential to result in long-term consequences (Centre for the Study of Traumatic Encephalopathy, 2009).

1.3.3. Second Impact Syndrome.

If athletes return to physical activity before their symptoms are completely resolved, existing post-concussive symptoms could be exacerbated (Majerske et al., 2008; Mayers, 2008), or a rare condition known as Second Impact Syndrome (SIS; Lovell & Collins, 2002) may occur. Second Impact Syndrome is when brain swelling occurs due to increased volume in extravascular compartments. Following a second impact the brain's ability to regulate its blood supply is compromised (Cantu, 1998; McCrory & Berkovic, 1998). Impairment of vascular regulation results in excessive blood supply to the brain, causing intracranial pressure which in turn can cause cistern obliteration. This is when the brain shifts across structures within the skull such as the flax cerebri, the tentorium cerebelli, and the foramen magnum. Cistern obliteration places extreme pressure on parts of the brain, cutting off blood supply, which can be fatal.

Initially when SIS occurs the athlete does not lose consciousness however, they appear dazed, remaining on their feet for 15 seconds to 1 minute, and then collapsing, appearing semi-comatose with rapidly dilating pupils, loss of eye movement, and respiratory failure (Cantu, 1998). Adolescents appear to be at the

highest risk for SIS, as all reported cases are of athletes under the age of 20 years (McCrory, 2001). Furthermore, severe football head injuries are three times more likely to occur in high school athletes than college athletes (Boden, Tacchetti, Cantu, Knowles, & Mueller, 2007). The reasons for the greater vulnerability in adolescents are unknown. However, it has been hypothesised that adolescents, compared to adults, have weaker necks, thus the impact to the body results in greater force being exerted onto the brain (Powell & Barber-Foss, 1999). Further, adolescent brains are still developing and injury may impede this developmental process.

Schneider (1973) was the first to describe the syndrome when he detailed the case of two athletes who died from minor head injuries while recovering from a previous concussion. Saunders and Harbaugh (1984) later reported the same scenario in a 19-year-old football player and coined the term SIS. Second Impact Syndrome results in fatality for 50% of sufferers and causes long-term brain damage in survivors. There is debate as to whether a second impact is necessary as there are documented cases in which no second hit was observed, yet the athlete still died (McCrory, 2001). It is suggested the second blow can be minor and often not observed (Cantu, 1998). For instance a blow to the chest, often common in contact sports, can indirectly affect the brain.

1.4. Summary

In summary, awareness of concussion in the sporting arena and the potential negative impact on functioning is slowly increasing (Green & Jordan, 1998). Though a universal definition is still lacking and the underlying pathological process is yet to be fully understood, there is a consensus regarding the general areas of impairment following concussion (McCrory et al., 2013). Concussive symptoms can occur in the physical, cognitive, emotional, and sleep-related domains. While these symptoms often resolve spontaneously, there is potential for long-term consequences, usually within the cognitive domain (McCrea et al., 2013; Rees & Bellon, 2007). Returning to play prematurely can exacerbate symptoms and increase the likelihood of long-term impairment. Although our understanding of concussion in the sporting arena remains in its infancy, research

in this area is proliferating, leading to gains in our understanding of sport concussion.

Chapter 2: Assessment and Management of Sport Concussion

There are numerous methods and tools available for the assessment and management of sports concussion, with no one method prevailing as the gold standard (Halstead & Walter, 2010). These methods and tools have evolved over the years as new research emerged and they will continue to do so as long as research in this field continues. Initial attempts to assess concussion were in the form of *grading scales*, which classified concussion by severity levels and the level of severity informed recovery guidelines (Johnston, McCrory, Mohtadi, & Meeuwisse, 2001; Meehan & Bachur, 2009). As research emerged that highlighted the large individual variation in concussion presentation, assessment moved from group-based grading scales to more individualised, multi-method forms of assessment and return-to-play protocol (McCrory et al., 2005; Meehan & Bachur, 2009). Individualised assessment often involves self-report of symptoms, neuropsychological testing, and, in some cases, postural stability assessment (McCrory et al., 2009). In this chapter past and present forms of concussion assessment and management are reviewed.

2.1 Traditional Methods of Concussion Assessment

2.1.1 Grading and dichotomous scales.

Traditionally, grading scales were the predominant method for assessing and managing concussion in sport (Meehan & Bachur, 2009). Over 25 concussion grading scales have been developed, and most evolved from expert opinion rather than empirical evidence (Johnston, McCrory, Mohtadi, & Meeuwisse, 2001). They use self-report symptoms and explicit player observations, such as loss of consciouness (LOC), to deduce concussion severity. The three most commonly used grading scales were those developed by Cantu (2001), the Colorado Medical Society (Schneider & McGrew, 2012), and the American Academy of Neurology (AAN; Kelly & Rosenberg, 1997). All these scales consist of three levels of concussion severity but differ in their classifying indicators.

Cantu's guidelines were first developed in 1986 (as cited in Halstead & Walter, 2010) and then updated in 2001. They classify a concussion with no LOC and less than 30 minutes of symptoms as Grade I. A Grade II concussion is defined by

LOC for less than 1 minute and/or amnesia present for 30 minutes, but no longer than 24 hours. A grade III concussion is diagnosed if LOC persists longer than 1 minute and/or post-concussive symptoms are present for longer than 1 week. The Colorado Medical Society's guidelines suggest only confusion is present in a Grade I concussion, both confusion and amnesia are present in a Grade II concussion, and Grade III involves confusion, amnesia, and LOC (Schneider & McGrew, 2012). The AAN grading scale (Kelly & Rosenberg, 1997) is a revised version of the CMS scale (Colorado Medical Society, as cited in Halstead & Walter, 2010). Confusion lasting less than 15 minutes with an absence of LOC is indicative of a Grade I concussion. In a Grade II concussion these symptoms last longer than 15 minutes. If LOC occurs, a Grade III concussion is diagnosed. The grade of concussion was used to inform RTP decisions. For example if a Grade I concussion was diagnosed, according to Cantu's grading scale, the player was allowed to return to play if they remained asymptomatic for a period of one week (Halstead & Walter, 2010). This was a blanket recommendation for all players classified as Grade I, without considering individual factors such as gender, age, and number of prior concussions.

The failure of grading scales to consider individual differences appeared to hinder their validity. For instance, it was demonstrated that the hypothesised recovery time based on the AAN (1997) and Colorado grading system guidelines was largely incorrect (McClincy, Lovell, Pardini, Collins, & Spore, 2006). If return-to-play decisions were based solely on the grading system guidelines, 80% of concussed athletes in McClincy's (2006) sample would have returned to play prior to cognitive symptoms being resolved. Furthermore, most grading scales use LOC as one of their main indicators of concussion grade (Cantu, 2001). However, inferring the duration of LOC is largely guesswork, as by the time the team physician comes in contact with the player they could have already lost consciousness for some time or had lost consciousness but have since regained consciousness.

More recently, a dichotomous classification scale was developed, diagnosing concussion as either simple or complex (McCrory et al., 2005). A simple concussion resolved within 10 days whereas a complex concussion was defined as either a concussion in which symptoms persisted for longer than 10 days or if

concussive convulsions, prolonged LOC, and/or multiple previous concussions were present. Hence, the simple/complex classification is diagnosed retrospectively. However, recommendations from the 2008 Concussion in Sport Conference suggested that all classification/grading scales be abandoned given their lack of empirical support and their failure to consider individual factors (McCrory et al., 2009). For instance, the scales did not account for the fact that some individuals recover faster, are more susceptible to concussion, or report more symptoms than others (Makdissi, 2009). It is probable that the lack of consideration of individual factors contributes to the poor validity of grading scales and dichotomous classification systems.

2.2 Individual differences in the presentation and recovery of concussion

Growing evidence indicates that age is an individual difference that requires consideration. Adolescents and children have been found to recover slower from concussion than adults. In a large sample ($N = 1631$) of college football players, McCrea and colleagues (2004) found the cognitive effects of concussion, evidenced by neuropsychological testing, subsided within 5 to 7 days post-concussion, whereas in a sample of high school students 14 days post-concussion was needed for most athletes to return to their cognitive baseline performance (McClincy et al., 2006). Furthermore, Field, Collins, Lovell, and Maroon (2003) compared recovery rates of high school and college athletes to healthy controls. They found college athletes' cognitive performance matched that of healthy controls at day three post-concussion. However, high school athletes continued to perform significantly worse than age-matched controls at seven days post-concussion, suggesting that younger age is associated with slower recovery time. The prolonged deficit observed in high school athletes versus college athletes is despite college athletes sustaining more severe concussions throughout the season. Lastly, a study comparing recovery times of National Football League (NFL) and high school players found that NFL players recovered within a week with most showing no symptoms of concussion at two days after the injury. By contrast, high school athletes continued to experience difficulties in reaction time and processing speed (Pellman, Lovell, Viano, & Casson, 2006), suggesting they take longer to recover than adults.

The underlying mechanisms of prolonged recovery observed in younger individuals compared to adults are unclear; however, several hypotheses have been proposed. Firstly, there may be a selection bias in that high school athletes who are more prone to concussion or who recover slower do not go on to play during their adult years (Pellman et al., 2006; Pellman, Lovell, Viano, Casson, & Tucker, 2004). There is also evidence of concussion resulting in more diffuse and prolonged cerebral swelling in younger individuals compared to adults (Lang, Teasdale, Macpherson, & Lawrence, 2009; Pickles, 1950). And lastly, it appears children's brains are 60 times more sensitive to glutamate, an excitatory neurotransmitter excessively released during the immediate post-concussion stage (McDonald & Johnston, 1990). All of these are potential explanations for the prolonged recovery observed in younger individuals.

Gender is also a potential source of individual variation in the presentation of concussion, particularly among high school athletes (Dick, 2009). It has been reported that adolescent girls are almost twice as likely to sustain a concussion compared to boys in soccer, ice hockey, basketball (Dick, 2009), and softball/baseball (Powell & Barber-Foss, 1999). However, the incidence of concussion among college-aged soccer (Green & Jordan, 1998) and basketball players (Echemendia, Putukian, Mackin, Julian, & Shoss, 2001a) was invariant across genders. The mechanisms by which gender influences the presentation of concussion, among adolescents at least, remain unknown.

Concussion history has also been noted as a source of individual variation, which has been linked to concussion risk and duration of post-concussion recovery (Guskiewicz et al., 2003). As described in the previous chapter, two or more concussions appear to increase an individual's risk of sustaining subsequent concussions. If a subsequent concussion does occur, the symptoms are more severe and recovery time is longer than an individual with a first time concussion (Guskiewcz et al., 2003).

2.3 Individualised Assessment and Management

Modern sport concussion assessment and management adopts an individualised, multi-method approach, typically consisting of self-reported symptoms, neuro-cognitive functioning, and/or postural stability (Broglio & Puetz, 2008).

2.3.1 Self-report symptom scales

The assessment of self-report symptoms is often completed with checklists or scales and has been a consistent component of concussion management, preceding postural and neuropsychological assessment (Meehan & Bachur, 2009). Self-report scales/checklists measure an individual's perceived symptoms at a specific point in time and have repeatedly demonstrated their ability to detect concussion (Macciocchi, Barth, Alves, Rimel, & Jane, 1996; McCrea, Guskiewicz, Marshall, & et al, 2003). They are the most practical option for concussion screening given their simplicity to administer, both in regards of resources and time.

Numerous concussion symptom scales exist, but most originate from six core scales (Alla, Sullivan, Hale, & McCrory, 2009). These include the Pittsburgh Stellers Post-Concussion Scale (17-items; Maroon, 2000), the Post-Concussion Symptom (PCS) Questionnaire (10-items; Cameron, Yunker, & Austin, 1999), the Concussion Resolution Index (CRI; 15-items; Erlanger, Feldman, & Kutner, 1999), the Signs and Symptom Checklist (34-items; Pellman, Lovell, Viano, Casson, & Tucker, 2004), the Sports Concussion Assessment Tool (SCAT; 25-items; McCrory et al., 2005), and the Concussion Symptom Inventory (CSI; 12-items; Randolph et al., 2009).

The Immediate Post-concussion Assessment and Cognitive Test (ImPACT) battery used in the current study includes a symptom scale that is a modified version of the Pittsburgh Steelers Post-Concussion Symptom Scale (PCSS) which was originally developed in the late 1980s (Lovell et al., 2006). The modified version used in ImPACT is referred to the Post-Concussion Scale (PSC) and consists of 22 items, as opposed to the 17 original items (Lovell et al., 2006). Athletes indicate which symptoms, if any, they have experienced over the past 72 hours in addition to indicating the severity of each symptom on a 7-point Likert scale. The PCSS is currently used to monitor symptoms on the sideline and for subsequent follow up. Factor analysis of the PCS, employing 11 to 18 year olds, indicated a four factor structure with items measuring somatic, cognitive, sleep, and affective factors (Pardini et al., 2004). The PCSS has demonstrated robust internal consistency, with the *Cronbach alpha* reported to be .88 to .94 for a control group and .83 for a concussed group (Lovell et al., 2006).

2.3.2 Postural stability assessment.

Postural stability assessment appears to be the least commonly used method of assessing concussion, perhaps because it is time-consuming to administer and assesses only one symptom of concussion. Impairment in postural control has been found following concussion. It is believed to result from a sensory integration deficit in the balance mechanism (Guskiewicz, Riemann, Perrin, & Nashner, 1997; Guskiewicz, Ross, & Marshall, 2001). Although there is limited research on postural controls associated with concussion, Broglio and Puetz (2008) found concussion had a large effect on postural control immediately after ($d = -2.56$) as well as 14 days post-concussion ($d = 1.16$). Disturbances in balance usually resolve within 3-5 days post-injury (McCrea, et al., 2003). A commonly used postural control assessment is the Balance Error Scoring System (BESS, Guskiewicz, 2003). This consists of a series of stances, including double and single-legged stances on a firm, foam, or tremor box surface. Participants attempt to hold each stance for 20 seconds with their hands on their hips and eyes closed (Wilkins, Valovich McLeod, Perrin, & Gansneder, 2004).

2.3.3 Neuropsychological assessment

Neuropsychological testing is the third method often used in concussion assessment. Neuropsychological tests are tasks designed to measure specific cognitive or psychological functions that are typically known to be associated with particular brain structures or pathways (Lezak, 1995). Neuropsychological testing has been termed the cornerstone of concussion management by the Concussion in Sport Concensus Group (McCrory et al., 2009), given its ability to detect subtle effects of concussion in the absence of self-reported symptoms (Van Kampen, Lovell, Pardini, Collins, & Fu, 2006). Although neuropsychological testing provides an objective measure of brain function, it is only one component of concussion assessment and should not be independently used to diagnose concussion or inform return-to-play decisions (Ellemberg, Henry, Macciocchi, Guskiewicz, & Broglio, 2009).

Barth and colleagues (1989) were the first to investigate the utility of neuropsychological testing in detecting concussion within the sports arena. They conducted a large study, employing a baseline model in which they tested athletes

pre-season and post-concussion. They found neuropsychological testing was effective at detecting concussion. In 1990, Lovell adapted Barth's baseline model of assessment and implemented it as part of the Pittsburgh Steelers' concussion management programme (Lovell, 1999). This marked the first clinically-orientated project in professional sport to use neuropsychological testing to assess concussion and contribute to return-to-play decisions. The NFL took notice of Lovell's initiative and in 1994 created the Mild Traumatic Brain Injury committee (Pellman, Viano, Tucker, Casson, & Waeckerle, 2003). Initially neuropsychological testing was voluntary within the NFL; however in 2007 baseline and post-injury neuropsychological assessment was mandated for all NFL athletes (Lovell & Solomon, 2011).

The use of neuropsychological testing for sports concussion gradually spread to other sporting codes such as ice hockey (Lovell & Burke, 2000), automobile racing (Olvey, 2002), soccer (Matser, Kessels, Lezak, Jordan, & Troost, 1999), rugby (Shuttleworth-Edwards & Radloff, 2008), skiing (Lovell & Solomon, 2011), and Australian Rules football (Makdissi et al., 2001). At present almost all professional sports teams incorporate neuropsychological testing into their concussion management procedures (Lovell & Solomon, 2011; Pellman et al., 2006; Van Kampen et al., 2006). At lower levels however such as in high schools, where concussion is most prevalent, neuropsychological testing is rare. This is perhaps due to financial limitations and lack of knowledge regarding potential consequences of concussion.

The NFL battery consisted of mainstream neuropsychological tests which, instead of examining full cognitive function, focused on specific areas believed to be affected by concussion, such as attention, memory, and processing speed (Ellemberg et al., 2009). Each test and the cognitive area assessed are listed in Table 2. The traditional paper and pencil neuropsychological tests that comprised the NFL battery were resource intensive and time-consuming as they required a neuropsychologist to administer them to each athlete individually and manually score each test (Lovell & Solomon, 2011). Furthermore, some of the tests such as Digit Span were insensitive to the subtle effects of concussion (Ellemberg et al., 2009). Additionally, there were no alternative forms of each subtest, therefore

practice effects were prominent (Strauss, Sherman, & Spreen, 2006), meaning that validity and reliability were weak.

Table 2

Areas of Cognition Assessed by Individual Tests that comprise the National Football League Battery

Test	Areas assessed
HVLT-R	Verbal memory (learning, delayed, recognition)
BVMT-R	Visual memory (learning, delayed, recognition)
TMT-A	Visual scanning & processing speed
TMT-B	Visual scanning, processing speed, & cognitive flexibility
COWA	Verbal fluency
WAIS; Symbol Search	Visual processing speed, perceptual organisation, & planning
Digit Symbol Coding	Psychomotor processing speed
Digit Span	Working memory & complex attention

Note. HVLT-R = Hopkins Verbal Learning Test − Revised; BVMT-R = Brief Visuospatial Memory Test − Revised; TMT = Trail Making Test; COWA = Controlled Oral Word Association; WAIS = Wechsler Adult Intelligent Scale.

2.3.3.1 Computerised neuropsychological screening.

To address the limitations of paper and pencil tests, computerised neuropsychological tests specific to sports concussion were developed (Patel, Shivdasani, & Baker, 2005). There are many advantages of computerised testing. They do not require a neuropsychologist to administer them, thus they are cost effective as they can be administered to a large group at the same time (Patel et al., 2005). Computerised testing ensures standardised administration, which increases reliability. Furthermore, they provide a more accurate measure of reaction time, measuring up to 0.01 of a second. They can randomise test items and generate alternative forms of the same test, minimising practice effects in

repeat administrations. Lastly, most computerised tests automatically process raw scores into standardised scores (Lovell & Collins, 2002), removing the need for manual test scoring. Although the benefits over paper and pencil testing are vast, a disadvantage of computerised testing is that it does not allow for the administrator to observe the idiosyncratic behaviours during testing that often provide valuable information.

Several computerised neuropsychological assessment programmes exist. The most commonly used are the Automated Neuropsychological Assessment (ANAM; Cernich, Reeves, Sun, & Bleiberg, 2007), Cogstate Sport (Makdissi et al., 2001), Headminders Concussion Resolution Index (CRI; Erlanger et al., 1999), and Immediate Post-concussion Assessment and Cognitive Testing (ImPACT; ImPACTonline, 2013). The ANAM was initially developed by the US Department of Defence to assess the effects of chemical exposure and environmental stressors on military personal (Cernich et al., 2007). The ANAM Sports Medicine Battery (ASMB) is a subtest within the ANAM that is sensitive to the cognitive effects of concussion and has been used by the US military academy in their concussion management programme since 1999 (Reeves, Winter, Bleiberg, & Kane, 2007). The ASMB has demonstrated adequate test-retest reliability and internal consistency in addition to validity when compared to traditional paper and pencil neuropsychological tests (Cernich et al., 2007).

The Concussion Resolution Index (CRI; Erlanger et al., 1999) was developed by a company named HeadMinder, to be used specifically in the sports arena. It has a three-factor structure, measuring processing speed, simple reaction time, and complex reaction time. It too has demonstrated adequate reliability and validity and is currently used by many athletic organisations at different levels (Schatz & Zillmer, 2003).

Cogstate was developed in Australia, originally validated on 300 professional Australian Rules football players and hundreds of control participants (Makdissi et al., 2001). It too has demonstrated adequate reliability and validity (Schatz & Zillmer, 2003), but there is, however, a paucity of normative data (Collie, Darby, & Maruff, 2001).

ImPACT was created in the US by Mark Lovell and Joseph Marron (ImPACT online, 2013). It was the first computerised test battery developed to assess sport concussion and appears to be the most widely-used battery. The developers of ImPACT claim it to be the most scientifically validated neuropsychological computerised assessment for sport concussion (ImPACTonline, 2013). ImPACT has been used with high school, college, and professional level athletes (Lovell et al., 2006). It was developed to address limitations of the traditional paper-and-pencil NFL battery. It was initially developed as a desktop program in 2000, with an online version released in 2008 (Elbin, Schatz, & Covassin, 2011a). ImPACT follows a baseline model in which athletes act as their own controls, which individualises concussion assessment. Athletes complete the ImPACT battery pre-season and this acts as a baseline measure to which post-concussion testing is compared. This model has been evaluated and supported as an effective method of assessing cognitive impairment following a concussion (Barth et al., 1989; Iverson, Brooks, Collins, & Lovell, 2006; Schatz & Zillmer, 2003).

ImPACT consists of three main sections; demographic and health history information, a symptom inventory, and the neuropsychological test modules (ImPACTonline, 2013). The symptom inventory is the PCS mentioned previously in which athletes indicate on a 7-point Likert scale the severity of any symptoms they have experienced in the past 72 hours. From this a total symptom score is calculated. There are six different test modules (Word-memory, Design-memory, X's and O's, Symbol Match, Colour Match, and Three Letters) which were derived from traditional neuropsychological tests included in the NFL battery (ImPACTonline, 2013). Scores from the six test modules combine to produce five composite scores; Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. A total score is not produced as test developers believe it would most likely be insensitive to concussive symptoms as not all concussed athletes will show decline in all cognitive areas assessed (ImPACTonline, 2013). The content of the ImPACT battery is discussed in more detail in the methodology section.

2.4 Summary

The way in which sport concussion is assessed and managed has changed over the years due to gains in our knowledge regarding its presentation and course of recovery. Assessment has shifted away from a simple, generalised approach toward a more individualised, multi-method assessment (McCrory et al., 2005; McCrory et al., 2009). The Concussion in Sport Consensus group have suggested that assessment should include both a self-report of symptoms in addition to neuropsychological testing (McCrory et al., 2009). To increase efficiency of administration, neuropsychological testing is now computerised and can be completed online. The ImPACT battery includes both neuropsychological tests and a self-report symptom scale. It appears to be the most widely used sport concussion battery. Notwithstanding, it is not without its own limitations, particularly regarding the lack of theoretical foundations and empirical research supporting its accuracy and reliability in detecting the cognitive effects of concussion. Hence the following two chapters focus on potential theoretical underpinnings and psychometric properties of the ImPACT battery.

Chapter 3: The Theoretical and Empirical Underpinnings of Concussion Induced Cognitive Impairment

Previous chapters have reviewed the literature pertaining to concussion in sport and, more specifically, the neuropsychological assessment of concussion within the sporting arena. The neuropsychological assessment instrument of interest in the current study is the Immediate Post-concussion Assessment and Cognitive Test (ImPACT). Assessment tools, such as ImPACT should be grounded in theory (Buckendahl & Plake, 2006), yet the theoretical foundation underpinning the development of ImPACT could not be located in the literature or in ImPACT's technical manual (ImPACTonline, 2013). Although the current study is a psychometric evaluation, it is an evaluation of theoretical constructs and thus a theory is needed to guide the inquiry in addition to providing a framework from which to interpret the results. The Cattell-Horn-Carroll theory was chosen given its robust support in the literature. It provides a taxonomy for understanding and studying cognitive constructs which are, by nature, inter-related and sometimes hard to separate (Keith & Reynolds, 2010). However, it is comprehensive and inclusive of all academic and cognitive abilities, all of which are not affected by concussion. Thus one must also look to empirical research regarding the specific domains affected by concussion to guide and provide a context for the current study (Buckendahl & Plake, 2006).

3.1 The Cattell-Horn-Theory of Intelligence

The Cattell-Horn-Carroll (CHC) Theory is currently the most comprehensive and empirically supported psychometric theory of cognitive and academic abilities (Keith & Reynolds, 2010; McGrew, 2005, 2009). It was originally developed from the work of McGrew (1997) and represents a taxonomy regarding the structure of cognitive abilities. Keith and Reynolds (2010) claim that most recently developed or revised intelligence tests are based on CHC theory, or at least acknowledge it in their development. For example, the Woodcock-Johnson test (2001) of cognitive abilities is explicitly grounded in CHC theory (Woodcock, McGrew, & Mather, 2001; Woodcock, McGrew, Mather, & Schrank, 2003) and the highly regarded Wechsler intelligence batteries, which were originally atheoretical, have referenced the CHC theory in their more recent test manuals

(Wechsler, 2003, 2008). Furthermore, *confirmatory factor analyses* of the Woodcock-Johnson test of cognitive abilities (McGrew, Werder, & Woodcock, 1991; McGrew & Woodcock, 2001), the Stanford-Binet test (Roid, 2003), the Differential Ability Scale (Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007; Stone, 1992), and the Wechsler tests (Chen, Keith, Chen, & Chang, 2009; Keith, Fine, Taub, Reynolds, & Kranzler, 2006) have identified structures which are compatible with the CHC theory. For example, McGrew and Woodcock (2001) completed several confirmatory factor analyses using the Woodcock-Johnson III that compared the CHC model to other models of intelligence. They found that the CHC theory, compared to other models, provided a superior fit to the data.

The CHC theory evolved from McGrew's (1997) attempt to resolve the differences between Cattell-Horn's (1991) Gf-Gc theory and Carroll's (1993) Three-Stratum theory. The result was an amalgamation of the two theories of intelligence and became known as the Cattel-Horn-Carroll (CHC) theory (McGrew, 1997). The traditional Gf-Gc theory postulated that intelligence incorporates approximately 100 abilities that interact in varying ways in different people and that these abilities separate into two broad forms of intelligence; Crystallised Intelligence (Gc) and Fluid Intelligence (Gf). Cattell (1941) stated that Crystallised Intelligence refers to acquired knowledge, often evidenced by one's vocabulary and general knowledge. Fluid intelligence on the other hand refers to an individual's ability to reason, think logically, and problem-solve. Cattell claimed that Crystallised Intelligence continues to increase with age, whereas Fluid Intelligence increases until approximately age 20 at which point it begins to gradually decline. Cattell's hypothesis continues to be supported by current literature (e.g., Brevik, 2012; Cavanaugh & Blanchard-Fields, 2006; Rogers, Kang, & Miller, 2007). Horn (1965) later expanded on Cattell's (1941) original dichotomous model to include eight forms of intelligence, Visual Perception or Processing (Gv), Short-term Memory (Gsm), Long-term Memory (Glr), Speed of Processing (Gs), Auditory Processing (Ga), Reaction Time and Decision Speed (Gt), Quantitative Ability (Gq), and Broad Reading/Writing Ability (Gw). This expanded theory is now known as the Cattell-Horn Gf-Gc theory of intelligence (Horn, 1991) and is depicted in Figure 1.

The Three Stratum Theory resulted from Carroll's (1993) extensive and systematic *exploratory factor analysis* (EFA) of over 460 cognitive ability data-sets. His research was innovative as it was the first empirically-based taxonomy of cognitive ability presented in a single organised framework. Carroll (1993) proposed cognitive abilities could be best understood via three various strata. Stratum III was the broadest strata and represented general intelligence consistent with Spearman's (1927) concept of 'g', which encapsulated eight broad (Stratum II) abilities; Crystallized Intelligence (Gc), Fluid Intelligence (Gf), General Memory and Learning (Gy), Broad Visual Perception (Gv), Broad Auditory Perception (Gu), Broad Retrieval Ability (Gr), Broad Cognitive Speediness (Gs), and Processing Speed (Gt). These eight broad abilities could further be divided into 73 narrow (Stratum I) cognitive abilities, such as Language Comprehension, Memory Span, Language Development, and General Sequential Reasoning.
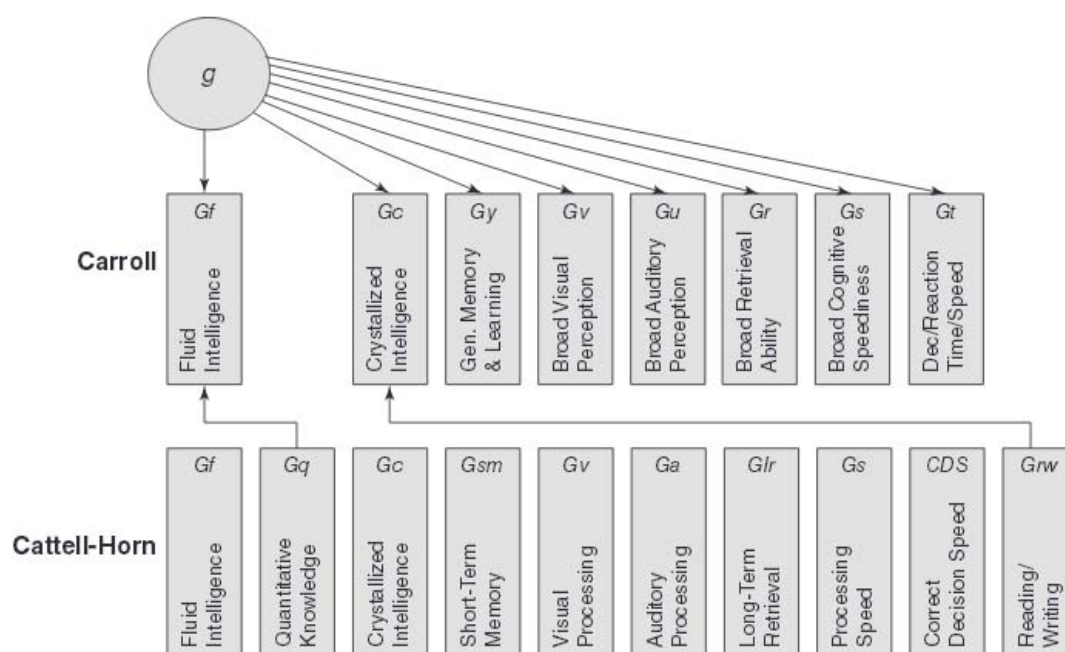
Figure 1. Comparison of Cattell-Horn Gf-Gc theory and Carroll's Three-Stratum theory. Source: Flanagan & McGrew (1997).

The CHC model currently consists of 16 broad abilities and over 80 narrow abilities (Schneider & McGrew, 2012). The original 10 abilities included Fluid Intelligence (Gf), Crystallised Intelligence (Gc), Reading/Writing Ability (Grw),

Short-term Memory (Gsm), Quantitative Knowledge (Gq), Long-term Storage and Retrieval (Glr), Visual Processing (Gv), Auditory Processing (Ga), Processing Speed (Gs), and Decision Speed/Reaction Time (Gt). The model does not include 'g', unlike Carroll's (1993) original theory. There remains debate regarding whether 'g' exists or not, however given most cognitive tests aim to identify specific areas of strengths and weaknesses, a construct of 'g' is generally considered not to be of importance (Schneider & McGrew, 2012). In the most recent revision of the CHC theory, Schneider and McGrew (2012) added six broad abilities, General Knowledge (Gkn), Olfactory Abilities (Go), Tactile Abilities (Gh), Psychomotor Abilities (Gp), Kinaesthetic Abilities (Gk), and Psychomotor Speed (Gps). Most intelligence tests do not measure these abilities as they contribute little to the prediction of traditional concepts of achievement, the predominant aim of intelligence tests (Flanagan, Ortiz, & Alfonso, 2013). However, in tests that aim to assess the impact of neurologically-based disorders, such as Traumatic Brain Injury (TBI), these motor and sensory abilities are perhaps more relevant. The purpose of a test battery will determine what cognitive domains should be assessed, and, given the extensiveness of the CHC theory, it is unlikely any one test battery would assess all CHC specified domains. Instead the differentiation of cognitive abilities should adhere to the CHC theory while the specific areas chosen for assessment should be informed by prior research in the area of interest.

3.2 Cognitive Domains Affected by Concussion

In the sport concussion literature there appears to be a general consensus that memory, attention, working memory, processing speed, and executive functions are the predominant areas affected by concussion (Barr & McCrea, 2001; Bruce & Echemendia, 2003; Echemendia, Putukian, Mackin, Julian, & Shoss, 2001; McCrea, 2003; McCrea, Kelly, Kluge, Ackley, & Randolph, 1997). These areas are indicative of one's level of fluid intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Fry & Hale, 2000). Crystallised abilities such as vocabulary and general knowledge are typically robust against head injuries, specifically mild forms such as concussion (Barth et al., 1989; Levin, Benton, & Grossman, 1982). Thus, it is uncommon for crystallised abilities to be measured by concussion assessments.

In a meta-analysis of 21 studies, researchers examined the effect of sport concussion on six different cognitive domains; orientation, attention, executive functions, memory acquisition, delayed memory, and global cognitive ability (Belanger & Vanderploeg, 2005). Overall, the study included 790 concussed individuals and 2014 control participants. The concussed group exhibited significant deficits in all cognitive domains except for attention and executive functions. Global cognitive ability demonstrated the largest effect size ($d = .81$), followed by memory acquisition ($d = .78$), delayed memory ($d = .60$), and orientation ($d = .27$). When cognitive performance was separated by time of assessment, there was a significant effect size for attention at 24-hours ($d = .51$) and 1-7 days ($d = .35$) post concussion. However, the effect size for executive functions at both time points remained non-significant for studies that compared concussed individuals performance to healthy controls. Of the 21 studies included in the meta-analysis, 19 included an assessment of attention, 10 measured delayed memory, seven measured memory acquisition and executive functions, six measured global cognitive abilities, and four assessed orientation.

Attention was measured by most of the studies included in the meta-analysis (Belanger & Vanderploeg, 2005). This suggests that the various researchers believed that impaired attention is frequently present in concussion. The neuropsychological tests the authors included in the 'attention' category measure not only attention but also working memory and processing speed. For example, the attention category included tests such as the reaction time and processing speed items of the ImPACT battery, the processing speed, complex and simple reaction time subtests of the Concussion Resolution Index (CRI; Erlanger et al., 1999), the Digit Span and Digit Symbol subtests from the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008) and the Paced Auditory Serial Addition Task (PASAT; Gronwall, 1977). To complete these tasks, although one cognitive ability may be the focus, all three abilities are needed to some extent (Gronwall, 1977). Thus, it is extremely difficult to separate out the constructs of attention, working memory, and processing speed, although they are distinct abilities. Nevertheless the meta-analysis highlights that test administer's generally assume that attention, processing speed, and working memory are important areas to assess in concussion (Belanger & Vanderploeg, 2005).

3.2.1 Attention, Processing Speed, and Working Memory

The importance of attention, processing speed, and working memory in concussion assessment can be better understood by looking at the older adult literature regarding cognitive decline (Bashore & Richard, 2002). In this literature, Salthouse (1996) highlights how the constructs of attention, working memory, and processing speed are often used interchangeably or lumped together in one category, despite being distinct functions. In his research he separates out these constructs and postulates that processing speed underlies the deficits seen in other cognitive domains among healthy aging adults (Kail & Salthouse, 1994; Salthouse, 1996). For instance, Kail and Salthouse (1994) demonstrated that 60% of cognitive decline among healthy older adults could be explained by decrements in speed of processing. The cognitive sequel observed in TBI is similar to that seen with natural aging, suggesting that the two processes may have an impact on common underlying neural mechanisms (Bashore & Richard, 2002; Hicks & Birren, 1970; Miller, 1970). Based on the author's perusal of the literature, it appears that the literature pertaining to the cognitive effects of aging are more developed than that of the cognitive effects of TBI. Thus the adult literature can perhaps provide additional insight into cognitive sequela following TBI.

Salthouse (1996) proposed that slowed information processing speed contributes to deficits in other areas via two distinct mechanisms, a time-limited mechanism and a simultaneity mechanism. The time-limited mechanism refers to the fact that, if an individual experiences a reduction in the speed at which they can process information then, in a finite period of time, the amount of information processed will be less than usual (Salthouse, 1996). Deficits of this type would be apparent in tasks, which require an individual to complete as many tasks as possible in a pre-specified time. The simultaneity mechanism, in which slowed processing speed can impact performance of other cognitive tasks, occurs in a context in which two tasks need to be completed. Due to slowed processing speed, earlier cognitive actions or information may not be available by the time later or more complex cognitive actions of the same task are supposed to be taking place. This can affect tasks that have an infinite time period in which they may be completed. For example, this mechanism can affect a task that requires information to be held in working memory and used to complete a series of actions. If one is slow at

completing the required actions due to slowed processing speed, the information held in working memory is more likely to decay compared to someone without slowed processing speed who can complete the actions faster. Thus, fewer actions will be completed, as without the information held in working memory one does not have the necessary information to complete the actions. Consequently, slowed processing speed is evident in both timed and untimed tasks (Salthouse, 1996).

This model of information processing speed provides a potential explanation of how slowed processing speed may contribute to deficits in other cognitive domains among healthy aging adults. Salthouse's (1996) processing speed hypothesis of cognitive impairment is also supported by the TBI literature. For example, studies examining severe TBI patients performance on neuropsychological tasks found that, once they controlled for simple processing speed, deficits in divided attention (Brouwer, Ponds, Van Wolffelaar, & Van Zomeren, 1989; Spikman, Zomeren, & Deelman, 1996), focused attention (Spikman et al., 1996), switching attention (Brouwer et al., 1989), and executive functions of planning and inhibition were not apparent (Brouwer et al., 1989; Veltman et al., 1996). Slowed processing speed is postulated to result from diffuse axonal shearing. Axonal shearing is when the fibres that transfer information in the brain are damaged and thus the speed at which information can travel via them is impeded (Keith & Reynolds, 2010; McGrew, 2009; van Zomeren & Brouwer, 1994).

Slowed information-processing speed is one of the most commonly cited and most disruptive cognitive impairments following TBI (Azouvi, Jokic, Der Linden, Marlier, & Bussel, 1996; McDowell, Whyte, & D'Esposito, 1997). Wilberger and colleagues (1993) found that, in a group of high school football players who sustained a concussion, 75% demonstrated deficits in information processing speed at 24 hours, 61% at one month, and 55% of the sample continued to show deficits at three-months post-concussion. Furthermore, McCrea and colleagues (2004) assessed 1631 concussed college football players processing speed at 2 and 7 days post-injury using traditional paper and pencil tests. At both time points participants demonstrated mild impairment in their processing speed. The robust relationship between impaired processing speed and concussion is supported by the ability of processing speed measures to differentiate between severity levels of

TBI, and between TBI patients and healthy controls (i.e., criterion validity; Donders et al., 2001; Hawkins, 1998; Martin, Donders, & Thompson, 2000). For example, in Martin and colleagues' (2000) study, the Processing Speed index of the WAIS-III was the only index that demonstrated criterion validity, in that it was sensitive to the severity of TBI. Thus, information processing speed should be a prominent area of interest in the cognitive assessment of concussion.

3.2.1.1. Simple versus Complex Processing Speed. Information processing speed refers to the time required to complete a cognitive task, or if given a finite time, the proportion of work that can be completed (Lezak, Howieson, & Loring, 2004). As mentioned previously, confusion exists in the literature regarding the definition of processing speed and the distinction between itself and other similar constructs such as complex attention, working memory, reaction time, and information processing. According to CHC theory, processing speed is distinct from the construct termed reaction/decision time (Flanagan & Harrison, 2012). Flanagan and Harrison (2012) define reaction time as the speed at which one can respond to a single stimulus, whereas processing speed is the speed at which one can respond to a series of stimuli items, completed in succession. The main difference is that in a reaction time task the stimuli or trials are presented one at a time so that the speed of perception and quickness of response can be gauged each time. Importantly, the examiner controls the pace of presentation, often allowing brief breaks between trials. In contrast, processing speed tasks present all stimuli simultaneously, the examinee determines the pace of progression and must continuously shift attention from one item to the next therefore requiring sustained attention for the duration of the task. This distinction was supported by Chiaravalloti, Christodoulou, Demaree, and DeLuca's (2003) factor analytic study that found reaction time tasks and processing speed tasks (i.e., Paced Auditory Serial Addition Test; PASAT) loaded onto two distinct latent factors.

Salthouse's (1996) theory of information processing speed is consistent with the CHC theory in that both describe two forms of processing speed. Salthouse argues that there is simple processing speed, which can be measured by reaction time tasks and is synonymous to CHC's reaction/decision time construct. Whereas what he refers to as complex processing speed appears synonymous to CHC's

processing speed construct. Both DeLuca and Kalmar (2013), and Lezak (1995) claim that reaction time is a behavioural indicator of processing speed. Simple processing speed (i.e., reaction time) involves only sensory perception of the stimuli and a motor response. Complex processing speed on the other hand often requires working memory and executive attention. For instance, the PASAT is frequently used to assess complex processing speed even though validation studies consistently demonstrate that it also measures working memory and executive attention (Lezak, 1995; Roman, Edwall, Buchanan, & Patton, 1991; Sherman, Strauss, & Spellacy, 1997). One could thus argue that complex processing speed is not a pure measure of processing speed, as tasks used to assess it rely heavily on other cognitive functions. Thus, perhaps reaction time tasks or simple processing speed, are the best indicator of general processing speed.

3.2.1.2 Attention. Attention refers to several abilities or processes concerned with how an individual becomes receptive and responsive to stimuli in their environment (Lezak et al., 2004). It requires both alertness and arousal and is often a pre-requisite for higher-level cognitive functions. There are varying theories of attention, however common features include the ability to orientate to a stimulus, selectively attend to a stimulus while ignoring others, and maintaining focus on a stimulus for the required time (Scott, 2011). Concentration is a term closely associated with attention. It refers to the capacity to sustain attention while ignoring irrelevant stimuli (i.e., distractions; Scott, 2011).

Attention is a complex process involving multiple brain regions (Scott, 2011). Predominant areas include the orbital prefrontal cortex, which is important for sustaining attention, the dorsolateral prefrontal cortex, important for the initiation of attention, and the general pre-frontal cortex, required for voluntary initiation and sustaining of attention, rapid alternation, and shifting of attentional resources (Scott, 2011). The superior and inferior colliculi regulate automatic orientation to visual and verbal stimuli, respectively. The ascending reticular activating system stimulates arousal in the cortex needed to initiate and sustain attention. Other important areas include the parietal lobes, thalamus, and limbic system (Scott, 2011). Impairment in attention can result following damage to any part of the brain, however frontal lobe damage in particular appears to have the most

catastrophic effect (Schoenberg & Scott, 2011). Attentional deficits are almost universally present in TBI presentation regardless of severity (Gronwall, 1991). However, previously mentioned studies (e.g. Brouwer et al., 1989; Spikman et al., 1996) that failed to find deficits in attention once processing speed had been controlled for cast doubt on this statement. Given most measures of attention also require processing speed (e.g., PASAT), it may be that it is the deficit in processing speed and not attention that is leading to the diminished performance observed.

   3.2.1.3 Working Memory. Working memory refers to the ability to hold information in one's awareness temporarily and to manipulate it if required (Knudsen, 2007). The amount of information able to be held at any one time is limited to seven, plus or minus two pieces of information (Miller, 1956). Baddeley and Hitch (1974) postulated that working memory includes subsystems which store and manipulate visual and verbal information, in addition to a 'central executive' which co-ordinates these subsystems. The regions of the brain predominantly involved in working memory include the frontal and parietal regions (Rottschy et al., 2012). Working memory and processing speed, although closely related are distinct constructs as evidenced by factor analytic studies (Arnau & Thompson, 2000; Chiaravalloti et al., 2003). Processing speed is thought to impede working memory in that, slowed processing speed allows more time for the content in working memory to decay (Chiaravalloti et al., 2003; Demaree, DeLuca, Gaudino, & Diamond, 1999).

   3.2.2 Memory

Memory is the process in which information is encoded, stored, and retrieved. Impairment can occur at any stage of this process (Scott & Schoenberg, 2011). Memory is complex and involves multiple brain regions, some of the most important structures being the medial temporal lobes, entorhinal cortex, hippocampus, amygdala, cingulate cortex, basal forebrain, diencephalic structures, and the anterior temporal cortex (Schoenberg & Scott, 2011). There is some evidence for a lateralized effect of memory, with the left temporal lobe associated with verbal memory, memory for words, stories, and numbers, whereas the right temporal lobe is thought to be prominently associated with visual memory,

including faces and figures (Schoenberg & Scott, 2011). A distinction can also be made between immediate, delayed, and recognition memory. The latter is when information had been encoded but unable to be spontaneously retrieved. When possible, assessment should include an investigation of immediate, delayed, and recognition memory in both the visual and verbal modalities (Schoenberg & Scott, 2011).

The only division of memory in the CHC theory is that of short-term and long-term memory. However, memory literature often divides memory into two major categories, declarative and procedural (Anderson, 2013; Lum, Conti-Ramsden, Page, & Ullman, 2012; Scott & Schoenberg, 2011). Declarative memory refers to memory of facts and knowledge, which can be consciously recalled, also known as explicit memory (Anderson, 2013). Declarative memory further divides into semantic and episodic memory. Episodic memory is the memory of autobiographical events, such as times and places, and is a spatio-temporal record of a person's experiences (Anderson, 2013). Episodic memory is the memory system most affected by TBI (Armstrong & Morrow, 2010). However, given its idiosyncratic nature, computerized cognitive tests such as those used in the assessment of sport concussion, are unable to assess episodic memory. Therefore such assessments often test the other form of declarative memory, known as semantic memory, and this is often through word lists. Episodic memory can further be divided into non-verbal and verbal memory which can be presented visualy or auditory, and Morris (2010) recommends that all areas should be assessed in TBI.

Factor analysis studies of neuropsychological batteries are not consistent on whether or not memory should be treated as a single factor or that a division between visual and auditory/verbal or immediate and delayed should exist. A common cause for the inconsistency is perhaps due to the variety in measures employed in different studies. Factor analysis studies with the Weschler Memory Scale - Revised (WMS-R; Bowden, Carstairs, & Arthur, 1999), and the Wechsler Memory Scale, Thrid Edition (WMS-III; Bradley Burton, Ryan, Axelrod, Schellenberger, & Richards, 2003), a psychometrically strong measure, do not support a division between immediate and delayed memory, but indicate that visual and verbal/auditory memory should be considered separately. Conversely,

another factor analysis study that performed a principle component analysis on an extremely large neuropsychology battery, including the Wechsler Adult Intelligence Scale (WAIS) and the WMS in addition to several other memory tests, found only a General Memory factor (Leonberger, Nicks, Larrabee, & Goldfader, 1992). Yet both verbal and visuospatial factors were produced although not specific to memory (Ardila, Galeano, & Rosselli, 1998). Lastly, a study examining the psychosocial outcomes of head injury reported memory as a single factor in their outcome measures (Dikmen, Ross, Machamer, & Temkin, 1995) and Larrabee and Curtiss (1995) found that visual and verbal memory items load onto a single General Memory factor. These findings are thus inconclusive. They do however suggest that there is little distinction between immediate and delayed memory in TBI and that these should possibly be considered as part of a single factor. In contrast, verbal and visual memory could potentially represent two distinct constructs and require further investigation prior to being combined.

Memory impairment following concussion is often transient (McCrory et al., 2013) and can be seen in both visual and verbal modalities (Iverson et al., 2006; McClincy et al., 2006). Lezak (2004) suggests that memory impairment in mild TBI is generally a result of difficulty at the encoding stage rather than at the storage or retrieval stages. This is consistent with the hypothesis that slowed processing speed contributes to deficits in memory as slowed processing speed decreases the efficiency at which information can be encoded (Salthouse, 1996). This will be especially disadvantageous if materials are presented for a brief period of time, as occurs in the ImPACT memory subtests (ImPACT, 2013).

3.2.3 Executive functions

Executive functions refer to a number of self-regulatory functions that control and manage other cognitive processes, emotions, and behaviours (Lezak et al., 2004). These include the ability to initiate or inhibit behavior, inhibit competing actions or stimuli, select relevant task goals, plan and organize, solve complex problems, shift problem solving strategies, monitor and evaluate behavior, and regulate emotions (Morris, 2010). Their neural correlates are predominately within pre-frontal cortex that has projections to all other lobes in the brain, which allows it to co-ordinate the various lobes (Schoenberg & Scott, 2011). Individuals who sustain

concussion often report difficulties with everyday tasks that may reflect executive dysfunction, for example, difficulty in organizing and carrying out daily activities or difficulty anticipating consequences (Morris, 2010).

Standardised brief neuropsychological tests' ability to detect impairment in executive functions is somewhat limited (Morris, 2010). They lack sensitivity to impairment as they are most frequently administered in structured settings that fail to elicit the everyday difficulties commonly reported by TBI patients. For example, in Belanger and Vanderploeg's (2005) meta-analysis, effect sizes for the impact of concussion on executive functions failed to reach significance, indicating that either concussion does not influence executive functions or that the tests lacked sensitivity. Common standardised measures of executive functions, such as the Stroop Colour and Word Test and the Trail Making Test (Reitan & Wolfson, 1994) appear sensitive to general cerebral pathology but fail to demonstrate specificity regarding pre-frontal impairments, common in TBI (Cicerone, Levin, Malec, Stuss, & Whyte, 2006). However, studies that have compared TBI patients to non-injured controls have found the TBI group to be inferior at detecting and correcting errors made in everyday actions (Hart, Giovannetti, Montgomery, & Schwartz, 1998). Thus while it may be useful to include a measure of executive function in the assessment of concussion, it must be acknowledged that no one test will be sensitive to all possible forms of executive dysfunction given the multi-factorial structure of such functions (Bennett, Ong, & Ponsford, 2005; Chaytor & Schmitter-Edgecombe, 2007).

3.3 Summary

The CHC theory provides a taxonomy of cognitive functions for understanding and studying cognitive constructs, which are by nature inter-related and sometimes hard to separate. While the CHC theory provides a framework for investigating cognitive areas influenced by concussion, empirical research within the concussion literature is needed to identify the specific areas impacted by concussion. Based on the above review of concussion-induced cognitive impairment, it can be argued that in any neuropsychological assessment of concussion it is essential to include a measure of both simple and complex processing speed. Furthermore, memory is also an important area to assess.

Although there should be an assessment of immediate and delayed memory in addition to visual and verbal memory, these do not necessarily need to be separate constructs, especially in brief assessments when the number of tests able to be administered is limited. Lastly, though other areas such as executive functions (e.g., impulse control) may provide some information regarding the presence of concussion, brief forms of standardised tests lack sensitivity to everyday executive function impairment, and typically provide little additional information. This review of the literature provides a theoretical and empirical base upon which the ImPACT battery can be compared and evaluated.

Chapter 4: Reliability, Validity, and the ImPACT battery

The Immediate Post-concussion Assessment and Cognitive Test (ImPACT) has been demonstrated to be a reliable and valid instrument for assessing the cognitive effects of sport concussion (Iverson et al., 2006, 2006; Schatz, Kontos, & Elbin, 2012). This chapter describes and critiques the empirical research pertaining to the psychometric properties of the ImPACT battery in addition to explaining the concepts of reliability and validity within the context of evaluating a measure. The empirical evidence regarding the ImPACT batteries' reliability and validity is mixed. The developers of ImPACT claim it to be psychometrically robust and this is supported by several published papers (Elbin, Schatz, & Covassin, 2011b; Iverson, Lovell, & Collins, 2002; Schatz & Putz, 2006). Conversely, recent studies not supportive of ImPACTs reliablity and validity (e.g., Mayers & Redick, 2012a, 2012b) state that there is not enough evidence to support the current widespread use of ImPACT.

4.1 Test-retest reliability

Reliability is an instrument's ability to consistently produce similar results when applied under similar circumstances (Cronbach, 1988; DeCoster, 2005). Given that ImPACT's purpose is to track cognitive performance over time, investigations into the longitudinal reliability of the measure are pertinent to its credibility (Mayers & Redick, 2012b). Investigations have examined test-retest reliability as one route to infer ImPACT's stability over time. Test-retest reliability is the ability of an instrument to produce consistent results when the same entities are tested at two or more different time points (Field, 2009). The test-retest statistic reported for ImPACT has been either Pearson's *r* or an intraclass correlation coefficient (ICC). Pearson's *r* is a standardised measure of the strength of relationship between two variables. Possible scores range from -1 to +1; the sign in front of the value represents the direction of the relationship whereas the value indicates the strength of the relationship, with values closer to -/+1 indicating a stronger relationship (Rodgers & Nicewander, 1988). Intra-class correlations on the other hand, represent the consistency between the measures' items (Koch, 1982). They are believed to be a superior measure of retest

reliability as they are not biased by sample size, unlike Pearson's *r* (Wilk et al., 2002).

With regard to ImPACT, test-retest reliability studies tend to support the consistency of the measure over time. For instance, Iverson, Lovell, and Collins, (2003) reported Pearson's *r* correlations across a 13-day interval for healthy individuals, ranging in age from 15 to 22. The four main ImPACT composites demonstrated strong temporal reliability evidenced by correlations of .70 for Verbal Memory, .67 for Visual Memory, .86 for Processing Speed, and .79 for Reaction time. Schatz (2010) found similar reliabilities, in the form of ICCs, for Visual Memory, Processing Speed, and Reaction Time, which ranged from .65 to .74. However, Verbal Memory demonstrated weak reliability, evidenced by an ICC of .46. The test-retest period used was two years and reliabilities were calculated for a sample of 95 athletes.

These former studies employed the desktop version of ImPACT, whereas Elbin, Schatz, and Covassin (2011b) examined the test-retest reliability of the online version. They found comparable reliabilities (.62 Verbal Memory, .70 Visual Memory, .85 Processing Speed, .76 Reaction Time) over a period of 0.5 – 2.35 years (mean 1.2 years). Conversely, low reliabilities (ICC), ranging from .23 to .38 for the four composites were found for a test-retest period of approximately 45 days (Broglio, Ferrara, Macciocchi, Baumgartner, & Elliott, 2007). This study required participants to complete four distinct computerised cognitive assessment batteries consecutively at each testing interval. As a result, participants may have lost motivation or became fatigued, thus influencing results. Furthermore, whether or not alternative forms of the test were used at each testing interval could perhaps explain the variability in reliability results. For instance, Shuttleworth-Edwards and Radloff (2008) used alternative forms at pre- and post-season testing, minimising practice effects. They found no significant improvement in either the rugby or control group's performance. Conversely, Whitefield (2006) administered the exact same test at pre- and post-season and found a significant increase in the control groups performance on the Visual Motor Speed composite, most likely attributable to a practice effect. Lastly, studies investigating significant change between scores across testing intervals among non-concussed athletes found either no significant change (Miller, Adamson, Pink, & Sweet,

2007) or minimal significant change (i.e. < 5%), but no more than what is to be expected with variation in human performance over time (Elbin et al., 2011b; Schatz, 2010).

The test-retest approach to assessing the longitudinal reliability of a measure is limited as the correlations calculated include all item variance, that is, common variance, shared variance, and measurement error (Hambleton, Swaminathan, & Rogers, 1991). A correlation attempts to assess the strength of association based on the shared variance between two items (Field, 2009). This relationship can be distorted, as the computation is not based on the composites' shared variance alone, but also included is variance and error specific to each item associated with the composite. In order to get a true representation of the strength of the relationship between two constructs, one must separate out the specific variance and error. A *structural equation modelling* (SEM) technique known as *confirmatory factor analysis* (CFA) is able to separate the common variance shared by all items of the construct from the variance and error specific to each item (Brown, 2006). This allows the relationship between two constructs to be examined without distortion. This method is used in the current study and thus is discussed in depth in the Method Section 6.3.3.3.

4.2 Internal Consistency

Another form of reliability is internal consistency. It relates to the consistency or relatedness of items that make up a measure or subscale (Streiner, 2003). Given ImPACT is a multi-dimensional battery, the items within each subscale should be sufficiently related to one another, but distinct from those of other subscales. The internal consistency of a test structure can be assessed via *Cronbach's alpha*, *corrected item-total correlations*, and *factor analysis*. Cronbach's alpha is the most common statistic reported for internal consistency, but is sensitive to sample size and thus should not be viewed in isolation (Nunnally & Bernstein, 1991).

4.3 Content Validity

Content validity is concerned with the degree to which an instrument includes items that are representative of all aspects of the construct it is attempting to measure (Field, 2009). The content of a test includes wording, format, content,

and item themes. With regard to ImPACT, content validity would be present if all areas affected by concussion were sufficiently assessed by the included items, and enough items were included to adequately represent each construct. Furthermore, items that are not representative of the construct of interest should not be included. The domains affected by concussion and thus those that should be included in the scale should be informed by theory, empirical research, and/or expert opinion (Buckendahl & Plake, 2006). To the author's knowledge no published studies have specifically examined the content validity of the ImPACT measure.

4.4 Construct Validity

Construct validity is concerned with whether the instrument accurately measures the construct it purports to measure (Messick, 1995). If ImPACT could demonstrate that it is in fact measuring the cognitive domains it claims to measure, that is, Verbal Memory, Visual Memory, Reaction Time, Visual Processing Speed, and Impulse Control, then construct validity would be supported. Construct validity is often examined by looking at the relationship between the measure of interest with other instruments purporting to measure the same theoretical construct in addition to instruments purporting to measure distinct theoretical constructs (Messick, 1995). These techniques are referred to as convergent and divergent validity, respectively.

4.4.1 Divergent validity

Divergent validity is the degree to which a hypothesised construct differs to theoretically dissimilar constructs (Messick, 1995). It ensures that a substantial overlap with distinct constructs is not present. With regard to ImPACT, divergent validity between the composites has be investigated as opposed to between ImPACT and other independent measures. The composites should differ sufficiently to indicate that they are measuring distinct constructs, while still demonstrating slight association given they are all measures of cognitive functions and thus should be related. Studies investigating the divergent validity of ImPACT provide mixed support. Iverson, Lovell, and Collins (2002) reported that among a sample of 120 healthy high school and college athletes, weak correlations between the composites were obtained. As a result the authors

concluded the composites had minimal shared variance. However, Iverson and colleagues (2002) failed to report the numerical value of the correlations and they used a previous version (1.0) of ImPACT. Schatz and Putz (2006) also investigated divergent validity among healthy subjects. They found no significant correlations between the composites. However, this study used a small sample size ($N = 30$) and failed to report the values of all but one correlation. Despite the gaps in reported findings, these studies that have used healthy controls appear to support the presence of divergent validity among composites.

Research employing concussed athletes have concluded that ImPACT composites do not demonstrate divergent validity. Significant correlations between the four composites have been reported (Iverson, Franzen, Lovell, & Collins, 2003) and Iverson, Lovell, and Collins, (2005) suggest that Processing Speed and Reaction Time are measuring a similar underlying construct, based on factor analysis. It should be borne in mind however, that the composite scores are derived from the same six tests and all are measuring some form of cognitive function, thus some shared variance should be expected.

4.4.2 Convergent validity

Convergent validity on the other hand refers to the extent to which different measures assessing the same or similar theoretical construct are related (Brown, 2006). This is usually assessed via correlation. The few studies reported tend to support ImPACT's convergent validity. Iverson and colleagues (2003) demonstrated convergent validity among a sample of 25 concussed amateur athletes who completed ImPACT and traditional neuropsychological measures 20 days post-concussion. They reported medium, significant correlations for the Brief Visuospatial Memory Test total score ($r = 0.5$, $p < .05$) and delayed recall score ($r = .85$, $p < .05$) with the two ImPACT Memory composite scores. Furthermore, Trails A ($r = -.49$, $p < .05$) and Trails B ($r = -.60$, $p < .05$) were both negatively correlated with ImPACT Processing Speed. Iverson, Lovell and Collins (2005) reported that Symbol Digit Modality Test (SDMT), a measure of scanning and tracking aspects of attention and processing speed, correlated with Processing Speed and Reaction Time composites to a greater degree than with the Memory composites.

Furthermore, an exploratory factor analysis employing the four ImPACT composites and SDMT, found Processing Speed, Reaction Time and SDMT all loaded onto the same factor and thus appear to measure the same underlying construct. A more recent study found the Verbal Memory composite significantly correlated with traditional verbal memory tests, such as the Hopkins Verbal Learning Test-Revised total recall ($r = .27$, $p = .01$) and Digit Span ($r = .29$, $p = .00$). The Visual Memory composite was significantly related to the Brief Visuospatial Memory Test (BVMT) total recall ($r = .38$, $p = .00$; Allen & Gfeller, 2011).

4.4.3 Limitations of divergent and convergent validity

While convergent and divergent validity provide an indication of a measure's construct validity, they possess limitations (Messick, 1995). Firstly, both these techniques rely on having a 'gold standard' against which to compare item scores. A 'gold standard' is impossible given the improbability of an observed measure perfectly representing a theoretical construct.

Furthermore, both convergent and divergent validity are assessed via correlations. Correlations include all of the item's variance, common variance, item-specific variance, and measurement error (Field, 2009). Thus the strength of the relationship between the two items can be biased in either direction due to variance not associated with the construct of interest. For instance, the correlation of two items that are measured in the same way may be artificially inflated because their method of measurement is the same (i.e., method effect), whereas this erroneous inflation would not be present between items measured by distinct methods. For example, all three ImPACT reaction time items are measured in the same way (e.g., timed), however this method of measurement is different to that used for memory items (e.g., quantity of correct items, untimed). Lastly, if one measure possesses large measurement error and produces a significant correlation with another item, one cannot be sure if the association is due to the common variance or a result of measurement error (Messick, 1995).

4.4.4 Factorial validity

Factorial validity is another method of evaluating construct validity which is not affected by the aforementioned limitations (Brown, 2006). Factorial validity is concerned with the structure of the measure, that is, the pattern and magnitude of relationships between items and their associated factors (i.e., the construct they purport to measure). Unlike convergent and divergent validity which look at the relationship between two observed variables, factor analytic techniques consider the relationship between an observed item and a hypothesised construct (i.e., the latent factor). A structural equation modelling (SEM) technique, termed confirmatory factor analysis (CFA) is often used in the assessment of factorial validity. Confirmatory factor analysis allows the relationship between the latent construct and the observed item to be modelled while adjusting for error (Brown, 2006). Therefore the observed relationships are not biased by measurement error. Confirmatory factor analysis, in addition to exploratory forms of factor analysis, are commonly used in the development of measures to ensure unidimensionality of subscales and construct validity (Brown, 2006). It can also be used to assess the factor structure of an existing measure. For example, it is often used to ascertain whether the proposed factor structure upholds in different contexts (See Method section 6.3.3.3. for a detailed description of CFA).

With regard to ImPACT, factorial validity is important because a specific combination of items contribute to several overall composite (i.e., factor) scores. If factorial validity is not achieved, in that items hypothesised to contribute to a specific composite do not, the overall composite score and any subsequent interpretations would be misleading. For example, if one of the three items that contribute to the Verbal Memory composite is measuring a construct other than verbal memory and a low score on that item occurs, it could erroneously reduce the overall score to make an individual's verbal memory performance appear poorer than it is.

Whether or not factor analytic techniques were used in the development of the ImPACT battery is unknown. The author could not locate an explanation of the psychometric development of the scale in the current test manual (ImPACT, 2013) or empirical literature. Schatz and Maerlender (2013) indicated that factor

analytic studies were conducted on the earlier version of ImPACT (1.2) and were included in an earlier test manual. The earlier version of ImPACT did not contain the Visual Memory composite and the test manual is no longer accessible. One published article was found that detailed an initial exploration of ImPACT's factor structure through *principle component analysis* (PCA; Allen & Gfeller, 2011). The researchers employed a sample of 100 colleague students and identified a five-factor structure that did not mirror the item-factor relationships proposed by ImPACT developers. The factor structure they identified had several limitations. Firstly, one factor only had one indicator where, by definition, a factor must possess at least two indicators although three is preferred (Kline, 2011). Furthermore, negative loadings occurred for two items, one item demonstrated substantial (i.e., >.32) cross-loading, and two factors were defined by only two indicators each, all of which indicate the model is problematic. Negative loadings are problematic as they indicate the item is changing in the opposite direction of the latent factor, which should not be the case if the item is truly representative of the latent construct. Furthermore, if an item has only two indicators it is unlikely a true solution will be found as the factor is not sufficiently represented. Given the problems associated with this factor solution, in addition to the limitations of PCA (See Method section 6.3.3.1.), it is unlikely that this solution would hold up if one attempted to validate it among a distinct population. Two other factor analysis studies were identified in the literature; however both of these used composite scores as the first-order indicators. Composite scores represent a third level score, following second level subscale scores, and first-level raw scores. Factor analysis is designed to use raw scores as first-item indicators in the analysis and when appropriate subscale or parcel scores may be used as first-item indicators (Kline, 2011). Thus, using composite scores in a factor analysis is erroneous and any results should be interpreted with extreme caution. Notwithstanding, the two studies both found ImPACT to have an underlying two-factor structure. One factor represented Memory and was defined by the Verbal Memory and Visual Memory composites and the second factor represented Speed defined by the Visual Motor Speed and Reaction Time composites (Iverson et al., 2005; Schatz & Maerlender, 2013).

4.5 Criterion Validity

Criterion validity refers to a measure's ability to predict an outcome, or determine the presence or absence of an outcome based on item responses (Field, 2009). With regard to ImPACT, criterion validity is concerned with the battery's ability to distinguish between those who have sustained a concussion and those who have not. Two components of criterion validity are sensitivity and specificity. Sensitivity refers to the measure's ability to identify the presence of concussion when a concussion has occurred, whereas specificity refers to the measure's ability to determine the absence of concussion when in fact no concussion has occurred (Schatz, Pardini, Lovell, Collins, & Podell, 2006). A measure with strong criterion validity should demonstrate sensitivity and specificity values close to 100% (Field, 2009). The aim is to develop a test that is sensitive enough to detect the presence of concussion, yet specific enough not to conclude a concussion is present when in fact it is not (i.e., a false positive).

Studies examining the sensitivity and specificity of the ImPACT battery appear supportive. Schatz and colleagues (2006) reported ImPACT sensitivity as 81.9% and specificity as 89.4% among high school athletes. A more recent study found that among high school male football players a combination of symptom scores and the four main ImPACT composite scores produced higher sensitivity (65.22%) and specificity (80.36%) compared to symptoms or cognitive test scores individually (Lau, Collins, & Lovell, 2011). This study provided evidence for the utility of neuropsychological testing in the assessment of sports concussion as specificity increased by 24.41% when neuro-cognitive testing was used in addition to self-report symptoms scores. This is consistent with van Kempen and colleagues' (2006) study in which they evaluated the added utility of ImPACT relative to symptom monitoring alone among 122 concussed and 72 non-concussed high school and college athletes. They found ImPACT increased sensitivity by 19% (sensitivity = 92%) compared to symptom monitoring alone, however they did not comment on how this influenced specificity.

4.6 Cross-Cultural/Country Validity

Investigations of cross-cultural or cross-country validity of a measure are concerned with its applicability to a specific culture or country distinct to that in

which it was developed. Cultural differences in behaviour, language, and meaning have been well-documented in cross-cultural psychology literature (Berry, Poortinga, & Pandey, 1997; Nell, 2000; Segall, 1986). The validity of an instrument and its normative data only upholds to the original population (Brown, 1996). Therefore if a measure is to be used with a population distinct to the population it was developed in, it must first be validated within the population of interest. The New Zealand Code of Ethics for Psychologists (Code of Ethics Review Group, 2012) indicates that ethical practice involves the clinician ensuring the assessment measure used is evidence-based and the context in which it is being used is supported in the literature. Therefore, using the ImPACT measure within a New Zealand context or any other context where ImPACT has not yet been validated would be questionable practice and would potentially lead to erroneous interpretations. For instance, as a result of comparing an individual to another population's norms, it may appear they have recovered when in fact they have not and, as a consequence, put themselves at risk of serious injury by returning to play prematurely. Conversely it may appear a player has not recovered when in fact they have and thus they miss out on valuable game time.

Only a small number of studies have investigated the cross-cultural or cross-country validity of ImPACT. Review of the literature identified two studies that evaluated the equivalence of ImPACT scores across athletes of two countries. Tsushima and colleagues (2008) developed Hawaiian normative data based on 728 adolescent athletes ranging in age from 13 to 18. The authors visually compared the Hawaiian composite norms, means, and standard deviations with US normative data ($N = 424$) and concluded that the two samples performed similarly. No formal statistical comparison of difference was employed, rather the conclusion was based merely on visual analysis. The other cross-country comparison study located in the literature was that by Shuttleworth-Edwards and colleagues (2009). They compared white South African rugby players' ImPACT performance with that of age-matched US football players. Based on a series of independent t-tests computed for each composite, no significant differences between the groups were found. However, South African participants self-reported more concussive symptoms. It was concluded that US ImPACT

normative data were an appropriate comparison for English first-language South African athletes.

Two studies have looked at the cross-cultural validity of ImPACT within a country. Kontos, Elbin, Covassin, and Larson (2010) compared the performance of African American and White American high school and college athletes' baseline and post-injury performance on ImPACT. A series of Analysis of Variance (ANOVA) calculations revealed no significant differences between the groups at baseline and post-injury with the exception of Processing Speed post-injury. A second cross-cultural study reported by Horsman (2010) compared a sample of South African adolescents from an Afrikaans school to existing South African English normative data. Findings supported a general trend of poorer performance by the Afrikaans sample compared to the English sample. However the only statistically significant differences in performance, identified via t-tests, were found for Visual Memory in the 14 to 16 year age group and Reaction Time in the 17 to 21 year age group.

While these studies are useful in providing a platform for investigations of cross-country and cross-cultural comparisons of ImPACT, their conclusions of equivalence between two countries (e.g., Shuttleworth-Edwards et al., 2009) and two different cultures (e.g., Horsman, 2010; Kontos et al., 2010), are premature. These studies have established equivalence of test scores but any conclusions of equivalence beyond that, such as equivalence of the latent constructs the items purport to measure, would be misplaced. T-tests, as with other statistical methods used for group comparisons, assume that the instrument of measurement is operating in exactly the same way. That is, the ratio of change between the item and the latent factor is invariant and the underlying construct has the same theoretical structure and psychological meaning across the groups of interest. However, most studies, including those previously mentioned, fail to statistically test these assumptions prior to performing group comparisons.

A statistically sophisticated technique for looking at differences between groups is a structural equation modelling (SEM) technique which allows for the comparison of groups within a multiple group confirmatory factor analysis (CFA) framework. This technique is often referred to as measurement invariance. This is an

extension from the single-group CFA described previously in the factor validity section. It employs a multiple-group design that allows for the comparison of two distinct groups' latent variables with associated items and measurement error specified in the model. This method is discussed in detail in the method section. For now, however it is worth noting that it demonstrates advantages over t-tests and ANOVAs as it examines group differences while controlling for measurement error, thus providing greater accuracy. Additionally, it can identify the source of invariance, whether the invariance is at the item or construct level.

4.7 Summary

Several forms of validity and reliability must be demonstrated for a measure to be considered psychometrically robust, in that it provides consistent and accurate results that can be interpreted with confidence. Investigation into the ImPACT batteries reliability and validity is at a rudimentary stage. While attempts have been made to assess its psychometric properties, it appears some fundamental forms of validity have been overlooked. Given validity is cumulative, if one form of validity is omitted, confidence about which test scores can be interpreted is diminished (Buckendahl & Plake, 2006). Furthermore, the statistical methods used in the psychometric studies described in this chapter represent basic analyses which, although provide some useful information, need to be supported by more sophisticated analyses not subject to the same methodological limitations. Furthermore, if clinicians are to continue to use ImPACT in the assessment of sport concussions, they must first evaluate the validity and reliability of the measure within their specific population.

Chapter 5: The Current Study

Concussion and its potentially harmful consequences is becoming an increasingly popular topic within the sporting arena and neuropsychology. Our understanding of concussion, the nature of its effects, the neural correlates, and recovery patterns are still far from comprehensive (McCrory et al., 2013). However, due to increased interest in the topic of sport concussion, research in this area is quickly proliferating. At present, the general consensus regarding the assessment of concussion is that it should be multifaceted, including physical examination, self-report of symptoms, and an assessment of neurocognitive function and balance (Littleton & Guskiewicz, 2013; McCrory et al., 2013). A multifaceted approach to assessment is important given the large individual variation in both the presentation and recovery of concussion.

The use of neuropsychological tests in the assessment of sport concussion is a relatively new approach, with the first concussion programme to implement neuropsychological testing dating back to only 1990 (Lovell et al., 1996; Lovell., 1999). Since then, neuropsychological testing has become computerised, with the most popular battery being the Immediate Post-concussion Assessment and Cognitive test (ImPACT; ImPACT online, 2013). Based on the author's review of the literature, empirical support regarding ImPACT's psychometric properties is inconclusive, with important forms of validity and reliability yet to be assessed. This is worrying given that the utility of ImPACT in the assessment of concussion relies on the demonstration of sound psychometric properties. If psychometric properties are poor the clinician cannot be certain whether individual scores reflect true cognitive ability or are instead an artifact of measurement variance.

The overall aim of the current study is to extend the current psychometric literature regarding ImPACT by examining psychometric properties of ImPACT not yet assessed. Furthermore, structural equation modeling (SEM) techniques will be used, which overcome the methodological limitations of the more basic statistical analyses that have been previously used in the sport concussion literature. This will allow a more sophisticated and in-depth examination of ImPACT's psychometric properties.

Structural Equation Modelling is a theoretically driven form of statistical analysis. An important theoretical difference between SEM techniques and those currently used in the assessment of ImPACT's psychometric properties (e.g., t-tests, ANOVA) is the theoretical system each technique tests (Bollen & Lennox, 1991; Thompson & Green, 2006). The Latent Variable System, used by SEM techniques, assumes that the measured variable (i.e., the indicator) is a linear combination of an unobserved variable (i.e., the latent factor) and error (Thompson & Green, 2006). Thus variation among indicator scores is explained by variation in the factor plus error (Kline, 2011). Conversely, the Emergent System that is inherent in t-tests and ANOVA conceptualises measured variables as causal variables. That is, fluctuations observed among indicator scores explains fluctuations in the aggregated or latent factor score (Thompson & Green, 2006). While these more basic analyses assume an Emergent System, this is often inconsistent with the theoretical approach of the research in which it is used (Hancock & Samuelsen, 2008). ImPACT hypothesises that the items are indicators of unobservable (i.e., latent) cognitive functions (ImPACT online, 2013) and therefore a Latent Variable System is more appropriate. Another advantage of SEM is that it distinguishes between specific and unique variance in addition to measurement error, therefore providing a more accurate account of affairs.

Structural Equation Modelling techniques often require subjective decisions based on the available data, especially in the exploratory stages. Thus it is important that these decisions be guided by theory and are not simply based on the researcher's opinion. Despite a search of the literature regarding ImPACT, the ImPACT website, and technical manual, no theoretical framework was found. It is uncertain whether the construction of ImPACT was atheoretical or whether the theoretical framework used in ImPACT's development was never published. Either way, the lack of theoretical guidance resulted in the current study, choosing a cognitive theory from the literature to both guide the methodological process and provide a framework from which results could be interpreted. The Cattell-Horn-Carroll (CHC) theory was chosen given its robust support within the literature (Keith & Reynolds, 2010; McGrew, 2005, 2009). The Cattell-Horn-Carroll theory provides a taxonomy for understanding and studying cognitive constructs, which are by

nature inter-related and thus difficult to separate (McGrew, 1997). However, it is comprehensive and inclusive of all academic and cognitive abilities. Therefore the researcher also considered previous empirical research regarding the specific domains affected by concussion to guide decisions and provide a context for the current study.

The current study has three main objectives. These are individually described below along with a rationale for each objective and their potential contribution to the literature.

Objective 1A: To identify an underlying factor structure of the ImPACT battery.

Identifying and validating the underlying factor structure of ImPACT (i.e., factorial validity) will add to ImPACT's construct and content validity literature. Content validity refers to the degree to which ImPACT includes a sufficient number of items that are representative of the constructs affected by concussion and excludes irrelevant items. No studies were identified that specifically evaluated the content validity of ImPACT. As discussed in Chapter 3, a review of the literature indicates that any measure examining the cognitive effects of sport concussion should at a minimum include measures of processing speed and memory. Furthermore, both CHC theory and empirical research suggest complex processing speed and simple processing speed (i.e., reaction time) be assessed as separate constructs. However whether visual and verbal memory should be treated as separate constructs, or combined to form a General Memory factor, remains ambiguous based on the current literature.

Construct validity, that is, ImPACT's ability to measure the cognitive areas it purports to measure (Verbal Memory, Visual Memory, Reaction Time, Processing Speed, Impulse Control), has thus far been inferred from investigations of convergent and divergent validity (Iverson, Franzen, et al., 2003; Iverson et al., 2005; Iverson et al., 2002). As discussed in Chapter 4, finding evidence for these types of validity is limited by their analysis. The major limitation is that the correlation analysis is between items that include all of the items' variance and error. Thus the current study uses SEM techniques to assess construct validity in the form of factorial validity, which controls for error and distinguishes between

unique and common variance, and is thus not subject to the same limitation as simple correlation.

As discussed in Chapter 4, validity is cumulative. Thus, if one form of validity is lacking, conclusions based on other forms of other validity are potentially erroneous. Factorial validity is a fundamental form of validity. Empirical studies investigating the factorial validity of ImPACT are limited. Of note, factor analytic techniques for identifying the structure of item-factor relationships (i.e., factorial validity) are recommended to form part of the development of a scale (Brown, 2006). However, neither the psychometric nor the theoretical rationale or foundation employed to guide the development of the ImPACT battery is cited in the test manual (ImPACT, 2013) or the literature. Efforts to contact the developers resulted in no response. Thus it is unknown how the test developers concluded that five factors account for the variation in test items and also how they established the pattern of item-factor associations. Furthermore, upon examination of the item content, the researcher's opinion is that the model is mis-specified. Firstly, at face value two items (i.e., XO-total-correct-interference and XO-reaction-time) appear extremely similar and thus one is most likely redundant. Secondly, the Symbol Match – Memory item is loaded with the Verbal Memory factor when its content, at face value, appear representative of Visual Memory. A review of the literature revealed a small number of studies investigating the factor structure of ImPACT. However, they produced solutions inconsistent with the current scoring structure and are replete with methodological limitations. Thus, there is an obvious need for additional investigations of ImPACT's factor structure as empirical support for the current structure is lacking. Based on the above rationale the first objective of the current study is to explore the underlying factor structure of the ImPACT battery so as to identify what factors are present, and the relationship between those factors and observed variables.

Objective 1B: To determine whether the factor structure identified in the current study (1A) or Allen and Gfeller's (2011) structure, best fits the data.

To strengthen the confidence in the results from objective 1A, an attempt will be made to validate the identified factor structure via confirmatory factor analysis (CFA) among a distinct sample. Furthermore, Allen and Gfeller's (2011) factor

structure will also be submitted to CFA to identify whether their five-factor structure depicted in Figure 3, or the one identified in the current study (i.e., Objective 1A), best accounts for the variance in observed items. It is hypothesised that Allen and Gfeller's (2011) model will not provide an adequate fit to the data, but the factor structure identified in 1A will adequately fit the data. To the author's knowledge the current study will be the first to conduct a 'pure' exploratory factor analysis (EFA) in addition to a CFA using items from the ImPACT battery.

Objective 2: To assess the longitudinal stability of the identified ImPACT structure

Longitudinal stability refers to the consistency of an instrument's structure over time (Marsh, 1993). Although it is rarely assessed in psychological research, it is a fundamental aspect and implicit assumption of analyses that assess temporal change of a construct (e.g., test-retest reliability). When one investigates the stability of an overall instrument, this is referred to as measurement invariance. If measurement invariance over time is not empirically tested then observed temporal change of a construct may not be due to true change but may be the result of changes in the structure or measurement of the construct. Distinguishing between the potential causes of invariance is not possible in the absence of measurement invariance.

Given that a search of the literature failed to identify any study that investigated the measurement invariance of the ImPACT battery over time, conclusions regarding the temporal reliability of ImPACT constructs are potentially erroneous. Such conclusions were based on test-retest reliability, which calculated Pearson's *r*, a correlation co-efficient between each of the five composites with themselves at a different time point. In these studies (e.g., Elbin et al., 2011b; Iverson, Lovell, et al., 2003; Schatz, 2010) the temporal change or lack of it has been interpreted as a true score change of each construct (i.e., ImPACT composite), when in fact it may be that the structure of the construct or the precision of measurement may have changed over time. For this reason an assessment of measurement invariance should precede any other form of longitudinal analyses.

Thus the second objective of the current study was to strengthen the previous findings by investigating the longitudinal stability of the identified ImPACT structure. To improve upon previous methodological limitations in the literature regarding ImPACT's longitudinal reliability, SEM techniques will be used for the analysis to allow invariance to be tested for the true score of the construct, the meaning of the construct (i.e., the number of factors that represents the construct), and the measurement properties (i.e., item-level scores) of ImPACT, over time. Thus, measurement invariance in addition to longitudinal stability and differences in latent means will be assessed for the ImPACT structure over time.

Objective 3: To validate the identified ImPACT structure within a New Zealand adolescent sample and assess ImPACT's cross-country validity.

Brown (1996) states that the validity and reliability of an instrument only upholds to the population it was tested in or populations with demographically similar characteristics. One cannot assume that because a study has demonstrated adequate validity of ImPACT among a US population that it will be valid among samples from countries other than the US. To the author's knowledge investigations regarding ImPACT's psychometric properties have yet to be conducted with a New Zealand sample. There have been a few studies that have examined the cross-country validity of ImPACT, such as those described in Chapter 4, between the US and South Africa (Shuttleworth-Edwards et al., 2009), and the US and Hawaii (Tsushima, Oshiro, & Zimbra, 2008). However, like previous longitudinal investigations, these studies are limited by their methodology. Analyses used were t-tests and ANOVA, methods that assume the instrument of measurement is operating in exactly the same way and that the underlying construct has the same theoretical structure and psychological meaning across the groups of interest. Yet, this assumption is still to be empirically tested for the ImPACT battery. Another important limitation is that these methods do not account for measurement error, which has the potential to bias results.

Due to the above rationale the third objective is to attempt to validate the ImPACT structure identified in Objective 1 among a sample consisting of individuals from a different country (i.e., New Zealand) to the sample (i.e., South Africa) in which the structure was identified and initially validated. In doing so it

is hoped additional evidence for the ImPACT structure identified by Objective 1 will be provided, supporting its validity over other proposed structures.

Chapter 6: Method

The current study sought to assess the validity of the Immediate Post-concussion Assessment and Cognitive Test (ImPACT) among a sample of adolescent athletes. Stage One attempted to identify the underlying factor structure of ImPACT as developers have not specified the theoretical or psychometric foundation upon which its current scoring structure is based upon. Additionally independent explorations of ImPACT's factor structure are scarce. Therefore Stage One initially took an exploratory approach. A series of *exploratory factor analyses* (EFA) were conducted to identify a theoretically and psychometrically strong factor structure. The superior factor structure was then validated on a separate sample via confirmatory factor analysis (CFA). Stage Two sought to assess the longitudinal stability of the identified ImPACT structure across three time points. Structural stability, differential stability, and latent mean stability, across time were assessed. Lastly, Stage Three investigated the cross-country measurement invariance of the identified ImPACT structure across New Zealand and South African male adolescents.

6.1 Participants

An adolescent sample was chosen given the dearth of literature regarding adolescents and concussion, despite this age range exhibiting the highest prevalence of sport concussion. Furthermore, a New Zealand sample was chosen as the original aim of this research was to assess whether the ImPACT battery could be used with New Zealand adolescents as currently New Zealand high school rugby teams do not use neuropsychological assessment in the management of sport concussion, despite it being the recommended best practice (McCrory et al., 2009). The South African sample is one of convenience. Originally the aim was to compare a New Zealand sample to that of the American ImPACT normative data, however attempts to obtain American data were unsuccessful. Furthermore, given the difficulty recruiting New Zealand participants the South African sample was used as the primary data set for the EFA and two CFA's as it was a larger sample, a required feature of factor analysis. South African baseline data was obtained for three years, 2011, 2012, and 2013.

6.1.1 Sample A. Sample A consisted of 611 South African male sports-playing students, ranging from 12 to 19 years of age ($M = 15.02$, $SD = 1.42$). The 2012 sample was chosen for Stage One of the study, as it was the largest sample. Participants were English-speaking students, all from the same South African high school. Participants were excluded from the analysis if they had sustained more than two concussions and/or had a diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) and/or a learning disability, as these are known to affect the cognitive domains measured by ImPACT (ImPACT online, 2013). Consequently, a total of 89 participants were excluded; 24 had a history of three or more concussions, 65 reported a diagnosis of ADHD, 11 of which also reported a diagnosis of dyslexia. The sample was then randomly divided to form two independent samples and will be referred to as 2012A ($N = 271$) and 2012B ($N = 257$). The sample was split so both an EFA and CFA could be conducted. Demographic information is presented in Section 7.1.

6.1.2 Sample B. Sample B initially consisted of 130 sports-playing, male students from two New Zealand high schools. Independent t-tests indicated no significant differences in age or performance between the two school samples on individual items, therefore the samples were combined for all further analyses (See Appendix D, Table 3). Exclusion criteria as outlined previously for Sample A resulted in the exclusion of 13 cases: Four participants reported a diagnosis of ADHD, one indicated the presence of ADHD and Dyslexia, and seven had sustained three or more previous concussions. A further nine cases were removed as they were identified as multivariate outliers (see section 6.3.4.1.2). The final sample consisted of 109 males, ranging in age from 12 to 17 ($M = 14.5$, $SD = 1.91$). The final samples demographic information is presented Section 7.1.

6.2 Measure: Immediate Post-concussion Assessment and Cognitive Test (ImPACT)

ImPACT is a 30 minute computerised concussion evaluation system, utilised in the clinical management of sports concussion (ImPACTonline, 2013). It is widely used in several countries such as, the United States, South Africa, Sweden, and Canada. It is also used across ages and competition levels, including high school, college, and professional athletes (ImPACT online, 2012). ImPACT takes an

individualised approach to concussion management by recommending a baseline method of assessment. That is, athletes are tested pre-season to obtain a baseline measure of their cognitive performance, to which any post-injury testing can be compared (McCrory et al., 2005). Post-injury testing can be implemented numerous times until the athlete's post-concussion cognitive function returns to baseline, indicating the individual has recovered. This is in comparison to simply comparing post-injury performance to normative data. However, if for some reason a baseline measure is not available, ImPACT provides normative data for which the individuals' performance can be compared to. ImPACT has developed reliable change indices (RCI) for the five composite scores to identify when the difference between a baseline score and a post-injury score falls outside of normal score variation (ImPACT technical manual, 2012). Percentile scores are presented alongside the composite scores to provide additional information for interpretation as although a score may not exceed the RCI, the change from baseline may still be clinically significant. ImPACT is not intended to be used in isolation to diagnose concussion rather it provides additional information for a medical practitioner to consult when managing an individual's concussion and their return-to-play (ImPACT technical manual, 2012).

ImPACT consists of three main sections. The first pertains to demographic information, such as height, weight, age, sports code, history of concussion, and/or learning disorder (ImPACT, 2013). The second section consists of a subjective concussion symptom inventory. Participants indicate which of a possible 22 symptoms they are experiencing and indicate the severity of each on a 7-point Likert scale. A total symptom score is produced by adding the individual items. The third section is the neuropsychological tests which consists of six modules, Word-memory, Design-memory, X's and O's, Symbol Matching, Colour Match, and Three Letter Memory, which together produce five composite scores, Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. These tasks represent commonly employed neuropsychological tests. For example the Colour Match task is similar to the well known Stroop Colour and Word test (Stroop, 1935), and the Symbol Match task is similar to the Digit Symbol Coding in the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV: Wechsler, 2008).

The Word-memory and Design-memory modules are discrimination tasks in which 12 words/designs are briefly (750 milliseconds) presented on the screen. Presentation occurs twice to facilitate learning. The 12 target words or designs are then presented individually along with 12 non-targets words or designs and test-takers are asked whether or not each word or design was one of those originally presented. In the X and O's task a random assortment of X and O's are visually presented, three of which are highlighted yellow. The participant must remember the location of the yellow X and O's and following a distractor task recall their location. The distracter task is a choice reaction time test. On the screen either a red circle or blue square appears. If it is a blue square, the test-taker must click the left mouse button, if it is a red circle, then the right mouse button. Test-takers must respond as quickly as possible.

The Colour Match task is also a choice reaction time task and measures impulse control and response inhibition. It is similar to the well known Stroop task (Stroop, 1935) as a colour word is displayed in either the same colour ink as the word or in a different coloured ink as the word. The participant is instructed to click the word as quickly as possible, but only if it is presented in matching ink. Lastly, the Three Letter Memory task requires test-takers to remember three displayed consonant letters. Following 18 seconds of a distracter task they are asked to recall the three letters. For the distracter task they are presented with a grid consisting of numbers 1 to 25 randomly placed. They must click in descending order as quickly as possible.

From the six modules twelve subscale scores are automatically produced, which contribute to the five composite scores (ImPACT technical manual, 2012). The subscale scores are either averages or totals of the items that comprise each subscale. The total composite scores of Verbal Memory and Reaction Time are comprised of three subscale scores each, whereas Visual Memory, Visual Motor Speed, and Impulse Control composites are each comprised of two subscale scores. The composites are averages of their constitute subscales. Table 4 outlines the scoring pattern and provides a description of the subscale scores.

Table 4

ImPACT Scoring Structure

Composite	Item	Description
Verbal Memory	Word-memory	Percentage of correct number of words recalled correctly
	Symbol-match-memory	Percentage of correct responses when the symbols were hidden from participant
	Three-letters-memory	Percentage of correct number of letters correctly recalled
Visual Memory	Design-memory	Percentage of correct number of designs recalled correctly
	XO-memory	Percentage of correct number of X and O position recalled
Visual Motor Speed	XO-total-correct – interference	Number of correct responses to either the red circle or blue square
	Three-letters-correct - interference	Average amount of numbers counted in correct order
Reaction Time	XO- reaction-time	Average time to respond correctly during interference task
	Symbol-match-reaction-time	Average time to respond correctly when symbols visible
	Colour-match-reaction-time	Average time to respond correctly
Impulse Control	XO-total-incorrect	Number of incorrect responses to either the red circle or blue square (interference task)
	Colour-match-total-commission	Number of incorrect responses

6.3 Procedure

6.3.1 Ethics. Ethical approval was obtained through Massey University Human Ethics Committee Northern (12/075). If a participant's performance on ImPACT was below the expected range for their age then a senior clinical psychologist would review their results and provide the individual and their family with recommendations if needed.

6.3.2 Data Collection

6.3.2.1 Sample A. The researcher contacted the head researchers of the concussion programme at Rhodes University in South Africa, requesting permission to utilise a portion of their ImPACT data in the current study. Rhodes university manages a sport concussion programme for several South African schools. The programme involves baseline testing each year and post-injury assessment and management should an individual sustain a sport concussion. Written into the contracts between Rhodes University and the schools whom utilise their concussion programme is that the schools give them consent to utilise the data for research purposes. A contract between Rhodes University and Massey University stipulating the purposes of the data was drafted by the university's legal departments and signed by all parties. Subsequently, Rhodes University provided the researcher access to ImPACT data collected from one South African school dating back to 2009.

For the purposes of the current study only baseline testing from 2011, 2012 and 2013 were included in the sample. This was for two reasons, firstly previous years utilised a different version of ImPACT (2.0) and secondly, the same individuals were not tested every year, thus as more years were included the smaller the sample became. This data was collected in accordance with the ImPACT guidelines for testing. Administration occurred in groups and was overseen by a trained supervisor. Athletes completed the battery individually on a computer which automatically recorded individuals' responses, collated their scores, and produced a print out of each individuals subscale scores.

6.3.2.2 Sample B. Letters of invitation, as presented in Appendix A, were sent to approximately 30 high schools in the Auckland district of New Zealand, these included both uni-sex and co-educational high schools. The letter outlined the purpose and benefits of the research and the requirements of the school and students if they should wish to participate. If no response had been made within 10 working days a follow up phone call was made to the school. Of approximately 30 schools contacted, one male-only school agreed to participate in 2012 and one co-education school agreed to participate in 2013. The researcher and colleague met with the sports co-ordinator of each school to confirm the details of the study

and organise the schedule for testing. Information sheets were distributed to parents and consent was sought. Data collection for Sample B1 occurred over three sessions whereas collection of Sample B2 data occurred over eight sessions. Participants were administered the ImPACT battery in groups, ranging in size from three to 20. Participants were seated at a computer, separated from others by at least one chair space. The purpose of the research was explained and informed consent was sought (see Appendix B and C). Participants were instructed to follow the on-screen instructions, work at their own pace, and to raise their hand if they had any questions. The ImPACT programme automatically recorded all data. Sessions took approximately 40 minutes and were supervised by the primary researcher.

6.3.3 Data Analysis Methods

Descriptive statistics and exploratory factor analyses were computed in Statistical Package for the Social Sciences version 20 (SPSS Inc, 2011), whereas the various *confirmatory factor analyses* were conducted in Analysis of Moment Structures version 21 (AMOS-21; Arbuckle, 2012).

6.3.3.1 Exploratory factor analysis. Exploratory factor analysis (EFA) is a statistical technique used to identify the underlying structure of a set of observed variables (Tabachnick & Fidell, 2012). It is used when the structure of a set of variables is unknown, ambiguous, or lacking theoretical and/or empirical support. Items are entered into the analysis that has no a priori hypothesis regarding item-factor relationships. Any item may be associated with any factor. There are several methods of EFA and thus the researcher must make a series of decisions regarding the specific steps taken based on the data they are using (Tabachnick & Fidell, 2012). In the current study, *factor analysis* (FA) was chosen over the more commonly used method of *principal component analysis* (PCA) for several reasons. Firstly, FA only examines common variance compared to PCA which includes all item variance (i.e., specific and common) in its solution (Field, 2009). PCA is merely a data reduction exercise in that it attempts to explain the most amount of variance with the fewest components. Therefore components are simply aggregates of items, they are not representative of the variance shared between items, as is the case with factors. Another limitation of PCA compared to

factor analysis is that its findings cannot be generalised beyond the tested sample (Field, 2009). This is because in PCA it is assumed the sample is the population, whereas in FA it is assumed the sample is a random selection of the wider population and thus findings can be generalised beyond the current sample.

The first step in FA is the extraction of factors. Maximum Likelihood (ML) was the method of extraction chosen in the current study. ML is the recommended extraction process. It allows for the computation of several goodness of fit indices for the model, the statistical significance of factor loadings, and the correlation among factors (Field, 2009). The factor loadings produced by ML are those that maximise the likelihood of observing the actual data (i.e., correlation matrix). Three major assumptions of ML are that the data are continuous, normally distributed, and the sample size is large (Byrne, 2009). In regards to sample size, absolute guidelines suggest that a sample size of 100 is the minimum required (Kline, 2005; Russell, 2002), a sample size between 100 and 200 is 'medium' and anything over 200 is large. According to these guidelines the current sample (2012A, $N = 264$) is of sufficient size for the intended analyses. With respect to normality the sample demonstrated both univariate and multivariate normality (See Results Section 7.2). Therefore the ML assumptions were satisfied and thus it was an appropriate method to be used for the current data set to be used.

Following extraction of factors, rotation of those factors occurs, which results in items loading maximally onto one factor therefore making interpretation easier. Importantly, rotation does not change the original solution (i.e., the fit between the hypothesised model and the observed data) it simply makes it easier to interpret. There are two major types of rotation, oblique and orthogonal rotation. In the current study a form of oblique rotation was chosen, direct oblimin. Direct oblimin rotation aims to simplify factors by minimising cross-products of loadings. Oblique rotation was chosen over orthogonal as it allows factors to correlate. The latent factors in the current study represent cognitive domains that are believed to be related, thus oblique rotation was appropriate.

6.3.3.2 Structural equation modelling. *Structural equation modelling* (SEM) is an umbrella term which refers to a set of multivariate, statistical techniques which test theoretically driven, hypothesised relationships among

variables (Ullman & Bentler, 2012). The relationships between variables are assumed to be causal and each represents a regression path. Multiple regression paths constitute a 'model', which is a hypothesised structure of how the variables interact. The hypothesised model can be compared to the actual relationships among variables, observed in the data to assess their similarity (Byrne, 2009). Structural Equation Modelling techniques are an extension of General Linear Modelling procedures such as the Analysis of Variance (ANOVA) and Multiple Regression (Kline, 2011). However SEM techniques not only allow for the modelling of multiple independent and dependent variables, but also considers latent variables which may be represented by clusters of observed variables (Savalei & Bentler, 2010). Latent variables are constructs that cannot be directly measured, such as memory. Instead they are indirectly measured through several observed variables (i.e., items) believed to represent the construct. As the construct of interest is not measured directly, measurement error is most likely present. Measurement error represents variation in the observed scores that is not attributable to the latent variable. An advantage of SEM over ordinary regression models is that measurement error is taken into consideration when calculations are performed (Ullman & Bentler, 2012). Because SEM models measurement error separately to the indicator, it is controlled for and thus provides a superior reflection of the true relationship between two or more latent constructs (Kline, 2011).

6.3.3.3. Confirmatory factor analysis. Confirmatory factor analysis (CFA) is a type of SEM that is often used to assess the validity of a hypothesised factor structure (Kline, 2011). Factor analysis identifies a small set of unobserved variables which can account for the covariance among a larger set of observed variables (Brown, 2006). Within a CFA model, hypothesised relationships between items (observed variables or indicators) and latent factors (unobserved variables) are specified a priori, based on theory and/or empirical evidence. CFA is based on the Common Factor Model which postulates that each indicator is a linear function of one or more common factors and one unique factor (Harrington, 2008). Common factors are unobserved (latent) variables that are defined based on observed variables (Brown, 2006). Thus, variation in the latent factors explains a proportion of the variations seen in the corresponding indicators, not the other

way around. The residual variation is unique variance, a combination of variance specific to the indicator in addition to random/measurement error (Brown, 2006). A CFA model always consists of a measurement model which includes a latent factor, several indicators (observed variables), and their associated measurement error (Byrne, 2001). Confirmatory Factor Analysis provides estimates of the magnitude and direction of the relationship between observed variables (items) and latent variables (factors), these are termed factor loadings or regression weights (Kline, 2011). Factor loadings represent how much of the variance in the indicator is explained by the latent factor. When there is more than one latent variable there is also a structural model which consists of the relationships between latent variables, the magnitude of which is protrayed through covariance estimates.

The aim of CFA is to assess how well the hypothesised model matches the covariance matrix of the data set (Brown, 2006). That is, how well the estimated parameters of the factor solution reproduce the observed relationship among the input variables. To assess this, the estimated covariance matrix is compared to the observed covariance matrix and a test statistic of difference is calculated (Tabachnick & Fidell, 2013). Since it is common for more than one model of a particular measure or construct to be specified in the literature, CFA can test which of the competing models provides a superior explanation for the observed items variance and covariance. The model consists of several equations, some of which the solutions are known and some unknown (Kline, 2011). The unknown equations are solved via an estimation process. The current study employed Maximum Likelihood (ML) estimation given it is the most commonly cited method in the literature (Kline, 2011, Tabachnick and Fidell, 2013), and importantly, it was most appropriate for the current data set. The previously mentioned, ML assumptions were also met by the data used in the series of CFA's. Specifically, the sample size (2012B, $N =257$) was sufficiently large, the data were continuous and were normally distributed.

6.3.3.4 Item parcelling.
6.3.3.4 Item parcelling. The current study used parcelled items as the lowest-order indicator variables in the CFA model. Parcels result from summing or averaging two or more items scores (Meade & Kroustalis, 2006). These are used in place of individual raw item scores in the analysis. In the current study

parcelled items were used, as raw data were not made available from the ImPACT company. The ImPACT programme automatically parcels items by averaging the subscale score; this is what test administers are provided with. The use of parcelled items or subscale scores appears to be common practice in factor analytic studies of neuropsychological batteries (Deary, 1993; Larrabee, Kane, & Schuck, 1983; Parker, 1983). In other research areas it is also not uncommon as highlighted by Bandalos and Finney (2001). In their review of 317 SEM studies conducted between 1989 and 2004, they found that 19.6% employed a form of parcelling and of these, 82.3% were CFA studies (Bandalos & Finney, 2001). Nevertheless, the use of parcelled items is controversial (Cattell & Burdsal, 1975; Little, Cunningham, Shahar, & Widaman, 2002; Nasser & Takahashi, 2003; Nasser & Wisenbaker, 2003; West, Finch, & Curran, 1995).

The main criticism of parcelling is its threat to validity. Parcelling compresses the specific variance and random error of each individual item, either eliminating it all together or at least reducing it. The parcelled items typically share variance, and this is what is emphasised in the aggregated score. As a result model fit is improved and mis-specification can be obscured (Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994). For example an item could originally load onto both a Memory and Reaction Time factor, if it is parcelled with other predominately memory loaded items, the aggregate or parcelled score may now only load on Memory and not Reaction Time. Thus, its association with the Reaction Time factor is concealed and the model is mis-specified.

In the current study, furthermore to having no other option than to use parcelled items, parcelling items was advantageous. ImPACT raw scores are dichotomous, and thus given the limited range of scores, they lack variance and normality is impossible. Parcelling the items increased the range of possible scores, creating continuous data with the potential for normality. Other advantages of parcelling are that the number of items to sample size decreased (Bagozzi & Edwards, 1998) and therefore parcelled items are more reliable than individual items (Cattell & Burdsal, 1975; Kishton & Widaman, 1994).

Guidelines regarding item parcelling suggest that if items are unidimensional then parcelling is acceptable (Kline, 2011). Whilst this assumption cannot be

statistically tested in the current study given the unavailability of raw scores, it is probable given the nature of the items. Within each subscale of ImPACT, the item (i.e., question or task) is predominantly the same. For example, in the Word-memory subscale each item is "Did the word _____ previously appear?". Whilst the specific word is different for each item the question is the same. In regards to the reaction time tests the items which make up each subscale or parcel are exactly the same, they are simply different trials of the same test. This is the typical format for tests of reaction time. Lastly, given the similar if not identical nature of the items within each parcel it is also likely that the specific variance and random error are also similar. However these are only hypotheses that could not be tested since raw data was not available.

6.3.3.5 Evaluation of model fit. An evaluation of how well a hypothesised model fits the data should be inclusive of the overall fit evidenced by goodness-of-fit statistics, localized areas of misfit, and the size and statistical significance of parameter estimates (Brown, 2006). Adequate fit of the overall model provides support for the validity of parameter estimates. Goodness-of-fit statistics provide a global measure of the models ability to reproduce the input covariance matrix, whereas localised strain and parameter estimates provide more specific information about the acceptability and utility of the solution (Byrne, 2001).

Several goodness-of-fit statistics exist for evaluating model fit (Brown, 2006). Some statistics are absolute in their assessment of model fit whereas others provide a continuous measure of model-data correspondence. Various test statistics have different strengths and weaknesses, thus several indices will be considered collectively as guides for evaluating model fit (Byrne, 2001). The following fit indices were chosen as they are widely accepted and were appropriate to use with the data of the present study.

6.3.3.5.1 Chi-square statistic. The chi-square (χ^2), which is absolute in nature, is the classic fit statistic (Kline, 2011). It tests the similarity between the implied and the observed covariance matrix, also known as the exact fit hypothesis (Kline, 2011). If the chi-square associated p-value is statically significant ($p < 0.05$) then the null hypothesis of exact fit is rejected (Raykov & Marcoulides, 2006). When a model provides a good fit to the data, the chi-square

value is lowered. However the chi-square statistic is inflated by large sample sizes and in the presence of a large sample, even minor differences between covariance matrices can produce a significant result (Cheung & Rensvold, 2002). A further limitation cited in the literature is that the standard of 'perfect fit' is not practically realistic in social science research (Kline, 2011). Given these limitations less emphasis will be placed on the chi-square statistic and it will be interpreted in conjunction with more practical fit indexes outlined below.

6.3.3.5.2 Comparative fit index and Tucker-Lewis index. Both the comparative fit index (CFI) and Tucker-Lewis index (TLI) are comparative fit indices as they compare the difference between the hypothesised model and an independence model in which all observed variables are uncorrelated (Kline, 2011). Their values indicate the ratio of improvement of fit of the observed model over the independence model. Values vary from 0 to 1, with those exceeding .90 indicative of adequate fit (Bentler & Bonett, 1980) and those exceeding .95 indicative of good fit (Hu & Bentler, 1999). Unlike the chi-square statistic the CFI and TLI are not influenced by sample size (Brown, 2006).

6.3.3.5.3 Root mean square error of approximation (RMSEA). The Root Mean Square Error of Approximation (RMSEA) is representative of the error of approximation within the population (Byrne, 2009). It examines the extent to which the hypothesised model fits the data "reasonably" well, opposed to the 'exact' fit as assessed by the chi-square likelihood ratio test (Brown, 2006). A RMSEA of less than .05 is considered a good fit, .05-.08, a reasonable fit, .08 to .10, a mediocre fit, and values greater than 0.10, a poor fit (Hu & Bentler, 1999). The RMSEA incorporates a penalty for poor model parsimony and is sensitive to misspecification. Confidence intervals can be calculated around RMSEA values providing an indication of the precision of RMSEA estimates (Byrne, 2009).

6.3.3.5.4 Modification indices. Whilst the fit statistics previously described are useful in assessing model fit, they should not be used in isolation (Kline, 2011). Goodness-of-fit statistics summarise the overall fit of a complete model in a single number. The model could contain a severe misspecification in one of the groups or for one specific parameter, but still have a reasonable overall fit (Kline, 2011). Therefore in addition to evaluating overall fit, it is equally important to

examine the model for localised areas of poor fit. Modification indices identify potential areas of local statistical poor fit (Brown, 2006). Modification indices are calculated if AMOS (Arbuckle, 2012) believes the addition of certain parameters would improve the fit of the model. That is, the data are suggestive of certain relationships that aren't currently specified in the model. A modification index indicates how much the chi-square value of a model would decrease and model fit improve, if the parameter were free instead of constrained (Brown, 2006). Modification indices are in fact chi-square tests for individual equality constraints; high values indicate that the respective parameter constraint is 'wrong'. However, given the chi-square statistic is sensitive to large sample size, only extremely high values, such as those exceeding 20, should be taken as strong evidence of poor fit. Because the chi-square values can be misleading, it is advisable to look at the *expected parameter change* (EPC) value as well (Kline, 2011). The parameter change indicates the extent to which a parameter would change if the equality constraint was removed. Parameter changes greater than .10 are considered substantial (Byrne, 2009). A caution regarding modification indices is that they are numerically driven, they are not theoretically based. Given the current study employs CFA, a theoretically driven technique, modification indices will not be heavily relied upon.

6.3.3.5.5 Standardised residuals. Lastly, standardised residuals can also provide information regarding model fit. In CFA the residuals represent the discrepancy between the hypothesised covariance matrix and the observed covariance matrix (Kline, 2011). Standardised residuals are not dependent on the scale of measurement as they represent an estimate of the number of standard deviations the observed residuals are from a residual of zero that would exist if the model were perfect (Brown, 2006). In large samples and a correctly specified model, standardised residuals follow a normal distribution, thus the majority should be less than two standard deviations. Byrne (2009) suggests the *z*-score values above 2.58 are considered large and may indicate areas of misspecification. This number is chosen as standardised residuals can be loosely interpreted as *z*-scores and a value that is equal to, or greater than 2.58 is two standard deviations away from the mean and represents a statistically significant value at $p < .01$.

6.3.3.6 Measurement invariance within a CFA framework. Measurement invariance is concerned with whether or not a measurement scale performs similarly across groups among similar circumstances (Kline, 2011). More specifically, whether the items (observed variables) are perceived in the same manner and thus represent the same underlying construct(s) across various groups. This is evidenced by equivalent relationships between the observed variables and latent variables across groups. Measurement invariance is an assumption of all group comparisons (e.g., t-tests, ANOVA) whether it is statistically tested or not (Vandenberg & Lance, 2000). In order to compare groups of individuals on their level of a construct one typically assumes that the measurement scale employed is measuring that same construct across groups (Schmitt & Kuljanin, 2008; Vandenberg, 2002). If this is not the case then differences in observed means may not reflect true difference of the construct, but rather, result from items measuring different constructs in each group (Meredith, 1993). Thus, for test scores to be comparable across distinct populations measurement invariance must be present. Test items must have invariant, quantitative relationships with the latent variable for each population of interest (Meredith, 1993). In regards to the current study, whether the items of the ImPACT battery were measuring the same constructs across time and across countries (New Zealand and South Africa) was tested.

The current study followed Meredith's (1993) hierarchy of invariance testing and thus configural invariance, weak factorial invariance, and strong factorial invariance were progressively assessed. This method involves testing increasingly stringent models and each level must be achieved for the subsequent model to be meaningful. The models are nested models in that they are progressions of one another. Essentially they are the same model with additional parameter constraints. Therefore, they are able to be statistically compared to one another to assess difference in fit. At each progression through the invariance hierarchy fit will decrease due to the increase in fixed parameters. If progression through each nested model does not result in significant change of model fit, then invariance is achieved.

Configural invariance is the lowest form of invariance and is concerned with the equivalence of a measures structural pattern across groups (Meredith, 1993; Widaman & Reise, 1997). More specifically, whether the structure of a measure

has the equivalent number of items and factors and that the same items load onto the same latent variables in each group. Configural invariance requires no equality constraints. It is achieved when the identical models are run simultaneously and adequate fit, evidenced by goodness-of-fit indices, is present (Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000).

Weak factorial invariance, also known as metric invariance is a higher level of invariance than configural (Selig, Card, & Little, 2008). It examines whether factor loadings, which represent the magnitude of item-factor relationships are equal across groups. Meredith (1993) defined weak factorial invariance, as existing when, for every value of the latent variable, the associated mean and variance of the indicator are independent of group membership. Metric invariance is tested by constraining factor loadings equal across groups, if the model adequately fits the data and the decrement in fit is not statistically significant then weak factorial invariance is achieved (Milfont & Fischer, 2010; Vandenberg & Lance, 2000).

Configural and weak factorial invariance are based on covariance structures, thus only parameters representing regression coefficients, variance, and covariance have been of interest thus far (Sousa & Chen, 2003). Covariance provides information regarding how items vary in relation to their latent structures, and these can be compared across groups. However the equivalence of covariance structures is not a sufficient predecessor for the comparison of latent means. If latent means are to be compared, a higher level of invariance, known as strong factorial (i.e., scalar) invariance, is needed (Cheung & Rensvold, 2002). In the analysis of covariance structures, observed variables are measured as deviations from their means of zero, thus intercept terms associated with the regression equations are irrelevant (Milfont & Fischer, 2010). In the analysis of mean structures however, item intercepts are the expected item scores for respondents that have a zero score on the latent variable. The invariance of item intercepts implies that all observed mean differences are a result of differences in the latent factor and are not a result of differences in item functioning or meaning across countries or over time. Strong factorial invariance is assessed by placing equality constraints on both the factor loadings and item intercepts (Meredith, 1993;

Widaman & Reise, 1997). This is a strict form of invariance and is seldom achieved. (Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012).

If at any level of invariance, measurement invariance for the full model is not achieved it is possible to test for partial invariance (Meredith & Teresi, 2006; Schmitt & Kuljanin, 2008). If partial invariance is achieved then Byrne (1989) suggests that it is appropriate to proceed with subsequent invariance testing. Kline (2011) suggests systematically freeing non-invariant parameters, guided by modification indices as they identify the items or intercepts that display the largest group differences. Each time an equality constraint is removed the model should be re-run and assessed for fit. For a model to be partially invariance there must be at least two invariant items per factor, including the marker item (Byrne, Shavelson, & Muthén, 1989; Vandenberg, 2002).

6.3.3.6.1. Assessing measurement invariance. Significant change between nested models was assessed via the chi-square difference test and the observed change in CFI estimates. The chi-square difference test evaluates the exact fit hypothesis for two hierarchical models (Kline, 2011). That is, it compares the model chi square statistic of the two hierarchical models and indicates whether the decrement in fit is statistically significant. The degrees of freedom of the chi-square difference test are the difference between the two nested models degrees of freedom (Kline, 2011). If the difference between the chi-square statistics of the two nested models is greater than, or equal to the critical value, found in the chi-square distribution table, then there is a statistically significant difference between the two models and measurement invariance is not present. As with the model chi-square the chi-square difference test is too affected by sample size. This means that in large samples, the chi-square difference statistic could be statistically significant, even though the absolute differences in parameter estimates are minimal. That is, the chi-square difference may imply lack of measurement invariance when the imposition of equality constraints across groups makes little difference to model fit. For this reason Cheung and Rensvold (2002) questioned the sole use of the chi square difference test. Cheung and Rensvold (2002) studied the characteristic of changes in the values of 20 different approximate fit indexes when invariance constraints were added. This was the first empirical attempt to investigate the sensitivity of fit indices to lack of measurement invariance. They

found that most were affected by model size and complexity, with the exception of the CFI (Bentler, 1990). Cheung and Rensvold (2002) suggest changes in CFI of .01 or below are indicative of measurement invariance. In a later study also employing the Monte Carlo method, Chen (2007) agreeably found the CFI to be sensitive to lack of invariance, specifically invariance of factor loadings, intercepts, and residual variances. Meade and colleagues (2008) questioned Cheung and Rensvold's cut-off of .01, stating that it was too strict for use with real-world data. Cheung and Rensvold (2002) stimulation studies used perfectly invariant data, an unlikely occurrence in the real world. Meade and colleagues (2008) suggest a less strict CFI change cut-off of .02 based on their stimulation studies which also used data known to be invariant, however not perfectly invariant. Meade and colleagues (2008) criteria for measurement invariance is used in the current study.

6.3.3.7 Latent means analysis. Means of latent variables cannot be directly measured. They must be inferred from items that can be directly measured and which are assumed to be representative of the latent variable. The analysis of latent means is based on both co-variance and mean structures given the inclusion of intercepts (Little, 1997; Milfont & Fischer, 2010). Estimation of intercepts is needed as they are used in the calculation of item means. Unlike the analysis of only covariance structures, item means take on a non-zero value in order to estimate difference in latent means (Meredith & Teresi, 2006). Because item means are functions of other parameters in the model (i.e. variance parameters), the intercept must be estimated jointly with all other model parameters (Jöreskog & Yang, 1996). In the analysis of covariance structures, item means are ignored as items are treated as deviations about their means of zero (i.e. variances).

Latent means are estimated within the scalar model. In this model, item intercepts are constrained equal across groups. As a result the factor intercepts have no definite origin, they are statistically undefined (Kline, 2011). Therefore the factor intercepts of one group, arbitrarily defined as the first group, must be fixed to zero to define the origin of the scale. However the factor intercepts of subsequent groups can be freely estimated. Thus, latent means are only interpretable in a relative sense, in that it can be detected if they differ between groups, but the

numerical mean for each factor of each group cannot be estimated (Steenkamp & Baumgartner, 1998). Thus the produced mean of a specific factor is the relative difference between itself and the factor that had its mean constrained to zero.

6.3.4 Data analysis procedure

6.3.4.1 Data preparation

6.3.4.1.1 Sample size. The accuracy of fit indices and parameter estimates in SEM is dependent on having an adequate sample size to generate the level of power required for the analysis (Kline, 2011). Generally the sample size needed increases as the complexity of model increases (MacCallum, Widaman, Zhang, & Hong, 1999). Smaller sample sizes in SEM tend to result in more pessimistic estimates of goodness-of-fit when compared to larger samples (Fan, Thompson, & Wang, 1999). Absolute guidelines suggest a sample size of 100 is the minimum required for CFA (Kline, 2011; Russell, 2002), between 100 and 200 is 'medium' and anything over 200 is large. Sample sizes in the current study ranged from 109 to 264, therefore all were of sufficient size for factor analysis.

6.3.4.1.2 Normality and outliers. Normality of each sample was assessed by examining skew and kurtosis values for each individual item. In line with West and colleagues (1995) guidelines, an item was considered to have an acceptable normal distribution if its skew value did not exceed two and kurtosis value did not exceed seven.

Multivariate normality and outliers were also assessed. Mahalanobis Distance (*D*) was used to identify potential multivariate outliers. It provides a relative measure of a data points distance from the median (Kline, 2011). Whilst there is no cut-off point to indicate a case is an outlier, Arbuckle & Wothke, (1999) suggest that cases should be considered as potential outliers if their *D* is well separated from other *D*'s. Multivariate kurtosis and thus normality was indicated by Mardia's (1970) co-efficient. According to Bollen (1989), multivariate normality is achieved if Mardia's coefficient is lower than p(p+2), where p is the number of observed variables.

6.3.4.2 Stage One. Initially a correlation matrix was computed utilising Sample 2012A including all twelve items. Factorability of the matrix was assessed

against several standards including the magnitude of correlations, Bartlett's (1954, as cited in Tabachnick & Fidell, 2012) test of sphericity, and Kaisers (1970, cited in Tabachnick & Fidell, 2012) measure of sampling adequacy. In order for a correlation matrix to be factorable it must contain several correlations that exceed .30, suggesting a number of items are sufficiently related to warrant factor analysis. However correlations exceeding .90 are undesirable as one of the items may be redundant given the similar nature of the two involved. Bartletts test of sphericity tests the null hypothesis that the off-diagonal correlations in the matrix are equal to zero. Thus if it is significant, the null is rejected, suggesting the correlations are large enough to proceed with factor analysis. Of note Bartletts test of sphericity is dependent on sample size, therefore in large samples it is likely to be significant even if correlations are low. Given the current sample uses relatively large samples it will be cautiously interpreted. Kaiser's measure of sampling (KMO) adequacy is the average degree of interrelatedness over all variables. The KMO statistic ranges between zero and one. Kaiser suggests a value .50 is the absolute minimum however values above .60 are desired (Tabachnick and Fidell 2013). The KMO can be calculated for the overall scale and for items individually.

Given the suitability of the data for factor analysis based on the above criterion, the twelve items were initially submitted to a *maximum likelihood* (ML) EFA with direct oblimin rotation. No limit was specified for the number of factors to be retained, instead factors were retained if their eigenvalues exceeded 1. An eigenvalue represents variance, therefore the greater the eigenvalue of a factor the greater amount of item variance is explained by that factor. An individual item's variance is equal to 1. Factors are only retained if their eigenvalues exceed 1, as it would explain more variance than one item alone.

Following rotation the magnitude of communalities and factor loadings were assessed. Factor loadings represent the correlation between the observed variable and the factor. Tabachnick and Fidell (2013) suggest only factor loadings that exceed .32 should be interpreted. Loadings above .80 are considered very high, those between .60 and .80 are high, those between .40 and .60 are moderate, and those below .40 are low. Communalities are squared correlations and represent the proportion of common variance possessed by a variable. Possible values range

from 0 to 1. Zero indicates the item has no common variance with other variables in the measure, suggesting all the items variance is item-specific and one indicates all variance is common variance. However values of 1 are problematic and are referred to as Heywood cases. The presence of Heywood cases suggest that there is a problem with the specification of the model (Tabachnick and Fidell 2013).

The first EFA's solution was problematic for reasons outlined in the results section 7.2.1. The factor analysis was re-run with ten items. The KMO and Bartletts test of sphercity were inspected for adequate factorability. The identified factors (i.e., subscales) internal consistency was assessed via Cronbach's alpha. Cronbach's alpha represents the level of consistency of responses across scale items (Nunnally, 1991). Values above .50 are acceptable yet poor, those above .70 are adequate, and above .80 are good (Kline, 2011). *Corrected item-total correlations* were also examined. They represent the correlation between an item with the total scale score, minus the value of the item the correlation is with. The item of interest is removed from the total score as if it were included it would inflate the correlation as a component of the total correlation would be the correlation of the item of interest with itself. Thus the corrected item-total correlation represents the strength of the relationship between the item of interest and the collective total of the other items in the scale (IBM, 2006). High correlations are desired as it is expected that if an individual performs well on one item they should also perform well on other items of a similar nature. Negative corrected item-total correlations are problematic as it is unlikely that an individual will perform poorly on one item but highly on another, which are both measuring the same construct. If the scale is realible all items should produce correlations of at least .30 with the total (Field, 2006). Lastly, *Cronbach's alpha if item deleted* was also reviewed. If this is significantly larger than the total items alpha, when a specific item is deleted then that item may not be sufficiently related to the other items within the scale.

The Processing Speed/Reaction Time factor's internal consistency was extremely low. Two items, Three-letters-average-correct and XO-total-correct were deleted. These two items represent complex processing speed according to ImPACT developers and thus are qualitatively distinct from the other items in the scale which are representative of reaction time (simple processing speed). This is

consistent with the CHC theory as they too distinguish between complex and simple (reaction time) processing speed.

The EFA was re-run with only eight items. Again the KMO and Bartletts test of Spherity were assessed for suitable factorability. Furthermore factor loadings were checked for suitable loading onto corresponding factors, and lastly the internal consistency of the Reaction Time subscale was assessed via Cronbach's alpha, in addition to corrected item-total correlations and alpha if item-deleted. The resulting solution was acceptable.

The second part of Stage One attempted to validate the hypothesised model identified in the previous EFA (i.e., Figure 2) in addition to that proposed by Allen and Gfeller (2011) via CFA. The second half of the 2012 sample (2012B, N = 257) was used in the computation. Allen and Gfeller's (2011) model was identified via a principal component analysis and purported that eleven items could be explained by five underlying dimensions. In their Model one factor was defined by only one indicator. In the current study this factor and hence indicator was removed, given that a minimum of two, preferably three indicators is needed to form a factor (Brown, 2006). CFA models were computed in AMOS graphics and were depicted using SEM notation. A latent variable was symbolised by an oval, an indicator as a rectangle, and error as a small circle. A single headed arrow represents causality, with factors causing indicators as hypothesised in a latent variable system. A curved, double headed arrow indicates a reciprocal relationship. Model fit was assessed via goodness-of-fit indices, modification indices, and standardised residuals. Given adequate fit statistics and thus validation of the model, factor loadings were assessed to ensure they were associated with their purported factors.
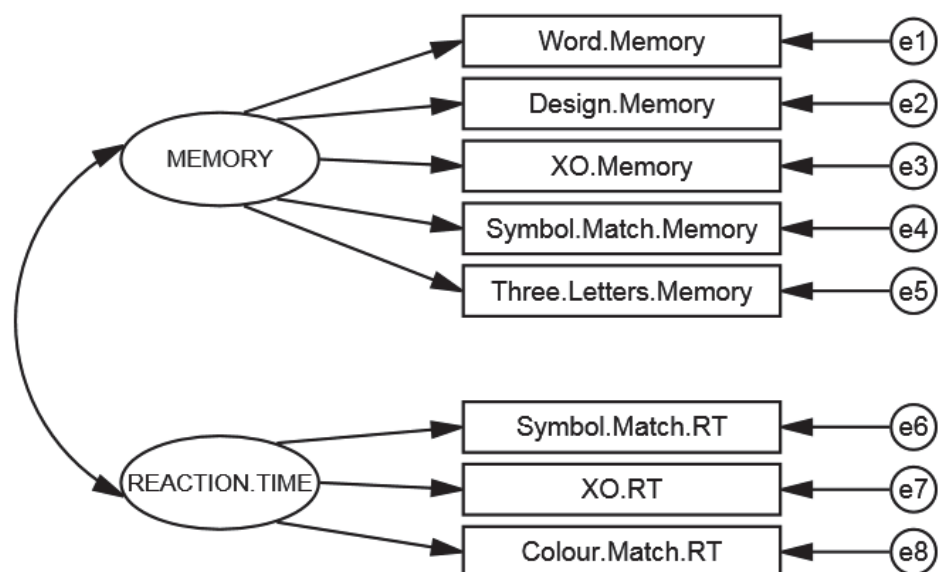
Figure 2. Model A: ImPACT model identified in current study.
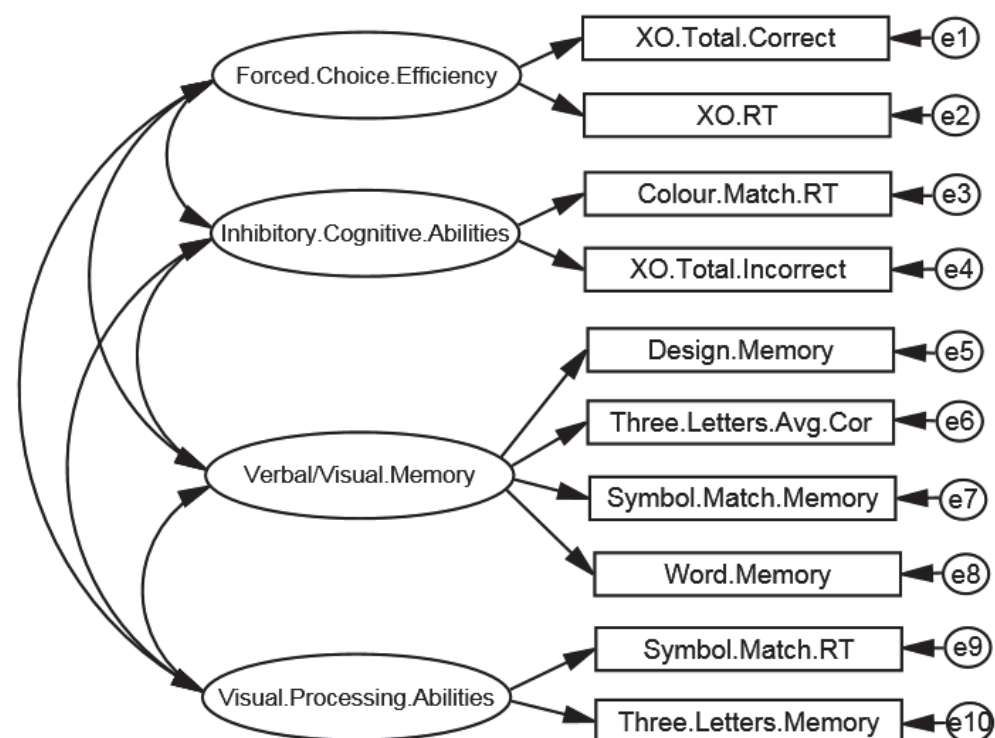
Figure 3. Model B: ImPACT model identified by Allen and Gfeller (2011).

6.3.4.3 Stage Two: Longitudinal stability. Longitudinal stability was assessed within a single factor repeated measures model. Three time points were included, 2011, 2012, and 2013 in the one model. Data from the 2011 (N = 595), 2012B (N = 257), and 2013 (N = 559) samples were compared to identify individuals who had been tested at all three time points. The 2012A sample was not included as it had been used to conduct the EFA. A total of 116 participants were identified as have being tested at all three time points and thus made up the final longitudinal sample.

Each factor (Memory and Reaction Time) was tested individually, therefore two independent repeated-measures were submitted to invariance testing. A hierarchical series of nested single-domain models were fitted to the data from all three time points simultaneously. The first item of each factor (i.e., Word-memory and Symbol-match-reaction-time) was chosen as the marker variable and its factor loading was fixed to one for both groups. Item errors were allowed to correlate with identical items at different time points. This specification was based on Marsh's (1993) suggestion as he found that item errors tend to correlate when the same measure (i.e., item) is administered at different time points. Failure to account for these correlations can erroneously inflate factor loadings. This model served as the baseline (i.e., configural) model for invariance testing. Invariance testing followed Meredith's hierarchy as explained previously (Section 6.3.3.4). Configural, weak factorial, and strong factorial invariance were assessed. Each superior level of invariance was only tested if the preceding level was achieved. Each successive model was assessed for adequate model fit based on the previously mentioned criteria (i.e., CFI > 0.90, TLI > 0.90, RMSEA < 0.08). Invariance was achieved if there was no significant change in model fit, primarily indicated by a change of CFI below 0.02.

Differential Stability was reported for the model which achieved the highest level of invariance. Differential stability is the correlation between a latent variable measure at different time points. It represents the consistency of individuals' rank-order over time. The higher the correlation, the greater consistency of an individual's relative standing across time. It does not provide information regarding the consistency of individuals' absolute score. This is gauged from an analysis of latent means. Relative differences in latent means were assessed for

the Memory factor but not the Reaction Time factor given strong factorial invariance was only achieved for the former. The mean of the Memory factor at Time 1 was fixed to zero to define the numerical origin of the scale. The means of the Memory factor at Time 2 and Time 3 were unconstrained, allowing for a relative estimate of mean differences. To estimate the difference between Time 2 and Time 3 the mean at Time 2 was fixed to zero and Time 1 was unconstrained.

6.3.4.4 Stage Three: Cross-country invariance. Cross-country invariance was assessed within a *multi-group confirmatory factor analysis* (MGCFA) framework. MGCFA is simply an extension of CFA, which tests the invariance of estimated parameters of two nested models across groups (Kline, 2011). The data files for Group One (New Zealand) and Group Two (South Africa) were specified to AMOS. The first item of each factor (i.e., Word-memory and *Symbol-match-reaction-time*) was chosen as the marker variable and its factor loading was fixed to one for both groups. The marker variable defines the scale of the latent variable, as they have no inherent scale (Stevens, 2002). The respecified, two-factor, eight-item model, identified in Stage One was used as the baseline model for cross-country invariance testing. Previous to invariance testing each model was run individually to ensure the data adequately fit the model independently of one another. The investigation of cross-country invariance followed Meredith's hierarchy of increasingly stringent models, explained in detail previously. Configural and weak factorial invariance were assessed, however given the latter was not achieved, a test of strong factorial invariance and hence latent mean analysis was omitted. Each model was assessed for adequate fit (i.e., CFI > .90, TLI > .90, RMSEA < .08) and the change in fit, particularly the change in CFI between nested models, was analysed.

Chapter 7: Results

This section is separated into three stages. In the first stage the factor structure of the ImPACT battery was investigated via *exploratory factor analysis* (EFA) techniques and reliability analyses (i.e., internal consistency). To strengthen the findings of the EFA, the identified model was subsequently validated on a different sample via *confirmatory factor analysis* (CFA). Both Stage Two and Stage Three report findings from an investigation of the measurement invariance of the previously identified ImPACT structure. The second stage report's findings regarding longitudinal stability of the ImPACT battery, assessed via a repeated measures model confirmatory factor analysis, across three time points. The final stage reports results regarding cross-country (South Africa and New Zealand) invariance of the ImPACT battery within a *multiple-group-confirmatory factor analysis* model.

7.1 Participant Demographic Information

Demographic data for the final New Zealand ($N = 109$) sample is presented in Table 5 This sample omits cases that met exclusion criteria and those that were multivariate outliers. The demographics of the final 2012 South African sample ($N = 519$), before it was randomly split is presented in Table 6. Since the 2011 ($N = 595$) and 2013 ($N = 559$) South African samples consist of the same individuals as those in the 2012 sample, demographic data for only the 2012 year is presented. For each descriptive item if there were multiple categories with less than three participants they were amalgamated into one 'other' category. For both samples, the majority of participants had sustained no previous concussions, played Rugby as their primary sport, were right handed, and English was their first language. No comparisons were made within demographic variables given the disproportionate sample size of the dominant category compared with the others.

Table 5

Demographic Information of the New Zealand Sample

		Frequency	Percentage
Number of Previous Concussions	0	87	79.80%
	1	17	15.60%
	2	5	4.60%
Sport			
	Rugby	94	86.20%
	Soccer	8	7.30%
	Other	7	6.40%
Native Country			
	New Zealand	76	69.70%
	Samoa	7	6.40%
	America	6	5.50%
	Tonga	5	4.60%
	Fiji	3	2.80%
	South Africa	3	2.80%
	Other	9	8.10%
First Language			
	English	91	83.50%
	Tongan	5	4.60%
	Samoan	4	3.70%
	Other	9	8.10%
Handedness			
	Right	95	87.20%
	Left	12	11.00%
	Ambidextrous	2	1.80%

Table 6

Demographic Information of theSouth African Sample

		Frequency	Percentage
Number of Previous	0	366	70.80%
Concussions	1	108	20.40%
	2	45	8.80%
Sport			
	Rugby	342	65.80%
	Hockey	103	19.80%
	Basketball	31	6.20%
	Water polo	18	3.50%
	Soccer	10	1.80%
	Squash	5	1.20%
	Other	10	1.80%
Native Country			
	South Africa	455	87.60%
	America	21	4.10%
	Zimbabwe	10	2.40%
	Zambia	7	1.50%
	Kenya	4	0.90%
	Other	18	3.60%
First Language			
	English	451	87.60%
	Afrikaans	6	1.20%
	Other	62	0.60%
Handedness			
	Right	467	89.70%
	Left	31	6.20%
	Ambidextrous	21	4.10%

7.2 Stage One: Exploring the structure of the ImPACT Battery

Normality was assessed for the two South African 2012 samples independently. Mardia's co-efficient for sample 2012A ($N = 264$) indicated multivariate normality, evidenced by a value of 35.39, well below the calculated cut-off of 168. According to Bollen (1989), multivariate normality is achieved if Mardia's coefficient is lower than p(p+2), where p is the number of observed variables. There were originally 12 observed variables, thus 12(12+2) = 168. Mardia's co-efficient should be below 168 for normality to be supported. Furthermore, there was no clear discontinuation of Mahalanobis d-square values. Univariate normality was also supported, evidenced by skew values less than two and kurtosis scores less than seven as demonstrated in Table 7. Sample 2012A items skew values ranged from 0.53 to 1.91 and kurtosis values ranged from 0.22 to 5.30 and are presented in Table 8 in Appendix E.

Both univariate and multivariate normality was supported for Sample 2012B ($N = 255$). Univariate skew values ranged from 0.02 to 2.04 and kurtosis values varied from 0.01 to 6.21. Sample 2012B was used in the confirmatory factor analysis, which had only eight observed variables. Thus based on Bollen's (1989) criteria, a Mardia's value below 80 is indicative of multivariate normality. Mardia's co-efficient for the current sample was 50.12, thus normality was supported.

Table 7

Mean and Standard Deviations for South African Samples 2012A and 2012B

	South Africa 2012A		South Africa 2012B	
	M	*SD*	*M*	*SD*
Word-memory	92.36	8.28	92.75	7.58
Design-memory	82.79	12.82	82.63	13.35
XO-memory	71.62	19.47	70.62	19.12
XO-correct-interference	28.36	2.21	28.38	2.16
XO-reaction-time	0.49	0.07	0.49	0.07
XO-total-incorrect	7.27	5.28	7.66	5.56
SM-reaction-time	0.52	0.11	0.52	0.12
SM-memory	67.00	23.43	69.67	22.87
CM-reaction-time	0.78	0.15	0.77	0.16
CM-total-commissions	0.39	0.61	0.77	0.16
TL-memory	88.18	15.08	86.56	15.59
TL-correct-interference	45.70	13.47	44.86	14.15

Note. SM = Symbol Match; CM = Colour Match; TL = Three Letters.

7.2.1 Exploratory factor analysis and internal consistency

Sample 2012A was subjected to an exploratory factor analysis (EFA) utilizing maximum likelihood as the extraction method with direct oblimin rotation. Factorability of sample 2012A's correlation matrix was supported. Firstly, it revealed several correlations above .30 and none that exceeded .90 as depicted in Table 9 (see Appendix F) The highest correlation was reported between XO-reaction-time and XO-correct-interference ($r = -0.84$, $p < 0.01$) and was suggestive of multi-collinearity. Two items (Colour-match-commissions and XO-total-incorrect), both hypothesised to load onto the Impulse Control factor failed to produce correlations greater than .30 with any other variable. Notwithstanding, these items were initially retained given their inclusion in the ImPACT scoring structure. The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy statistic was .56 for the total scale (12-items). The lowest individual KMO score was reported for XO-total-incorrect (.15) and highest was for Colour-match-reaction-

time (.92). The Barlett's Test of Sphericity statistic was significant (χ^2 [66] = 1541.38, $p < .001$) indicating the correlation matrix significantly differed from the identity matrix. On the basis of these findings it was appropriate to proceed with factor analysis.

7.2.1.1 Model 1. All twelve items were initially submitted to a factor analysis. Three factors with eigenvalues (3.48, 1.78, 1.28) greater than one were extracted. The first factor accounted for 28.97% of total variance, followed by a factor that accounted for 14.82% variance, and a third factor accounting for 10.69% variance. The factors represented a Reaction Time/Processing Speed factor, an Impulse Control factor, and a General Memory factor, respectively. Cumulatively, these three factors accounted for 54.48% of the total variance of the samples performance on the ImPACT battery. During iterations one or more estimates with communalities greater than one were encountered, therefore the solution of the previous iteration (i.e., 15th) was presented. This is problematic given the true solution was not found and it is strongly suggestive of misspecification, thus caution was taken in interpreting the results. Communality values at extraction ranged from .05 to .99 and are presented in Table 10. Five communalities were weak, evidenced by values less than .30, in addition to two communalities equal to one (i.e., Heywood cases). The rotated factor loadings were all above the desired .32 cut-off, suggested by Tabnachnick and Fidell (2012), with the exception of CM-commissions. As expected the Reaction Time items loaded negatively onto their factor as a higher score represented poorer performance, whereas with all other items a higher score represented superior performance. Two Heywood cases were produced and dominated their associated factor; XO-total-incorrect-interference produced a factor loading of .99 and XO-total-correct-interference had a factor loading of .95, which were substantially higher than the majority of other items within their associated factors.

Given the aforementioned difficulties with the current solution, it was decided to remove the two Impulse Control items, XO-total-incorrect and CM-commissions and re-run the EFA. These items were removed, firstly because they did not produce correlations greater than .30 with any other item, including each other. Secondly, XO-total-incorrect produced a low KMO (.15) and a communality equal to 1, suggestive of misspecification. And lastly, literature suggests that

artificial measures of impulse control fail to capture the everyday executive dysfunction observed in concussed individuals (Maillard-Wermelinger et al., 2009; Turkstra & Byom, 2010). This is supported by the fact that the ImPACT guidelines recommend the Impulse Control composite should not be used to inform clinical decisions (ImPACTonline, 2013).

Table 10

Factor Loadings and Communalities of Twelve Items of the ImPACT Battery

	Factor			Communality
Item	Reaction Time/ Processing speed	Impulse Control	General Memory	
XO-total-correct	.95			0.99
XO-reaction-time	-.91			0.98
TL-correct-interference	.50			0.33
SM-reaction-time	-.37			0.12
CM-reaction-time	-.35			0.22
XO-total-incorrect		.99		0.99
CM-commissions		.22		0.05
Design-memory			.76	0.58
Word-memory			.65	0.39
SM-memory			.57	0.34
TL-memory			.43	0.20
XO-memory			.34	0.19

Note: Factor loadings less than .32 were suppressed. SM = Symbol Match; CM = Colour Match; TL = Three Letters.

7.2.1.2 Model 2. The remaining ten items were submitted to another factor analysis, again using maximum likelihood with direct oblimin rotation. The KMO statistic was .77 indicating that underlying dimensions could explain a large quantity of co-variation among variables. Individual KMO statistics were all above the acceptable limit of .50 and ranged from .70 to .90 with a mean of .79. The Barlett's Test of Sphericity statistic was significant (χ^2 [45] = 799.20, $p <$ .001), suggesting items were sufficiently correlated to proceed with factor analysis. A two-factor solution was produced. The first factor had an eigenvalue of 3.48 and accounted for 28.97% of the total variance, whereas the second factor had an eigenvalue of 1.78 and accounted for 14.82% of the total variance. The

factors appeared to represent a Reaction Time/Processing Speed factor and a General Memory factor, respectively, and cumulatively accounted for 54.48% of the ImPACT battery's total variance. Each factor possessed five indicators each. The indicators' rotated factor loadings were all greater than .32, the suggested minimum, and no substantial cross-loadings were present. Rotated loadings for factor one ranged from .38 to .88 and loadings for factor two ranged from .37 to .76. These are presented in Table 11.

Table 11

Factor Loadings and Communalities of Ten Items of the ImPACT Battery

	Factor		Communality
Item	Reaction Time/ Processing speed	General Memory	
XO-total-correct	.87		.86
XO-reaction-time	-.88		.80
TL-correct-interference	.53		.36
SM-reaction-time	-.41		.15
CM-reaction-time	-.38		.23
Design-memory		.76	.58
Word-memory		.65	.39
SM-memory		.58	.34
TL-memory		.44	.20
XO-memory		.37	.17

Note. Factor loadings less than 0.32 were suppressed. SM = Symbol Match; CM = Colour Match; TL = Three Letters.

Assessment of the internal consistency of each subscale revealed mixed findings. The Memory subscale demonstrated acceptable, although low internal consistency, evidenced by a *Cronbach alpha* value of .63 for the five memory items. *Corrected item-total correlations* ranged from .33 to .56. Although low they exceeded .30, therefore supportive of the scales reliability. Furthermore if any of the items were deleted it would result in a decrease of the overall subscales reliability (see Table 12, Appendix G). Therefore the inclusion of all five items was supported. Conversely, the reliability for the Processing Speed/Reaction Time subscale was extremely poor, evidenced by a Cronbach alpha of .17. Analysis of

the *Cronbach's alpha if item deleted* indicated that deletion of any individual items would have produced minimal change in reliability.

Given the unacceptable internal consistency of the Reaction Time/Processing Speed scale another factor analysis was run with the deletion of Three-letters-average-correct and XO-total-correct. These two items were chosen to be deleted primarily because ImPACT conceptualises them as complex processing speed tasks and there is consensus in the literature that simple and complex processing speed are two distinct constructs (Chiaravalloti et al., 2003; Flanagan & Harrison, 2012; Salthouse, 1996). Thus compressing them together to form one unitary concept would be erroneous. Furthermore as explained in the method section it is of the author's opinion that XO-total-correct is representative of simple processing speed, not complex processing speed as outlined in the scoring structure. Furthermore it is redundant given its similar nature to XO-reaction-time which is a more direct measure of simple processing speed. Therefore only one complex processing speed item remained and one item is not a sufficient representation of a latent construct.

    7.2.1.3. Model 3. The remaining eight items were submitted to another factor analysis, again utilizing maximum likelihood as the extraction method, with direct oblimin rotation. The overall KMO statistic (.75) was above acceptable limits in addition to individual item KMO's (range: .57 – .84). Furthermore Bartlett's Test of Sphericity was significant (χ^2 [28] = 339.68, $p < .001$), indicating items were sufficiently correlated to proceed with factor analysis. Two factors were extracted based on eigenvalues greater than 1. The first factors eigenvalue was 2.59 and explained 32.40% of common variance. This factor possessed the five memory items and thus represented a General Memory factor. The second factor represented Reaction Time (i.e., simple processing speed) and had an eigenvalue of 1.45, explaining 18.08% of the shared variance. Cumulatively these two factors accounted for 50.48 of the overall variance. The rotated loadings were all above the suggested .32 minimum and ranged from .40 to .76, as demonstrated in Table 13.

Table 13

Factor Loadings and Communalities of Eight Items of the ImPACT Battery

	Factor		Communality
Item	General Memory	Reaction Time	
Design-memory	.76		.58
Word-memory	.62		.39
SM-memory	.59		.35
TL-memory	.44		.19
XO-memory	.40		.16
XO-reaction-time		.66	.47
SM-reaction-time		.58	.29
CM-reaction-time		.51	.37

Note. Factor loadings less than .32 were suppressed. SM = Symbol Match; TL = Three Letters; XO = X and O's; CM = Colour Match.

Internal consistency of the revised Reaction Time scale revealed minimally acceptable reliability, evidenced by a Cronbach alpha value of .54. The corrected item-total correlations for XO-reaction-time (.47), Symbol-match-reaction-time (.32), and Colour-match-reaction-time (.44) indicated that each item was related to the total scale score. Based on alpha if item deleted values, if any of the items were to be deleted the scales reliability would decrease (see Table 14, Appendix G). Lastly the internal consistency of the overall scale, inclusive of both Memory and Reaction Time items was .58.

7.2.2 Confirmatory factor analysis

Utilizing AMOS graphics (version 21; Arbuckle, 2012) two distinct models were submitted to independent confirmatory factor analyses (CFA) with South African sample 2012B ($N = 257$). The first model was identified through a principle component analysis conducted by Allen and Gfeller (2011) and the second was identified in the previous exploratory factor analysis (EFA).

In the present analysis, Allen and Gfeller's (2011) model as depicted in Figure 3 (Section 6.3.4.2) was under-identified, in that not all the information needed to calculate the unknown parameters were present. Therefore it did not converge and

no solution was found. Conversely the model identified in the previous EFA did produce a solution that demonstrated adequate fit to the data. A non-significant (p = .26) chi-square statistic of 22.46 with 19 degrees of freedom was produced along with an adequate TLI (.99), CFI (.98), and RMSEA (.03; .00–.06). No modification indices were produced and no problematic standardised residuals were revealed. The largest standardised residual was 1.99. All freely estimated unstandardised regression weights were statistically significant (p < .001). Standardised factor loading estimates revealed a moderate (i.e., .40 − .60) relationship for five indicators and a strong (i.e., >.60) relationship for three indicators with their purported latent factors as depicted in Figure 4. Additionally, a moderate negative relationship (-.41) between the two latent factors was observed.
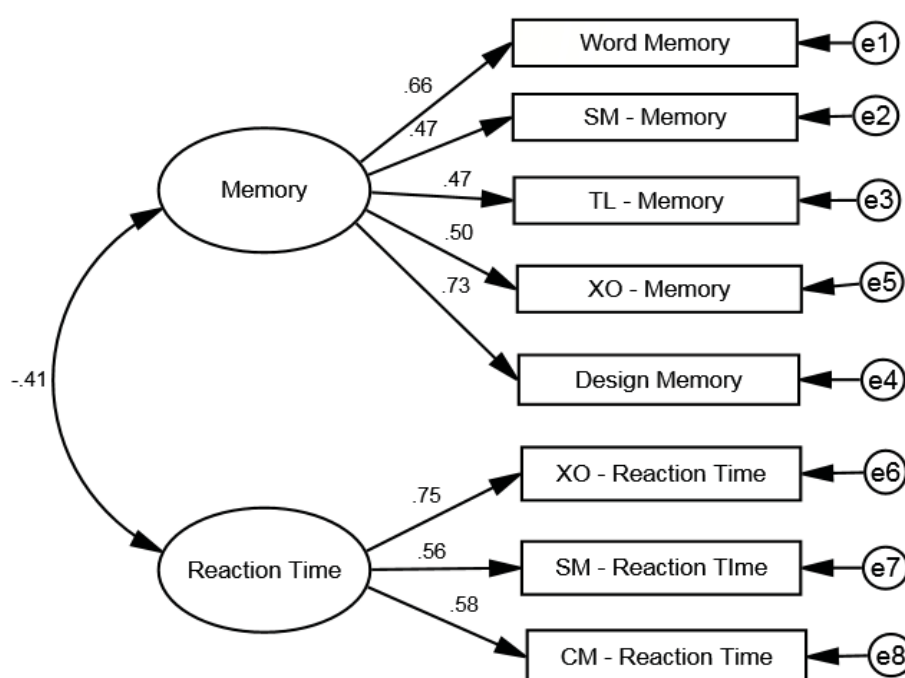
Figure 4. Confirmatory factor analysis for the eight item, two-factor correlated model.

7.3 Stage Two: Longitudinal Stability

Participants who were tested at all three time points (2011, 2012, 2013) were included in the investigation of longitudinal stability for the eight-item, two-factor

model identified and confirmed in Stage One. Each item's mean and standard deviation are reported in Table 15 and skew and kurtosis are presented in Table 16 in Appendix H, for each time point. These descriptive statistics are presented for all twelve items although only eight were included in the analysis. Univariate normality was supported for the eight included items, over the three time points, evidenced by a skew less than two and kurtosis value less than seven, with the exception of Three-letters-memory (skew 2.15) at Time 1 and Word-memory (2.17) at Time 3. Multivariate normality was also supported for the memory items, evidenced by a Mardia's coefficient of 57.90, a value less than the cut-off of 80 based on Bollens calculation (p[p+2]; p = number of observed variables).

Table 15

Mean and Standard Deviations for Each Item Across Time Points, South African Sample

	2011 ($n = 116$)		2012 ($n = 116$)		2013 ($n = 116$)	
Item	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word-memory	94.67	6.45	93.87	20.15	93.41	8.57
Design-memory	84.37	13.96	84.11	13.41	85.35	12.75
XO-memory	67.46	22.14	72.49	20.15	72.85	20.88
XO-total-interference	27.45	2.03	8.60	1.90	27.57	1.80
XO-reaction-time	0.51	0.07	0.48	0.06	0.51	0.06
XO-total-incorrect	8.60	6.5	7.66	5.05	7.54	6.05
SM-reaction-time	0.50	0.10	0.50	0.10	0.52	0.09
SM-memory	69.64	22.89	71.36	23.18	73.37	26.37
CM-reaction-time	0.82	0.16	0.76	0.13	0.80	0.13
CM-total-commissions	0.78	1.01	0.61	0.95	0.54	0.96
TL-memory	85.75	16.73	86.47	17.08	89.31	13.93
TL-correct-interference	39.64	13.82	46.35	13.10	52.16	12.49

Note. XO = X and O's; SM = Symbol Match; CM = Colour Match; TL = Three Letters.

Meredith's hierarchy of invariance testing was followed in that a series of increasingly stringent longitudinal models were tested. Longitudinal invariance and hence structural stability are reported for each unidimensional factor (i.e., Memory and Reaction Time) individually.

7.3.1 Memory factor

A baseline model with correlated errors was initially assessed to ensure each time point adequately fit the model, in the absence of any temporal constraints. All freely estimated unstandardised regression weights were statistically significant. The baseline model and standardised factor loadings for each of the three time cycles are presented in Figure 5. Design-memory produced the largest loading at Time 1 (.76) and Time 2 (.73) and at each time point demonstrated a strong relationship with the Memory factor. The Word-memory item also demonstrated a strong relationship with the Memory factor at Time 2 (.67) and Time 3 (.74), however a moderate (.55) relationship was observed at Time 1. XO-memory, Symbol-match-memory, and Colour-match-memory items all demonstrated a moderate relationship with the Memory Factor across all three time points. The correlation between the Memory Factor at Time 1 and Time 2 was .86, between Time 1 and Time 3 was .79, and between Time 1 and Time 2 was .92, suggestive of a strong relationship between all the time points.

The longitudinal baseline model was supportive of configural invariance, evidenced by a CFI of .96, a TLI of .95, and a RMSEA of .05 (.00-.08). Constraints were subsequently placed on all eight measurement weights, and the resulting solution indicated adequate fit to the data evidenced by a CFI of .95, TLI of .93, and a RMSEA of .06 (.03 - .08). The change in CFI (Model 2 – Model 3) was .01, below Meade and colleagues' (2008) cut-off of .02. However the difference in the chi-square statistics of the two nested models (16.35) slightly exceeded the critical value (χ^2 [7] =14.07). Given the limitations of the chi-square difference test as discussed in the Method section 6.3.3.6.1, it was concluded that on the basis of non-significant change in CFI, weak factorial invariance was achieved. Strong factorial invariance (i.e., Model 4) was also achieved as the subsequent model in which constraints were added to item intercepts provided adequate fit to the data (CFI = .95, TLI = .95, RMSEA = .05) and produced a

change in CFI that was less than .02 as displayed in Table 17. However, again the difference in chi-square (19.11) slightly exceeded the critical value (χ^2 [10] = 18.31).

Table 17

Goodness-of-Fit Statistics for Progressive Levels of Longitudinal Invariance for the Memory Factor

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	CFI	ΔCFI	TLI	RMSEA
M1:Correlated errors	92.09***	72			.96		.95	.05 (.00-.08)
M2:Equal loadings	108.45*	79	16.35	7	.95	.01	.93	.06 (.03-.08)
M3:Equal intercepts	114.93*	89	19.11	10	.95	.02	.95	.05 (.02-.08)

Note. M1 = Model 1; M2 = Model 2; M3 = Model 3.

Strong factorial invariance was the highest level of invariance achieved, thus differential stability coefficients were estimated for the model. Strong differential stability was achieved between all three time points. The strongest correlation was exhibited between Time 2 and Time 3 (.90), followed by that between Time 1 and Time 2 (.88), and Time 1 and Time 3 (.81).

Given scalar invariance was achieved longitudinal differences among latent means were assessed. The latent mean model fit the data well, evidenced by a CFI value of .94, a TLI value of .92, and an RMSEA of .06 (.03-.08) As can be seen in Table 18 there were no significant differences in the latent means of both Memory at Time 1 and Time 2 (-.13) and Memory at Time 1 and Time 3 (.72) as evidenced by critical ratio's below 1.96 and p-values above .05. To estimate the difference between Memory at Time 2 and Time 3, the mean at Time 2 was then constrained to zero and Time 1 was freely estimated. Again there was no significant difference between the Memory factor at Time 2 and Time 3 (.86, $p > .05$).

Figure 5. Baseline longitudinal stability model for the Memory factor with factor loadings, factor correlations, and error correlations.

Table 18

Differences Between Memory Latent Means Over Time

	Estimate	SE	c.r.	*p*
T1 - T2	-0.13	0.57	-0.23	0.82
T1 - T3	0.72	0.66	1.10	0.27
T2 - T3	0.86	0.54	1.58	0.12

Note. T1 = Time 1; T2 = Time 2; T3 = Time 3; SE = standard error; c.r. = critical ratio; p = significance level.

7.3.2. Reaction time factor

A longitudinal invariance baseline model for the Reaction Time factor supported the unidimensional structure across the three time points. All freely estimated unstandardised regression weights were statistically significant at the .001 level, with the exception of Symbol-match-reaction-time at Time 3, which was significant at the .01 level. The Reaction Time baseline model and standardised factor loadings for each time point are presented in Figure 6. The XO-reaction-time item consistently produced strong factor loadings at all three time points (.83, .68, .84), respectively. The Symbol-match-reaction-time item demonstrated a moderate loading at Time 1 (.54) and Time 2 (.53), however loaded poorly onto the Reaction Time factor at Time 3 (.32). Lastly, the Reaction Time factor at Time 2 (.55) and Time 3 (.49) loaded moderately onto Colour-match-reaction-time. The correlation between the reaction time factor at Time 1 and Time 2 was .87, between Time 1 and Time 3 was .75, and between Time 1 and Time 2 was .71.
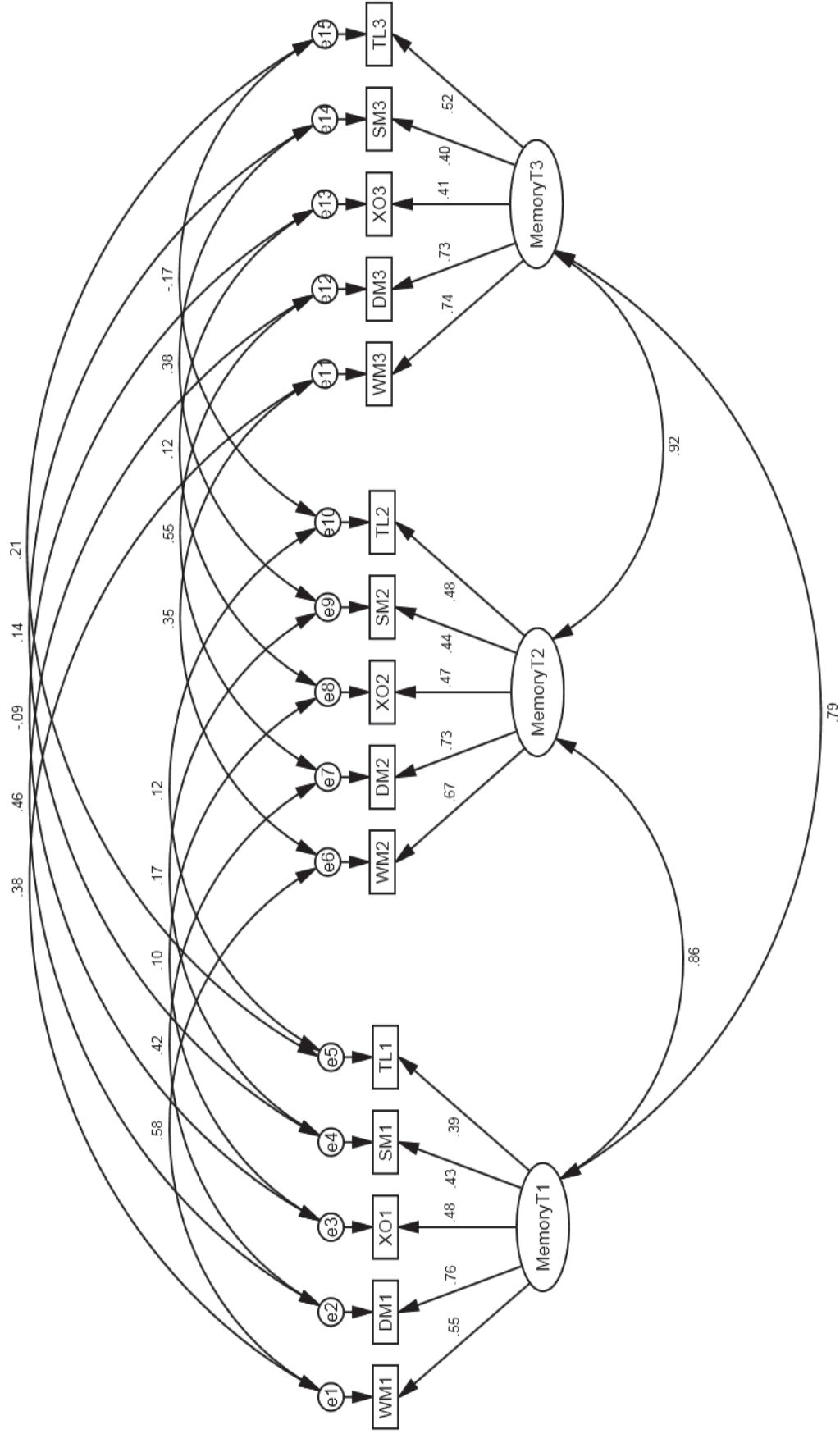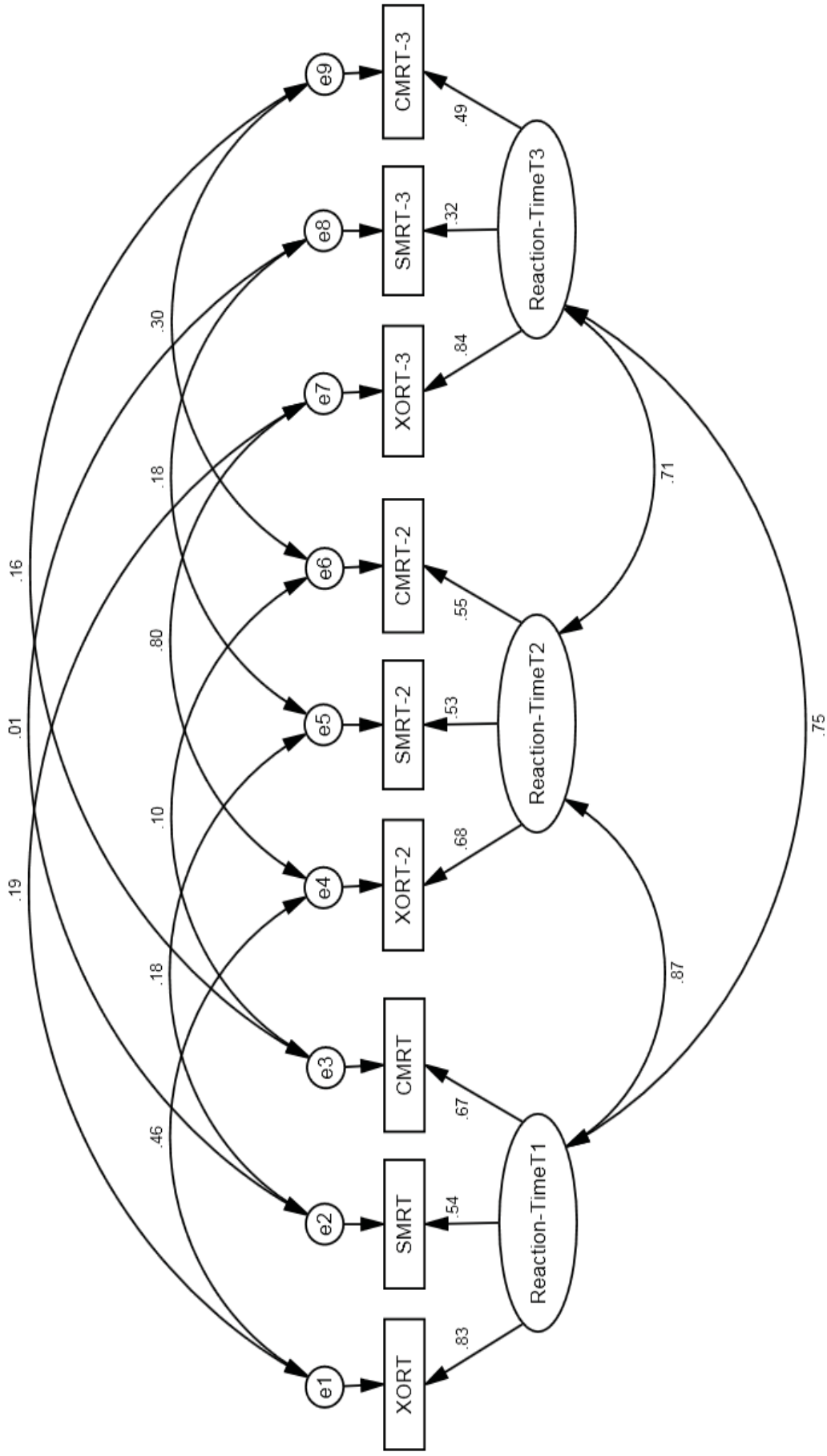
Figure 6. Baseline longitudinal model for the Reaction Time factor with factor loadings, factor correlations, and error correlations.

Fit statistics for the longitudinal baseline measure supported configural invariance, evidenced by a non-significant chi-square statistic (χ^2 [15] =16.97, $p < .05$), a CFI of .99, TLI of .99, and RMSEA of .04 (.00 − .09). The metric model in which factor loadings were constrained equal also provided adequate fit to the data (CFI .99, TLI .99, RMSEA of .03) as depicted in Table 19. Furthermore, metric invariance (i.e., weak factorial invariance) was supported, given that change in CFI was minimal (.001) and the chi-square difference (4.30) was smaller than the critical value (χ^2 [4] = 9.49). The scalar model in which all intercepts were constrained equal across groups failed to provide an adequate fit to the data and did not support strong factorial invariance. The change in CFI exceeded .02 and the chi square difference (83.63) exceeded the critical value (χ^2 [6] =12.59). Standardised residuals revealed no potential areas of poor fit, as the highest residual was 1.55. Three intercept modification indices were however produced for the XO-reaction-time item at each time point. None of the associated expected parameter of change (EPC) values exceeded 0.1 and only the modification index for XO-reaction-time at Time 2 exceeded 20. Therefore partial invariance was explored by re-analyzing Model 3 with the intercept for XO-reaction-time at Time 2 unconstrained (i.e., Model 4). The resulting model provided adequate fit to the data (CFI = .95, TLI = .93, RMSEA = .08), but failed to support partial scalar invariance, evidenced by a change in CFI of .04 and a non-significant ($p > .05$) chi-square difference test. The highest standardised residual was 1.05 and only one intercept modification index was produced, however it was below 20 and its EPC did not exceed 0.1. No localized areas of potential non-invariance were identified therefore invariance testing ceased and partial scalar invariance was not achieved.

Stability coefficients were estimated from the metric model as weak factorial invariance was the highest level of invariance achieved. The stability co-efficient reported between Time 1 and Time 2 (.84) was the strongest, followed by Time 2 and Time 3 (.76), and lastly Time 1 and Time 3 (.75).

Table 19

Goodness-of-Fit Statistics for Progressive Levels of Longitudinal Invariance of the Reaction Time Factor

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	CFI	ΔCFI	TLI	RMSEA [95% CI]
M1:Correlated Errors	16.97	15			.99		.99	.03 [.00-.09]
M2:Equal Loadings	21.27	19	4.30 (9.49)	4	.99	.00	.99	.03 [.00-.09]
M3:Equal Intercepts	104.89	25	83.63 (12.59)	6	.76	.23	.65	.17 [.13-.20]
M4:Equal intercepts (XORT-T2 unconstrained)	40.07*	24	18.80 (11.07)	5	.95	.04	.93	.08 [.03-.12]

Note: M1 = Model 1; M2 = Model 2; M3 = Model 3; M4 = Model 4; XORT-T2 = XO Reaction Time item at Time 2.
* $p < 0.05$

7.4. Stage Three: Cross-Country Invariance

Cross-country invariance was assessed between a New Zealand sample ($N = 109$) and the 2013 South African sample ($N = 116$). Each item's mean, standard deviation, skew, and kurtosis for the New Zealand sample are presented in Table 20. These descriptive statistics were previously reported for the South African sample in Table 7 in Section 7.2 and Table 8 in Appendix E. Among the New Zealand sample initial univariate skew and kurtosis values indicated five of the eight items were non-normal. Additionally the Mardia's coefficient (425.85) indicated extreme multivariate non-normality as based on Bollens (1989) calculation, any value which exceeded 80 was suggestive of non-normality. Mahalanobis Distance (*D*) values identified nine cases as multivariate outliers and these were subsequently deleted from the analysis. As a result univariate skew and kurtosis values fell within the

acceptable range as displayed in Table 20, and the Mardia's coefficient (10.94) indicated the data were now multivariately normal.

Table 20

Item Means, Standard Deviations, Skew and Kurtosis Values for the New Zealand Sample

Item	*M*	*SD*	Skew	Kurtosis
Word-memory	92.62	5.97	-0.96	0.87
Design-memory	74.30	13.87	-0.05	-0.49
XO -memory	66.28	18.54	-0.36	-0.22
XO-total-interference	26.12	2.80	-1.94	6.50
XO-reaction-time	0.55	0.07	1.14	2.71
XO-total-incorrect	8.10	6.26	1.11	0.83
SM-reaction-time	0.55	0.13	1.60	3.50
SM-total-memory	62.79	20.20	0.09	-0.67
CM-reaction-time	0.82	0.18	0.80	2.30
CM-total-commission	0.46	0.67	1.35	1.31
TL-letters-correct	86.69	17.21	-2.23	5.81
TL-correct-interference	41.40	14.07	-0.09	0.27

Note. XO = X and O's; SM = Symbol Match; CM = Colour Match; TL = Three Letters.

The hypothesised eight-item, two-factor model was computed with each sample individually. The model provided a good fit to the data of each country's sample. The New Zealand model produced a non-significant chi-square statistic (χ^2 [19] = 29.88), an adequate CFI (.93) and RMSEA (.72), although the RMSEA statistic possessed a wide confidence interval (.00 − .12). The TLI (.89) was slightly below the desired .90. All unstandardised estimates were statistically significant at an alpha level of .01. Standardised estimates were all above the minimum recommendation of .32, the minimum loading suggested by Tabachnick and Fidell (2012). Symbol-

match-reaction-time produced a low standardised loading (.33), Word-memory (.46), Symbol-match-memory (.55), and Three-letters-memory (.53) demonstrated moderate loadings, whereas Design-memory (.63), XO-memory (.61), XO-reaction-time (.82), and Colour-match-reaction-time (.66) loaded strongly with their purported factors. A moderate negative relationship (-.58) was found between the two latent factors, Memory and Reaction Time as displayed in Figure 7. No factor-item modification indices were produced and the highest standardised residual was 1.75.

The hypothesised model as shown in Figure 8 also provided an adequate fit to the South African sample (2012B). The chi-square statistic (χ^2 [19] =24.24) was non-significant and the CFI (.96) and TLI (.94) were both above the desired .90. Furthermore the RMSEA suggested good fit (.05). However, as with the New Zealand sample, the confidence interval was wide (.00-.10). All unstandardised regression weights were significant at an alpha level of .05. All standardised loadings were above the desired .32. Two items evidenced weak loadings, three items were moderate, and three items revealed a strong relationship with their purported factors. Only a small negative correlation (-.18) was reported between the latent factors of Memory and Reaction Time. No factor-item modification indices were produced and the highest standardised residual was 2.33.
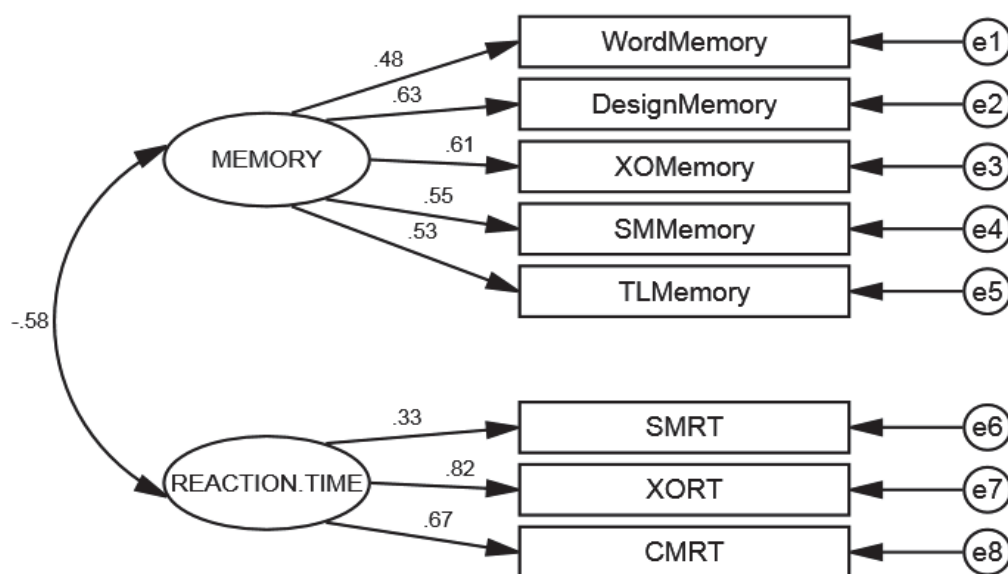
Figure 7. Baseline model with standardised regression weights for the New Zealand sample.
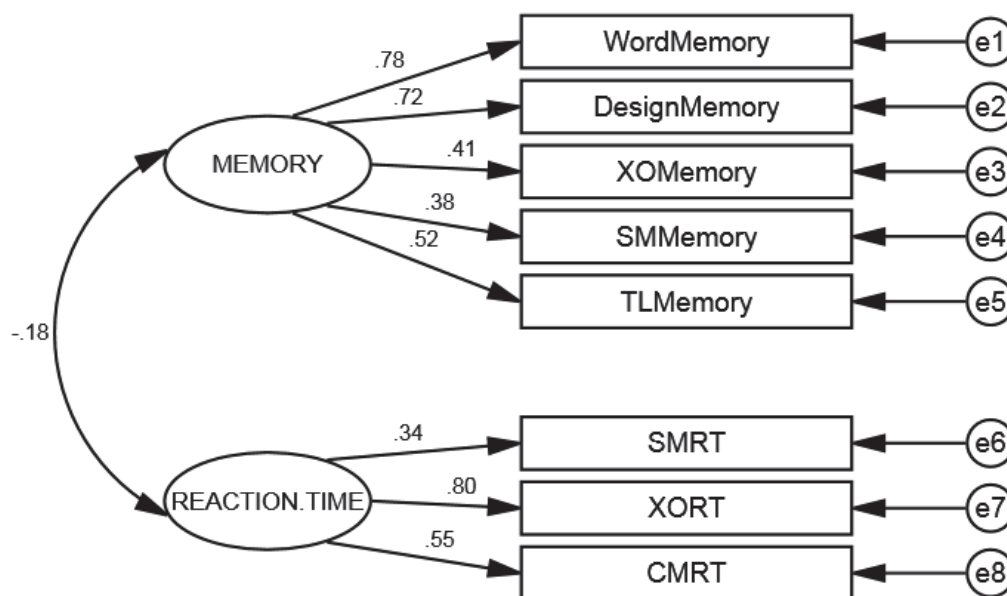
Figure 8. Baseline model with standardised regression weights for the South African sample.

To test for configural invariance (i.e., Model 1) the two models were run simultaneously. Configural invariance was supported as the model adequately fitted the data, evidenced by a CFI of .94, a TLI of .91, and a RMSEA of .04 (.01-.07) as displayed in Table 21. Equality constraints were subsequently placed on the regression weights to test for metric invariance. Whilst the model provided adequate fit to the data (CFI = .92, TLI = .89, RMSEA = .05) the change in CFI (.03) exceeded .01, therefore metric invariance was not supported. Only one factor-to-item modification index was produced and it was between the Reaction Time factor and Three-letter-memory item, among the New Zealand sample. No problematic standardised residuals (i.e., > 2.58) were evident for either group. Model 2 was re-run with Three-letter-memory regression weight unconstrained (Model 3). This resulted in a minute difference in model fit and the change in CFI (.03). Additionally the chi-square difference test (χ^2 [5] = 12.94) failed to support metric invariance. Model 3 did not produce any problematic standardised residuals or modification indices, thus invariance testing ceased.

Table 21

Goodness-of-Fit Statistics for Progressive Levels of Cross-Country Invariance

Model	χ^2	df	$\Delta\chi^2$	Δdf	CFI	ΔCFI	TLI	RMSEA
Baseline:								
New Zealand	29.88	19			.93		.89	.07(.00-.12)
(N=109)	24.24	19			.96		.94	.05(.00-.10)
South Africa								
(N = 116)								
M1: Configural model	53.12*	38			.94		.91	.05(.01-.07)
M2:Equal loadings	67.23*	44	14.11	6	.92	.03	.89	.05(.02-.07)
M3: Equal loadings, TL-memory unconstrained	66.06*	43	12.94	5	.92	.03	.89	.05(.02-.07)

Note: M1 = Model 1; M2 = Model 2; M3 = Model 3; TL = Three Letters.
* $p < .05$

Chapter Eight: Discussion

Neuropsychological assessment is now an integral component of concussion assessment and management within the sporting arena. The Immediate Post-concussion Assessment and Cognitive Test (ImPACT) is a commonly used neuropsychological battery to assess for the presence of cognitive dysfunction in concussion. The empirical support regarding ImPACT's psychometric properties has to date been inconclusive and incomplete, with only a small number of studies assessing limited areas of reliability and validity. This is concerning since the utility of ImPACT, or any other neuropsychological test, is reliant upon a foundation of robust reliability and validity. If psychometric properties are poor, clinicians cannot have confidence in interpretations that arise from the measure.

The aim of the current study was to extend the psychometric literature pertaining to the ImPACT neuropsychological battery. This was achieved using a sample of both New Zealand and South African male adolescent athletes. To date, psychometric evaluations of the ImPACT battery have been based upon simplistic statistical methods. This study represents the first investigation using advanced statistical techniques that control for methodological error, producing a more accurate picture of the state of affairs. The current study investigated three main objectives, each of which is presented below accompanied by a discussion of the findings and their implications.

8.1. Investigation of ImPACT's factor structure

The first objective was to identify the underlying factor structure of the ImPACT battery. Due to several concerns with the ImPACT scoring structure, initial factor analysis took an exploratory approach as opposed to attempting to confirm the current composite structure. A primary concern was that neither the theoretical nor psychometric foundations, upon which the battery was developed could be located by the author of the present study in the literature or test manual. Previous factor analytic studies have produced solutions that are inconsistent with the current scoring (i.e., composite) structure of ImPACT and these studies are replete with methodological limitations. An additional concern was that upon examination of item content, the researcher's opinion was that the model was mis-specified. Firstly, one item was redundant (XO-total-correct-interference) given its similarity to

another (XO-reaction-time) and the Symbol-match-memory item was loaded with the Verbal Memory factor when its content was representative of Visual Memory.

Three exploratory factor analyses (EFA) were run before the final model was reached. The final model possesses only eight items, with two factors. The first EFA with all twelve items produced a solution with three factors, representative of General Memory, Processing Speed, and Impulse Control. This model was unsatisfactory given some items shared very little variance (i.e., common variance) with other items within the battery and some items produced Heywood cases (i.e., communality greater than one) which are indicative of model mis-specification. The Heywood cases were present for XO-total-correct and XO-total-incorrect.

It was decided to remove two items; XO-total-incorrect and Colour-match-commissions, these were both Impulse Control items. They were removed firstly because XO-total-incorrect produced a Heywood case and, additionally, neither produced correlations greater than .30 with any other item, including each other. It is concerning that they failed to sufficiently correlate with one another given they are hypothesised to be measuring the same latent construct. Given the lack of association between the two items it appears they are measuring two distinct constructs. When items loading onto one construct are not sufficiently related in that at least one is not representative of the construct of interest, the erroneous item can bias the overall composite score leading to false interpretations regarding the construct of interest.

The literature suggests that standardised measures of executive function (e.g., impulse control) fail to capture the everyday executive dysfunction observed in concussed individuals (Lezak, 1982; Maillard-Wermelinger et al., 2009; Turkstra & Byom, 2010). This is supported by the fact that ImPACT guidelines state that the Impulse Control composite should not be used to inform clinical decisions. Additionally, studies that have investigated the psychometric properties of ImPACT typically omit the Impulse Control composite and its items.

The second EFA, with the removal of the above two items, produced an acceptable two factor model. Adequate model fit was achieved and all items were sufficiently associated with their purported factors. The two factors were the same two of three produced in the previous analysis, General Memory and Processing Speed. The current two-factor solution was also consistent with Schatz and Maerlender's (2013)

study in that a clear distinction was made between memory and speed. Importantly, the Processing Speed factor was inclusive of items measuring both reaction time (i.e., simple processing) and visual processing speed (i.e., complex processing speed). This however was consistent with a principal components analysis (PCA) conducted with the four main ImPACT composites and the Symbol Digits Modalities Test (SDMT), a measure of processing speed and attention (Iverson et al., 2005). Using a concussed sample Iverson and colleagues (2005) found two factors, the first consisting of the memory items and the second factor including the SDMT battery items, the reaction time items, and the processing speed items. Thus the authors concluded that these three types of items were measuring a similar underlying construct (Iverson et al., 2005).

The current EFA findings are partly inconsistent with the ImPACT composite structure, CHC theory, and previous cognitive literature, which argue that simple processing speed (i.e., reaction time) and complex processing speed are two distinct factors (ImPACT, 2013; McCrew, 1997). In the current study the distinction between simple and complex processing speed emerged only after the internal consistency of the Processing Speed factor was assessed via Cronbach's alpha. The inter-relatedness of items as evidenced by Cronbach's alpha was extremely poor. Perhaps the EFA grouped these five items together due to the scarcity of items and, with a greater number of items representative of each construct (i.e., simple processing speed and complex processing speed), the distinction may have emerged. Furthermore, if raw scores were used as first-level indicators instead of parceled items, this may have also led to a clearer distinction between simple and complex processing speed. Another explanation may lie in the similarity between XO-reaction-time and XO-total-correct. In the author's opinion XO-total-correct is a measure of simple processing speed. However, according to the ImPACT composite structure, it is a measure of complex processing speed as it is included in the Visual Processing Speed composite. A complex processing speed task is one in which all stimuli are presented simultaneously, allowing the examinee to determine the pace of progression and requiring them to continuously shift their attention from one item to the next. Simple processing speed tasks on the other hand have one stimulus (i.e., trial) presented at a time. Thus, the examiner controls the pace of presentation and often brief breaks occur between trials. The XO-total-correct score represents the

total number of trials the participant got correct on the interference XO task. The task was a choice reaction time test, representative of simple processing speed in that one trial was presented at a time. Hence, the XO-total-correct score provides very similar information to that of XO-reaction-time. This hypothesis is supported by the high and significant correlation between the two items (r = -.84, p = .01). Consequently, the other visual processing speed item was also removed since, although a measure of cognitive effects of concussion should ideally include a measure of complex processing speed, one item is not sufficient to represent the latent construct.

Given the above, the third EFA was run with eight items and an acceptable solution was produced in that the model adequately fit the data and items sufficiently loaded with their corresponding factors. The internal consistency (i.e., Cronbach's alpha) of each subscale was borderline acceptable. This was lower than expected given the robust factor loadings of individual items to their construct. The extremely small number of items per scale may explain the low internal consistency despite robust factorial validity. It has been found that Cronbach's alpha increases as the number of items in a scale increases (Nunnally, 1978). The identified sub-scales of Reaction Time and Memory only had three and five items each respectively. Also of note is that estimates of alpha lie at the lower limit of reliability and, as Sijtsma (2009) states, they are a gross underestimation of reliability. Thus the acceptability of the internal consistency of the current study's scales is debatable. Notwithstanding, more emphasise was placed on the factorial validity findings since, in addition to alpha's sensitivity to sample size, factorial validity was the main point of inquiry.

The two factors (i.e., subscales) identified were representative of a General Memory factor and a Reaction Time/Simple Processing Speed factor. Of note, there was no distinction between non-verbal and verbal memory. Previous research appears mixed on whether or not these should be considered individual factors or one General Memory factor (e.g., Ardila et al., 1998; Bowden et al., 1999; Bradley et al., 2003). However, the ImPACT composite structure does distinguish between verbal and visual memory, which is inconsistent with the current study's findings. The inability of the current EFA to distinguish between these two types of memory may be due to the lack of indicators. If it was possible to include the raw items as opposed to subscale scores (i.e., parceled items) the distinction between visual and verbal

memory may have emerged. Furthermore, the Symbol-match-memory task appears to be more indicative of visual memory than verbal memory, as stipulated by the ImPACT scoring structure. The task requires the participant to recall which symbol matches which number, thus it is clearly a task of visual memory. This hypothesis is supported by the higher correlation of Symbol-match-memory with Design-memory (i.e., visual memory) compared to Word-memory (verbal memory). Furthermore other cognitive tasks, which use symbol stimuli, are classified as visual tasks, not verbal tasks. For example, the Coding task in the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997) has an optional free recall task in which the participant recalls the symbol-number pairings. It is hypothesised that poor performance on this task should be mirrored by poor performance on visual memory tasks (Espe-Pfeifer & Wachsler-Felder, 2000). Thus it is unclear why the developers of ImPACT use the Symbol-match-memory score in the calculation of the Verbal Memory composite. One last potential explanation for the holistic General Memory factor may be that the Word-memory task, supposedly representative of verbal memory, was presented visually as opposed to audibly. Thus it may also be tapping areas of visual memory in addition to verbal memory.

Following the EFA, two confirmatory factor analyses (CFA's) were conducted employing a new data set. A CFA was run for the two-factor, eight-item structure previously identified through EFA, for the purpose of confirming the structure and thus to strengthen confidence in the findings. A second CFA was computed for Allen and Gfeller's (2011) model. It was the only other hypothesised factor structure identified in the literature that also used subscale scores as indicators. The structure identified was inconsistent with that identified in the current study's EFA as well as with ImPACTs scoring structure. As expected, Allen and Gfeller's (2011) model failed to converge and thus a solution was not found. Conversely, the data was supportive of the current study's two-factor, eight-item model, in that it produced a good fit to the data and item-factor relationships were robust. As expected, the two latent factors, Memory and Reaction Time, were moderately associated with one another. Greater memory performance was associated with decreased reaction times. The fact the structure identified in the current study was successfully validated provides robust evidence for its factorial validity and hence, construct validity.

This study represents the first to validate a hypothesised factor structure of the ImPACT battery through CFA. Confirming or validating a factor structure in a sample other than that in which it was originally identified is important as it strengthens the validity of the structure. It provides support for the construct validity of the Reaction Time factor and a General Memory factor. The structure identified and subsequently validated in the current study is inconsistent with the current ImPACT scoring structure. Importantly, this is not the first study to produce a factor solution through exploratory means that differs from that stipulated by ImPACT (e.g., Iverson et al., 2005; Allen & Gfeller, 2011). Thus it is uncertain why the ImPACT programme maintains their current scoring structure despite the absence of supporting evidence.

Based on the current findings it is strongly recommended that the developers of ImPACT revise the current items included in the battery and its scoring structure. The current findings suggest that there should be no distinction between visual and verbal memory, and that the ImPACT scoring structure should therefore only have one General Memory composite. This was the case in earlier versions of the ImPACT battery (Schatz & Maerlender, 2013). Furthermore, it is recommended that XO-total-correct-interference item be removed given its similar nature to XO-reaction-time. Lastly, although a Processing Speed factor was not produced, empirical research indicates it to be an important area to assess in concussion. This is supported by the fact that other popular neuropsychological concussion batteries, such as Cogstate and the Concussion Resolution Index (CRI), include a measure of processing speed. Thus, it is recommended that items be added so the processing speed construct is sufficiently represented. Furthermore, it is important that complex processing speed items not be combined with reaction time items (i.e., simple complex processing speed) to create one composite, given that the literature clearly demonstrates them to be two distinct constructs.

ImPACT has the potential to be a useful assessment tool for concussion as the items retained demonstrated robust validity. However, it needs to be modified to improve its efficiency. ImPACT is a brief measure; brief measures need to be efficient in that all items included should provide unique information, unlike full neuropsychological batteries, brief measures do not have the luxury of including many items. Thus items must target the most prominent areas affected by concussion and those which can be

accurately measured. Thus it appears, Impulse Control and Executive Function items should be omitted and focus should rather be on including a sufficient number of items that accurately and reliably represent Memory, Reaction Time, and Complex Processing Speed. This is consistent with another popular concussion battery, the CRI by Headminder, which also does not assess executive functions (Erlanger et al., 1999).

8.2. Longitudinal Stability

The second objective of the current study was to assess the longitudinal stability of the ImPACT structure identified in the previous analysis. Longitudinal stability was assessed for the performance of South African male adolescent athletes on ImPACT across three time points, with approximately 1-year intervals. The ImPACT structure at each of the three time points was combined to form a single longitudinal model upon which the stability of the measure over time was tested through confirmatory factor analysis (CFA). Each factor (i.e., Memory and Reaction Time) was analyzed individually for its structural stability, differential stability, and latent mean stability over time.

8.2.1. Memory

The Memory factor with five indicators achieved all three levels of longitudinal stability. The structural stability of the Memory factor was supported, as configural, weak, and strong factorial invariance were achieved. The presence of weak factorial invariance means that factor loadings were equivalent across the three time points. Thus, given their equivalence, a unit of change in any of the memory item's score would result in an equal unit change in the Memory factor, regardless of the time point. Furthermore, given the intercepts were also invariant (i.e., strong factorial invariance) any change that is observed in the indicator items can be interpreted as true change in the latent Memory factor. Thus, the measurement precision of the five ImPACT items reflective of memory, were stable across time. That is, the memory items are consistently measuring the same concept of memory over time.

Differential stability was supported for the Memory construct indicating the rank-order of individuals' scores were consistent over the three time points. Thus, each participant's memory performance relative to others was consistent over time. As

expected, the differential stability decreased as the time interval increased. The correlation between Time 1 and Time 2 was .88, between Time 2 and Time 3 was .90, and the correlation for the 2-year interval, between Time 1 and Time 3 was slightly lower, .81. Differential stability is very similar to test-retest reliability in that they both measure the consistency of variance over time. The difference is that, because differential stability is computed within a confirmatory factor model, measurement error is taken into account and hence it is a more accurate estimate of longitudinal reliability than that found with test-retest reliability.

The differential stability found in the current study is superior to the test-retest reliability (i.e., intra-class correlation coefficient; ICC) over a two-year period reported by Schatz (2010). They reported low test-retest reliability for both the Visual Memory (.65) and Verbal Memory (.46) composites of the ImPACT battery. A possible explanation for the discrepancy in findings may be the different average age of the samples as there are differences in concussion presentation between adults and adolescents (McCrea et al., 2004; Pellman et al., 2006). Schatz employed a college sample whereas the current studies sample consisted of adolescents. Furthermore, differential stability took into account error whereas the test-retest reliability in Schatz's (2010) study did not. Lastly, it may be that when the items of two ImPACT Memory composites are combined, as in the current study, they produce a more reliable and stable construct due to the increase in the quantity of indicators and the appropriate factor-item associations. This hypothesis is supported by Schatz and Maerlender's (2013) study in which they recalculated the test-retest reliability for a General Memory factor, inclusive of visual and verbal memory items, using the data from Schatz's (2010) earlier study. They found test-retest reliability as evidenced by ICCs was greater for a combined General Memory (.74) factor than for the Visual Memory (.65) and Verbal Memory (.46) composites alone.

Given strong factorial invariance was achieved, an assessment of differences in latent means was warranted. Comparison of the Memory latent mean across the three time points revealed no significant differences in absolute scores. That is, non-concussed, South African, adolescent males' memory ability remained relatively stable over a three-year period. Based on the absence of improved performance, practice effects appeared absent. This is consistent with previous studies which too found ImPACT to be free from practice effects (Lovell et al., 2009; Schatz, Pardini,

Lovell, Collins, & Podell, 2006). This is an important observation given ImPACT is often used within a baseline framework, in that return-to-play decisions are based on post-injury performance returning to that at baseline. If practice effects were present, observed improvements in performance (i.e., returning to baseline) may be reflective of practice effects rather than true improvement and recovery. The mean Memory performance at Time 3 was .85 units greater than that at Time 2 and .72 units greater than Time 1, and Time 1 was .13 units greater than Time 2. Overall there was a slight, though insignificant increase from Time 1 to Time 3. The increase was expected since the sample consisted of adolescents who are still cognitively developing; thus, it would be expected that their memory would improve slighty over time.

8.2.2. Reaction time

The Reaction Time factor, with three indicators, achieved both configural and weak factorial invariance. Weak invariance was sufficient evidence to support the presence of measurement invariance of the Reaction Time factor, as strong invariance is often hard to achieve (Chen et al., 2005). The achievement of weak invariance means that the scale of measurement for the Reaction Time factor remained consistent over time. That is, each unit of change in the reaction time items was equivalent to one unit of change in the Reaction Time factor, regardless of the year of testing. Given strong invariance was not achieved, item intercepts were not invariant across time. Therefore latent mean analysis was not conducted, as any change in the score of the Reaction Time factor over time could not undeniably be interpreted as the 'true' score change. Hence, item scores may vary over time due to factors other than changes in the latent construct of Reaction Time. This is important to consider in the clinical application of ImPACT. Clinicians should keep this limitation in mind when interpreting scores: the Reaction Time composite score is not a perfect reflection of an individual's reaction time capability. This caveat is not new, both the developers of ImPACT in addition to the Concussion in Sport Consensus Group state that neuropsychological assessment scores should only be one of many pieces of information that inform clinical decisions regarding the management of concussion (ImPACT online, 2013; McCrory et al., 2013).

Differential stability was assessed within the metric model, as this was the highest level of invariance achieved. Differential stability between all three time-points was supported. The strongest correlation was reported between Time 1 and Time 2 (.84), followed by Time 2 and Time 3 (.76), and lastly Time 1 and Time 3 (.75). Therefore, the rank-order of individuals' score on the Reaction Time construct relative to others, was relatively consistent over the three baseline testing sessions. These findings are substantially better than the test-retest reliability reported for the Reaction Time composite in Schatz's (2010) study. He calculated a Pearson's r of .52 and an ICC of .68 over a two-year interval. Despite the shorter test-retest interval, Iverson, Lovell, and Collins (2003) produced comparable findings to the current study, with a Pearson's r of .79 over a period of 5.8 days (range $1 - 13$ days).

This study has contributed to the literature, as it is the first to have statistically tested the assumption of measurement invariance over time. This is an assumption of all statistical methods that have assessed the stability of ImPACT performance over time, yet until now this assumption has not been tested. Since measurement invariance was supported for both the Reaction Time and Memory factors, greater confidence can be had in the findings of previous studies, which have assessed change over time (e.g., Broglio et al., 2007; Elbin et al., 2011b; Schatz & Maerlender, 2013; Schatz, 2010). That is, because measurement invariance implies that the measurement precision of the ImPACT instrument is consistent over time, confidence can be had that any change in item scores is the result of the same quantity of change in the latent factor, regardless of time. Having confidence that any change in observed scores is reflective of change in the latent constructs of Memory and Reaction Time is important for the clinical use of ImPACT as clinicians are using information provided by ImPACT to inform their decisions regarding whether or not a player is ready to return to play.

Another practical implication of the current findings pertains to the frequency of baseline testing. At present there are no official guidelines regarding how frequent baseline testing should be conducted among adolescents. However one would assume testing should be relatively frequent given adolescence is a time of rapid cognitive development. The current findings suggest conducting baseline testing every two years would suffice, given no significant changes were observed during this period. This is consistent with recommendations of Elbin and colleagues (2011)

who also suggest a two year interval between repeated baseline testing for adolescents is sufficient. Therefore, if for some reason individuals' baseline data for the current year were missing or invalid, comparing post-injury testing to the previous year's baseline performance would yield clinically relevant results.

8.3. Cross-Country Invariance

The third aim of the current study was to validate the identified structure within a New Zealand sample and to assess whether the structure was invariant across New Zealand and South African male adolescents. The two-factor, eight-item structure was successfully validated among a New Zealand sample, evidenced by adequate model-fit statistics and sufficient factor loadings. The fact that the previously identified structure was upheld within a new sample of individuals from a different country provides robust evidence supporting the factor structure identified in this study opposed to alternative proposed structures such as the Allen and Gfeller model and the ImPACT scoring structure. However, the structure failed to achieve measurement invariance between the two countries, as configural invariance was the highest level of invariance achieved. The presence of configural invariance means that the number of factors and the item-factor pattern was the same for both New Zealand and South African samples. Additionally it implies the latent factors being measured are similar, although not identical in each country. As mentioned previously, at least weak factorial invariance is required for measurement invariance to be supported. Given neither weak nor strong factorial invariance was achieved, measurement invariance for ImPACT between the two countries was not present. Thus test items appear not to be measuring the same latent construct in both groups and the scale of measurement is different for the two countries. Therefore any comparisons made between the Memory and Reaction Time composites between these two countries would potentially yield inaccurate results and meaningless interpretations (Billet, 2003). Because measurement invariance was not present between the two countries, any differences found between means or other statistics may not be accurate, but might reflect systematic bias of response across countries (Steenkamp & Baumgartner, 1998). With the exclusion of the current study, all studies comparing ImPACT performance between countries (e.g., Shuttleworth-Edwards, Whitefield-Alexander, Radloff, Taylor, & Lovell, 2009; Tsushima, Oshiro, & Zimbra, 2008) have not statistically tested the assumption of measurement

invariance prior to performing group comparisons. Therefore their findings may be erroneous if this assumption is found to be incorrect.

Possible reasons for non-invariance may be due to methodological issues such as small sample size. While both New Zealand ($N = 109$) and South African samples, ($N = 116$) met the minimum requirement of 100 participants (Kline, 2011), a larger sample may have produced more favourable results. While it was possible to use a larger South African sample, it was decided to use a sample that was similar in size to that of the New Zealand sample given disproportionate sample sizes between groups can cause results of the multiple-group model to be biased toward the model of the group with the largest sample size (Brown, 2006). Because the chi-square is affected by sample size, the group with the larger samples' contribution to the overall chi-square will be greater than that of the group with the smaller sample size. Furthermore, any other CFA statistics based on the chi-square (e.g., CFI) or that are sensitive to sample size will be disproportionately influenced by the unequal sample sizes (Brown, 2006). An alternative explanation for the absences of invariance between New Zealand and South Africa may be that measurement bias is present in these items so that different aspects of Memory and Reaction Time are being assessed, or potentially the items are measuring different constructs in each country. However, this seems unlikely given the objective nature of memory and reaction item. Lastly, it may be that the New Zealand sample performed poorer than the South African participates. This explanation appears more probable given the noticeable differences of item means between the two countries (see Table 8 and Table 20).

This investigation of cross-country invariance was unique as it was the first to evaluate measurement invariance across different country populations. Previous studies comparing individuals from different countries performance on ImPACT have employed less sophisticated techniques to evaluate equivalence (e.g., Shuttleworth-Edwards et al., 2009; Tsushima et al., 2008). Given the lack of measurement invariance in the current study a shadow of doubt is cast on the conclusion of equivalence from those previous studies. However, this is purely a hypothesis and measurement invariance between the specific populations of previous comparative studies would need to be statistically evaluated. Nevertheless, the current findings imply that if ImPACT were to be used in New Zealand, New

Zealand specific norms would need to be developed given the measure's lack of invariance with the South African population. While ImPACT recommends an individual approach using baseline data, the Concussion in Sport Group states that there is not enough evidence to endorse baseline testing (McCrory et al., 2013). Therefore, it is essential to have normative data available for the population in which ImPACT is used to provide a comparison for post-injury performance and guide return-to-play decisions.

The fact that the measure was variant across countries highlights the importance of first validating the test within the culture, context, and/or cohort it is going to be used in and not simply to generalise findings or information gathered from one group to another of differing characteristics. ImPACT is currently used in several countries, in many of which the test is being administered in the native language (ImPACTonline, 2014). In light of the current findings it is important to develop local normative data in each country. This is unless measurement invariance can be demonstrated with the country whose norms they wish to use.

8.4. Limitations and Future Research

A limitation of the current study was that parcelled items were used as the lowest-order indicator variables in the CFA models. Parcelled items were used in place of individual item scores (i.e., raw scores) as the latter were not available from ImPACT. The use of parcelled items is controversial: if they are used, an essential assumption is that the items that make up the parcel are unidimensional, as multidimensionality within a parcel is problematic (Little, Cunningham, Shahar, & Widaman, 2002). Although we suspect this assumption to be true given the identical nature of test items within each parcel, since the raw scores were not available this could not be statistically tested in the current study. Furthermore, parcelled items can be problematic as they threaten the validity of findings. Parcelling compresses the specific variance and random error of each individual item, either eliminating it altogether or at least reducing it. The parcelled items typically share variance, and this is what is emphasised in the aggregated score. As a result, the model fit is improved and mis-specification can be obscured (Bagozzi & Edwards, 1998; Bagozzi & Heatherton, 1994). For example, an item could originally load onto both a Memory and Reaction Time factor. If it is then parcelled with other predominately

memory-loaded items, the aggregate or parcelled score may now only load on Memory and not Reaction Time. Thus, its association with Reaction Time is concealed and the model is mis-specified. Basically, parcelling items reduces item-specific error and improves model fit. Without access to the raw scores we cannot be sure of the extent to which model fit was inflated due to item parcelling. Although the current study was of the opinion that parcelling raw scores was appropriate, future research should, if possible, assess the unidimensionality of each parcel (subscale) score.

The sample was one of convenience and thus this limited the size and diversity of participants. In both the New Zealand and South African samples female participants were non-existent. This appears to be a common limitation in the sport concussion literature. Nevertheless, gender invariance of the ImPACT structure could not be tested, thus it is unknown whether the structure identified in the current study is applicable to female athletes. Furthermore, the sample was limited to a non-clinical population in that no participants had recently sustained a concussion. It is important that the structure also be validated among a concussed sample because, as Delis and colleagues (2003) suggest, the item-factor relationships may differ between healthy and concussed populations. It may, for example, be possible that the distinction between visual and verbal memory emerges when one is concussed. That is, although the visual and verbal memory items appear to share variance in the healthy brain, this may not be the case for a concussed brain. Thus it is essential for future research to attempt to validate the identified ImPACT structure among a sample of concussed athletes given the purpose of the battery is to assess the cognitive effects of concussion and monitor their resolution. Furthermore, given the ImPACT battery is currently used within diverse populations, such as differing countries, cultures, genders, and sporting levels, future research should attempt to validate the battery within each of these populations as research of this kind is currently minimal.

The most pressing issue for future research is to re-evaluate the ImPACT battery. An analysis of the content should be conducted to ensure the test items are measuring what they purport to measure. Following from this should be a re-evaluation of the scoring structure, specifically ensuring that the item-factor relationships are correct. And lastly, more complex processing speed items should be added so that this construct is sufficiently represented. The adequate representation of processing speed

is important if Salthouse's (1996) theory claim, that processing speed deficits underlie impairments in other cognitive domains, is true.

The current study assessed the longitudinal stability over a one and two year period. While this length of temporal interval provides valuable information regarding the required frequency of baseline testing, evaluation of stability over shorter time intervals and with concussed athletes is an important avenue for future research. The reason for this is that, following a concussion, ImPACT is administered repeatedly over a short period of time, usually once a week until symptoms resolve. Thus it is important to ascertain the stability of the measure over this time period, in addition to assessing for practice effects which are typically more prominent during shorter intervals. This is also important as test-retest reliability is used in the calculation of Reliable Change Indices and ImPACT uses these to determine if any change in performance is clinically meaningful. Thus it is important that test-retest reliability for these shorter intervals be accurate. One step in ensuring the accuracy of the test-retest calculation is by statistically testing the important assumption that measurement invariance is present.

8.5. Concluding Summary

Sport concussion is currently a topic of interest, evidenced by increased media attention and empirical research. The growth within the sport concussion literature has centered on the utility of neuropsychological tests in assessing and managing sport concussion yet few studies have examined the psychometric properties of such tests. Neuropsychological batteries for sport concussion are now computerised and, with sophisticated marketing techniques, there is now widespread use of such testing throughout America and other developed countries. The widespread use of neuropsychological testing for the assessment of sport concussion is not however adequately supported in the literature. The use of neuropsychological testing, specifically of one particular battery, ImPACT, appears to have spread faster than research regarding its utility can be produced. This study took a step back and investigated the fundamental properties that should be present in order for this widespread use of ImPACT to be supported.

The study achieved its three objectives of identifying ImPACT's factor structure and assessing the structure's measurement invariance both over time and across two

populations. A factor structure was identified and later confirmed (i.e., validated) which was inconsistent with ImPACT's current scoring structure. As a result, it was strongly recommended that ImPACT revise the items included in the battery in addition to the scoring structure. Specifically it was proposed that a measure of executive function and impulse control should be omitted and instead focus should be on including a sufficient number of items to accurately and reliably measure memory, complex processing speed, and simple processing speed (i.e., reaction time) as these appear to be the more salient areas affected by concussion. Furthermore the identified structure demonstrated measurement invariance over three time-points with one-year intervals. This has practical implications for baseline testing in that it suggests baseline testing conducted two-yearly would suffice. Lastly, the structure failed to demonstrate invariance between populations of two countries, New Zealand and South Africa, emphasising the importance of having population specific norms.

The current study advanced knowledge regarding the psychometric properties of the ImPACT structure and the stability of the structure over time and across different country populations. The most salient strength of this study was methodological. Not only did it highlight the limitations of previous studies, it overcame these limitations and represented a more rigorous investigation of factor structure, stability over time, and stability across different populations. This was achieved through structural equation modelling (SEM) techniques. If this thesis is to achieve anything, it is hoped that it will raise awareness of concussion in the sporting arena. It should encourage individuals to critically evaluate the empirical research regarding ImPACT or any other neuropsychological battery used in the assessment of sport concussion and not simply accept the information provided by the test-sellers at face value.

References

Alla, S., Sullivan, S. J., Hale, L., & McCrory, P. (2009). Self-report scales/checklists for the measurement of concussion symptoms: a systematic review. *British Journal of Sports Medicine*, *43*(1), 3–12. doi:10.1136/bjsm.2009.058339

Allen, B. J., & Gfeller, J. D. (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. doi:10.3109/02699052.2010.541897

American Academy of Neurology. (1997). Practice parameter: The management of concussion in sports (summary statement). *Neurology*, *48*(3), 581–585. doi:10.1212/WNL.48.3.581

American Psychiatric Association. (2000). Diagnostic *and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC.

Anderson, J., R. (2013). *Language, memory, and thought*. Psychology Press.

Arbuckle, J. L. (2012). IBM SPSS Amos 21 User's Guide. Amos Development Corporation, Crawfordville, FL.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters.

Ardila, A., Galeano, L. M., & Rosselli, M. (1998). Toward a model of neuropsychological activity. *Neuropsychology Review*, *8*(4), 171–190. doi:10.1023/A:1021618218943

Armstrong, C. L., & Morrow, L. (Eds.). (2010). *Handbook of medical neuropsychology: Applications of cognitive neuroscience*. New York: Springer.

Arnau, R. C., & Thompson, B. (2000). Second order confirmatory factor analysis of the WAIS-III. *Assessment*, *7*(3), 237–246. doi:10.1177/107319110000700304

Aubry, M., Cantu, R., Dvorak, J., Graf-Baumann, T., Johnston, K., Kelly, J., … Schamasch, P. (2002). Summary and agreement statement of the first International Conference on Concussion in Sport, Vienna 2001. *British Journal of Sports Medicine*, *36*(1), 6–7. doi:10.1136/bjsm.36.1.6

Azouvi, P., Jokic, C., Der Linden, M. V., Marlier, N., & Bussel, B. (1996). Working memory and supervisory control after severe closed-head injury. A study of dual task performance and random generation. *Journal of Clinical and Experimental Neuropsychology*, *18*(3), 317–337. doi:10.1080/01688639608408990

Baddeley, A., D., & Hitch, G. J. (1974). Working memory. In G. H. Bower, *The Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*(1), 45–87. doi:10.1177/109442819800100104

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1,* 35-67

Bailes, J. E. (2009). Sports-related concussion: what do we know in 2009-a neurosurgeon's perspective. *Journal of the International Neuropsychological Society*, *15*(4), 509–511. doi:10.1017/S1355617709090936

Bandalos, D. L. , & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Barr, W. B., & McCrea, M. (2011). Diagnosis and assessment of concussion. In F. M. Webbe (Eds.), *Sport neuropsychology* (pp. 80-102). New York, NY: Springer Publishing Company.

Barth, J. T., Alves, W. M., Ryan, T. V., Macciocchi, S. N., Rimel, R. W., Jane, J. A., & Nelson, W. E. (1989). Mild head injury in sports: Neuropsychological sequelae and recovery of function. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Mild head injury* (pp. 257–275). New York, NY: Oxford University Press.

Barth, J. T., Freeman, J. R., Broshek, D. K., & Varney, R. N. (2001). Acceleration-deceleration sport-related concussion: The gravity of it all. *Journal of Athletic Training*, *36*(3), 253. Retrieved from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC155415/

Bashore, T. R., & Richard, K. (2002). Older age, traumatic brain injury, and cognitive slowing: Some convergent and divergent findings. *Psychological Bulletin*, *128*(1), 151–198. doi:10.1037/0033-2909.128.1.151

Bazarian, J. J., Blyth, B., & Cimpello, L. (2006). Bench to bedside: Evidence for brain injury after concussion—Looking beyond the Computed Tomography scan. *Academic Emergency Medicine*, *13*(2), 199–214. doi:10.1197/j.aem.2005.07.031

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society*, *11*(04), 345–357. doi:10.1017/S1355617705050411

Bennett, P. C., Ong, B., & Ponsford, J. (2005). Assessment of executive dysfunction following traumatic brain injury: Comparison of the BADS with other clinical neuropsychological measures. *Journal of the International Neuropsychological Society*, *11*(05). doi:10.1017/S1355617705050721

Bentler, P. ., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. doi:10.1037/0033-2909.88.3.588

Berry, J. W., Poortinga, Y. H., & Pandey, J. (1997). *Handbook of Cross-cultural Psychology: Basic processes and human development*. Needham Heights, MA: Allex and Bacon.

Bird, Y. N., Waller, A. E., Alsop, J. C., Chalmers, D. J., & Gerrard, D. F. (1998). The New Zealand rugby injury and performance project: Epidemiology of a season of rugby. *British Journal of Sports Medicine, 32*(3), 319-325. doi:10.1136/bjsm.32.4.319

Boden, B. P., Tacchetti, R. L., Cantu, R. C., Knowles, S. B., & Mueller, F. O. (2007). Catastrophic head injuries in high school and college football players. *The American Journal of Sports Medicine*, *35*(7), 1075–1081. doi:10.1177/0363546507299239

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bowden, S. C., Carstairs, J. R., & Arthur, E. (1999). Confirmatory factor analysis of combined Wechsler Adult Intelligence Scale—Revised and Wechsler Memory Scale—Revised scores in a healthy community sample. *Psychological Assessment*, *11*(3), 339–344. doi:10.1037/1040-3590.11.3.339

Bradley Burton, D., Ryan, J. J., Axelrod, B. N., Schellenberger, T., & Richards, H. M. (2003). A confirmatory factor analysis of the WMS-III in a clinical sample with crossvalidation in the standardization sample. *Archives of Clinical Neuropsychology*, *18*(6), 629–641. doi:10.1016/S0887-6177(02)00149-X

Brevik, E. J. (2012). *Fluid Intelligence in a Norwegian sample of middle-aged and older adults.* (Masters Thesis, The University of Bergen). Retrieved from http://bora.uib.no/handle/1956/6431

Broglio, D. S. P., & Puetz, T. W. (2008). The effect of sport concussion on neurocognitive function, self-report symptoms and postural control. *Sports Medicine*, *38*(1), 53–67. doi:10.2165/00007256-200838010-00005

Broglio, S. P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training*, *42*(4), 509–514. Retreived from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC2140077/

Brolinson, P. G., Manoogian, S., McNeely, D., Goforth, M., Greenwald, R., & Duma, S. (2006). Analysis of linear head accelerations from collegiate football impacts. *Current Sports Medicine Reports, 5*(1), 23–28. doi:10.1007/s11932-006-0025-0

Brouwer, W. H., Ponds, R. W. H. M., Van Wolffelaar, P. C., & Van Zomeren, A. H. (1989). Divided attention 5 to 10 years after severe closed head injury. *Cortex*, *25*(2), 219–230. doi:10.1016/S0010-9452(89)80038-3

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Brown, T., A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bruce, J. M., & Echemendia, R. J. (2003). Delayed-onset deficits in verbal encoding strategies among patients with mild traumatic brain injury. *Neuropsychology*, *17*(4), 622–629. doi:10.1037/0894-4105.17.4.622

Buckendahl, C., & Plake, B. (2006). Evaluating Tests. In S.M Downing, & T. M. Haladyna, *Handbook of Test Development* (pp. 725–738). New Jersey, NJ: Lawrence Erlbaum Associates.

Byrne, B. M. (2001). *Structural equation modeling With AMOS: Basic concepts, applications, and programming.* New York: Taylor & Francis.

Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming,* (2nd ed.,). New York: Taylor & Francis.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. doi:10.1037/0033-2909.105.3.456

Cameron, K. L., Yunker, C. A., & Austin, M. C. (1999). A standardized protocol for the initial evaluation and documentation of mild brain injury. Journal of Athletic Training, 34, 34–42.

Cantu, R. C. (1998). Second-impact syndrome. *Clinics in Sports Medicine*, *17*(1), 37–44. doi:10.1016/S0278-5919(05)70059-4

Cantu, R. C. (2001). Posttraumatic retrograde and anterograde amnesia: pathophysiology and implications in grading and safe return to play. *Journal of Athletic Training*, *36*(3), 244. Retreived from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC155413/

Carroll, J., B. (1993). *Human cognitive abilities*. New York, NY: Cambridge University Press. Retrieved from http://www.langtoninfo.com/web_content/9780521382755_frontmatter.pdf

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

Cattell, R. B., & Burdsal, C. A. (1975). The radial parcel double factoring design: A solution to the item-vs-parcel controversy. *Multivariate Behavioral Research*, *10*(2), 165–179. doi:10.1207/s15327906mbr1002_3

Cavanaugh, J. C., & Blanchard-Fields, F. (2006). *Adult development and aging* (3rd ed.,). Belmont, CA: Thomson Higher Education.

Centre for the Study of Traumatic Encephalopathy. (2009). 18 year old high school football player. Retrieved from http://www.bu.edu/cste/case-studies/18-year-old//

Cernich, A., Reeves, D., Sun, W., & Bleiberg, J. (2007). Automated Neuropsychological Assessment Metrics sports medicine battery. *Archives of Clinical Neuropsychology*, *22*(1), 101–114. doi:10.1016/j.acn.2006.10.008

Chaytor, N., & Schmitter-Edgecombe, M. (2007). Fractionation of the dysexecutive syndrome in a heterogeneous neurological sample: Comparing the Dysexecutive Questionnaire and the Brock Adaptive Functioning Questionnaire. *Brain Injury*, *21*(6), 615–621. doi:10.1080/02699050701426949

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 465-504. doi: 10.1080/10705510701301834

Chen, H.-Y., Keith, T. Z., Chen, Y.-H., & Chang, B.-S. (2009). What does the WISC-IV measure? Validation of the scoring and CHC-based interpretative approaches. *Journal of Research in Education Science*, *54*(3), 85–108. Retreived from http://www.edubook.com.tw/tw/file/pdf/1099/5403/04.pdf

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471–492. doi:10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. doi:10.1207/S15328007SEM0902_5

Chiaravalloti, N. D., Christodoulou, C., Demaree, H. A., & DeLuca, J. (2003). Differentiating simple versus complex processing speed: Influence on new learning and memory performance. *Journal of Clinical and Experimental Neuropsychology*, *25*(4), 489–501. doi:10.1076/jcen.25.4.489.13878

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardised assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: Moving from bench to bedside in patients with traumatic brain injury. *Journal of Cognitive Neuroscience*, *18*(7), 1212–1222. doi:10.1162/jocn.2006.18.7.1212

Code of Ethics Review Group. (2012). *Code of Ethics for Psychologists Working in Aotearoa/New Zealand*. Wellingtion, New Zealand.

Collie, A., Darby, D., & Maruff, P. (2001). Computerised cognitive assessment of athletes with sports related head injury. *British Journal of Sports Medicine*, *35*(5), 297–302. doi:10.1136/bjsm.35.5.297

Collins CL, Micheli LJ, Yard EE, & Comstock R. (2008). Injuries sustained by high school rugby players in the United States, 2005-2006. *Archives of Pediatrics & Adolescent Medicine*, *162*(1), 49–54. doi:10.1001/archpediatrics.2007.1

Collins, M. W., Field, M., Lovell, M. R., Iverson, G., Johnston, K. M., Maroon, J., & Fu, F. H. (2003). Relationship between postconcussion headache and neuropsychological test performance in high school athletes. *The American Journal of Sports Medicine*, *31*(2), 168–173. Retreived from http://ajs.sagepub.com.ezproxy.massey.ac.nz/content/31/2/168.short

Congress of Neurosurgens. (1968). Clinical neurosurgery. Proceedings of the congress of neurological surgeons, San Juan, Puerto Rico, 1966. *British Journal of Surgery*, *55*(4), 319–320. doi:10.1002/bjs.1800550431

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–83. doi:10.1016/S0160-2896(01)00096-4

Cronbach, L., J. (1988). Five perspectives on validity argument. In H. Wainer & H.I. Braun (Eds.), *Test Validity* (pp. 3–17). New Jersey: Lawrence Erlbaum Associates.

Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*(4), 558–575. doi:10.1177/0022022112438397

Deary, I. J. (1993). Inspection time and WAIS-R IQ subtypes: A confirmatory factor analysis study. *Intelligence*, *17*(2), 223–236. doi:10.1016/0160-2896(93)90029-5

DeCoster, J. (2005). Scale Construction Notes (University of Alabama). Retrieved from http://www.stat-help.com/scale.pdf

DeLuca, J., & Kalmar, J., H. (2013). *Information processing speed in clinical populations*. New York, NY: Psychology Press.

Demaree, H. A., DeLuca, J., Gaudino, E. A., & Diamond, B. J. (1999). Speed of information processing as a key deficit in multiple sclerosis: Implications for rehabilitation. *Journal of Neurology, Neurosurgery & Psychiatry*, *67*(5), 661–663. doi:10.1136/jnnp.67.5.661

Denny-Brown, D., & Russell, W. R. (1940). Experimental cerebral concussion. *The Journal of Physiology*, *99*(1), 153. Retrieved from http://jp.physoc.org/content/99/1/153.full.pdf

Dick, R. W. (2009). Is there a gender difference in concussion incidence and outcomes? *British Journal of Sports Medicine*, *43*(1), 46–i50. doi:10.1136/bjsm.2009.058172

Dikmen, S. S., Ross, B. L., Machamer, J. E., & Temkin, N. R. (1995). One year psychosocial outcome in head injury. *Journal of the International Neuropsychological Society*, *1*(1), 67–77. doi:10.1017/S1355617700000126

Donders, J., Tulsky, D. S., & Zhu, J. (2001). Criterion validity of new WAIS–III subtest scores after traumatic brain injury. *Journal of the International Neuropsychological Society*, *7*(07), 892–898.

Duff, M. C. (2009). Management of sports-related concussion in children and adolescents. Retrieved from http://www.asha.org.ezproxy.massey.ac.nz/publications/leader/2009/090714/f090714a.htm

Echemendia, R. J., Putukian, M., Mackin, R. S., Julian, L., & Shoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury: *Clinical Journal of Sport Medicine*, *11*(1), 23–31. doi:10.1097/00042752-200101000-00005

Elbin, R. J., Schatz, P., & Covassin, T. (2011a). 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. doi:10.1177/0363546511417173

Ellemberg, D., Henry, L. C., Macciocchi, S. N., Guskiewicz, K. M., & Broglio, S. P. (2009). Advances in sport concussion assessment: From behavioral to brain imaging measures. *Journal of Neurotrauma*, *26*(12), 2365–2382. doi:10.1089/neu.2009.0906

Erlanger, D., M., Feldman, D., J., & Kutner, K. (1999). *Concussion resolution index*. New York, NY: Headminder Inc.

Espe-Pfeifer, P., & Wachsler-Felder, J. (2000). *Neuropsychological interpretations of objective psychological tests*. New York: NY, Springer.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 56–83. doi:10.1080/10705519909540119

Fazio, V. C., Lovell, M. R., Pardini, J. E., & Collins, M. W. (2007). The relation between post concussion symptoms and neurocognitive performance in concussed athletes. *NeuroRehabilitation*, *22*(3), 207–216. Retreived from http://iospress.metapress.com.ezproxy.massey.ac.nz/content/4325027837215802/

Field, A. (2006). *Research methods: reliability analysis.* Retrieved from http://www.statisticshell.com/docs/reliability.pdf

Field, A. (2009). *Discovering statistics using IBM SPSS statistics* (3rd ed.). New York, NY: Sage Publications, Inc.

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

Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.,). New York, NY: Guilford Press.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment*. New Jersey, NJ: John Wiley & Sons.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*(1-3), 1–34. doi: 10.1016/S0301-0511(00)00051-X

Gessel, L. M., Fields, S. K., Collins, C. L., Dick, R. W., & Comstock, R. D. (2007). Concussions among United States high school and collegiate athletes. *Journal of Athletic Training*, *42*(4), 495. Retreived from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC2140075/

Giza, C. C., & Hovda, D. A. (2001). The neurometabolic cascade of concussion. *Journal of Athletic Training*, *36*(3), 228. Retreived from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC155411/

Green, G. A., & Jordan, S. E. (1998). Are brain injuries a significant problem in soccer? *Clinics in Sports Medicine*, *17*(4), 795–809. doi:10.1016/S0278-5919(05)70120-4

Gronwall, D. (1991). Minor head injury. *Neuropsychology*, *5*(4), 253–265. doi:10.1037/0894-4105.5.4.253

Gronwall, D. M. A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills*, *44*(2), 367–373. doi:10.2466/pms.1977.44.2.367

Guskiewicz, K. M. (2003). Assessment of postural stability following sport-related concussion: *Current Sports Medicine Reports*, *2*(1), 24–30. doi:10.1249/00149619-200302000-00006

Guskiewicz, K. M., Marshall, S. W., Bailes, J., McCrea, M., Cantu, R. C., Randolph, C., & Jordan, B. D. (2005). Association between recurrent concussion and late-life cognitive impairment in retired professional football players. *Neurosurgery*, *57*(4), 719–726.

Guskiewicz, K. M., Riemann, B. L., Perrin, D. H., & Nashner, L. M. (1997). Alternative approaches to the assessment of mild head injury in athletes. *Medicine and Science in Sports and Exercise*, *29*(7), 213–221.

Guskiewicz, K. M., Ross, S. E., & Marshall, S. W. (2001). Postural stability and neuropsychological deficits after concussion in collegiate athletes. *Journal of Athletic Training*, *36*(3), 263–273.

Guskiewicz, K. M., McCrea, M., Marshall, S. W., Cantu, R. C., Randolph, C., Barr, W., . . . Kelly, J. P. (2003). Cumulative effects associated with recurrent concussion in collegiate football players: The NCAA concussion study. *Journal of the American Medical Association*, *290*(19), 2549–2555. doi:10.1001/jama.290.19.2549

Halstead, M. E., & Walter, K. D. (2010). Sport-related concussion in children and adolescents. *Pediatrics*, *126*(3), 597–615. doi:10.1542/peds.2010-2005

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. New York, NY: Sage Publications.

Hancock, G., & Samuelsen, K. M. (Ed.). (2008). *Advances in latent variable mixture models*. North Carolina, NC: Information Age Publishing Inc.

Harrington, D. (2008). *Confirmatory factor analysis*. New York, NY: Oxford University Press.

Hart, T., Giovannetti, T., Montgomery, M. W., & Schwartz, M. F. (1998). Awareness of errors in naturalistic action after traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *13*(5), 16–28. Retreived from http://journals.lww.com.ezproxy.massey.ac.nz/headtraumarehab/Abstract/1998/10000/Awareness_of_Errors_in_Naturalistic_Action_after.4.aspx

Hawkins, K. A. (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III technical manual clinical samples data: A preliminary approach to clinical utility. *The Clinical Neuropsychologist*, *12*(4), 535–551. doi:10.1076/clin.12.4.535.7236

Hicks, L. H., & Birren, J. E. (1970). Aging, brain damage, and psychomotor slowing. *Psychological Bulletin*, *74*(6), 377–396. doi:10.1037/h0033064

Hinton-Bayre, A., Geffen, G., & Friis, P. (2004). Presentation and mechanisms of concussion in professional Rugby League football. *Journal of Science and Medicine in Sport*, *7*(3), 400–404. doi:10.1016/S1440-2440(04)80035-5

Horn, J., L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *WJ-R technical manual* (pp. 197–232). Itasca, IL: Riverside Publishing.

Horsman, M. (2010). *Concussion in contact sport: Investigatin the neurocognitive profile of afrikaans adolescent rugby players*. (Doctoral dissertation, Rhodes University). Retreived from http://eprints.ru.ac.za/2264/1/HORSMAN-MSocSc-TR11-137.pdf

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation*

modeling: A Multidisciplinary Journal, *6*(1), 1–55. doi:10.1080/10705519909540118

IBM. (2006). *Item-total correlations in SPSS*. Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg21477000

ImPACTonline. (2013). *Immediate post-concussion assessment and cognitive test (ImPACT).* Retrieved from http://www.impacttest.com/about/

ImPACTonline. (2014). *Immediate post-concussion assessment and cognitive test (ImPACT).* Retrieved from http://www.impacttest.com/about/

Iverson, G. L., Brooks, B. L., Collins, M. W., & Lovell, M. R. (2006). Tracking neuropsychological recovery following concussion in sport. *Brain Injury*, *20*(3), 245–252. doi:10.1080/02699050500487910

Iverson, G. L., Franzen, M., D., Lovell, M. R., & Collins, M. W. (2003). Construct validity of ImPACT in athletes with concussion. *Archives of Clinical Neuropsychology,* *29*(7), 961-962.

Iverson, G. L., Gaetz, M., Lovell, M. R., & Collins, M. W. (2004). Cumulative effects of concussion in amateur athletes. *Brain Injury*, *18*(5), 433–443. doi:10.1080/02699050310001617352

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

Iverson, G. L., Lovell, M. R., & Collins, M. W. (2005). Validity of ImPACT for measuring processing speed following sports-related concussion. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 683–689. doi:10.1081/13803390490918435

Iverson, G., Lovell, M., & Collins, M. W. (2002). Validity of ImPACT for measuring the effects of sports-related concussion. *Archives of Clinical Neuropsychology*, *17*(8), 769.

Johnston, K. M., McCrory, P., Mohtadi, N. G., & Meeuwisse, W. (2001). Evidence-Based review of sport-related concussion: clinical science. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *11*(3), 150–159.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides, & R. E. Schumacker (Eds.), A*dvanced Structural Equation Modelling: Issues and Techniques.* (pp. 57-88). New Jersey, NJ: Lawrence Erlbaum Associates Inc.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2), 199–225. doi:10.1016/0001-6918(94)90003-5

Katayama, Y., Becker, D., P., Tamura, T., & Hovda, D., A. (2010). Massive increases in extracellular potassium and the indiscriminate release of glutamate following concussive brain injury. *Journal of Neurosurgery, 112*(2), 889-900. Retreived from *http://thejns.org/doi/abs/10.3171/jns.1990.73.6.0889@col.2*

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure? *School Psychology Review*, *35*(1), 108–127.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635–650. doi:10.1002/pits.20496

Kelly, J. P., & Rosenberg, J. H. (1997). Diagnosis and management of concussion in sports. *Neurology*, *48*(3), 575–580. doi:10.1212/WNL.48.3.575

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, *54*(3), 757–765. doi:10.1177/0013164494054003022

Kline, R., B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, *30*(1), 57–78. doi:10.1146/annurev.neuro.30.051606.094256

Koch, G., G. (2006). Intraclass correlation coefficient. *Encyclopedia of Statistical Sciences*, *6*, 213–217. doi: 10.1002/0471667196.ess1275.pub2

Kontos, A. P., Elbin, R. J., Covassin, T., & Larson, E. (2010). Exploring differences in computerized neurocognitive concussion testing between African American and white athletes. *Archives of Clinical Neuropsychology*, *25*(8), 734–744. doi:10.1093/arclin/acq068

Lang, D., A., Teasdale, G., M., Macpherson, P., & Lawrence, A. (2009). Diffuse brain swelling after head injury: more often malignant in adults than children? *Journal of Neurosurgery,* 80(4), 675-680. doi:10.3171/jns.1994.80.4.0675

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

Larrabee, G. J., & Curtiss, G. (1995). Construct validity of various verbal and visual memory tests. *Journal of Clinical and Experimental Neuropsychology*, *17*(4), 536–547. doi:10.1080/01688639508405144

Larrabee, G. J., Kane, R. L., & Schuck, J. R. (1983). Factor analysis of the WAIS and Wechsler Memory Scale: An analysis of the construct validity of the wechsler

memory scale. *Journal of Clinical Neuropsychology*, *5*(2), 159–168. doi:10.1080/01688638308401162

Lau, B. C., Collins, M. W., & Lovell, M. R. (2011). Sensitivity and specificity of subacute computerized neurocognitive testing and symptom evaluation in predicting outcomes after sports-related concussion. *The American Journal of Sports Medicine*, *39*(6), 1209–1216. doi:10.1177/0363546510392016

Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59–66. doi:10.1080/00031305.1988.10475524

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. R. (1992). Factor structure of the Wechsler Memory Scale—Revised within a comprehensive neuropsychological battery. *Neuropsychology*, *6*(3), 239.

Levin, H. S., Benton, A. L., & Grossman, R., G. (1982). *Neurobehavioural consequences of closed head injury*. New York, NY: Oxford University Press.

Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*(1-4), 281–297. doi:10.1080/00207598208247445

Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). New York, NY: Oxford University Press.

Lezak, M. D., Howieson, D., B., & Loring, D., W. (2004). *Neuropsychological Assessment* (4th ed.,). New York, NY: Oxford University Press.

Little, T. D. (1997). Mean and covariance structures analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*(1), 53–76. doi:10.1207/s15327906mbr3201_3

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173. doi:10.1207/S15328007SEM0902_1

Littleton, A., & Guskiewicz, K. (2013). Current concepts in sport concussion management: A multifaceted approach. *Journal of Sport and Health Science*, *2*(4), 227–235. doi:10.1016/j.jshs.2013.04.003

Lovell, M. R. (1999). Evaluation of the professional athlete. In J. E. Bailes, M. Lovell, & J. Maroon, *Sports-Related Concussion* (pp. 200–214). New York, NY: CRC Press.

Lovell, M. R., & Burke, C., J. (2000). Concussion management in professional hockey. In R. Cantu (Ed.), *Neurological athletic head and spine injury*. (pp. 109–116). Philadelphia: WB Saunders.

Lovell, M. R., & Collins, M. W. (2002). New developments in the evaluation of sports-related concussion: *Current Sports Medicine Reports*, *1*(5), 287–292. doi:10.1249/00149619-200210000-00006

Lovell, M. R., Collins, M. W., Iverson, G. L., Field, M., Maroon, J. C., Cantu, R., … Fu, F. H. (2009). Recovery from mild concussion in high school athletes. *Journal of Neurosurgery,* 98(2). Retreived from http*://dx.doi.org/10.3171/jns.2003.98.2.0296*.

Lovell, M. R., Iverson, G. L., Collins, M. W., McKeag, D., & Maroon, J. C. (1999). Does loss of consciousness predict neuropsychological decrements after concussion? *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *9*(4), 193–198. Retreived from http://journals.lww.com.ezproxy.massey.ac.nz/cjsportsmed/Abstract/1999/10000/Does_Loss_of_Consciousness_Predict.2.aspx

Lovell, M. R., Iverson, G. L., Collins, M. W., Podell, K., Johnston, K. M., Pardini, D., … Maroon, J. C. (2006). Measurement of symptoms following sports-related concussion: Reliability and normative data for the post-concussion scale. *Applied Neuropsychology*, *13*(3), 166–174. doi:10.1207/s15324826an1303_4

Lovell, M. R., & Pardini, J., E. (2010). Neuropsychological assessment and sports-related mild traumatic brain injury (concussion). In T. D. Marcotte, & I. Grant (Eds.), *Neuropsychology of everyday functioning* (pp. 331–356). New York, NY: Guilford Press.

Lovell, M. R., & Solomon, G. S. (2011). Psychometric data for the NFL neuropsychological test battery. *Applied Neuropsychology*, *18*(3), 197–209. doi:10.1080/09084282.2011.595446

Lum, J. A. G., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154. doi:10.1016/j.cortex.2011.06.001

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. doi:10.1037/1082-989X.4.1.84

Macciocchi, S. N., Barth, J. T., Alves, W., Rimel, R. W., & Jane, J. A. (1996). Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery*, *39*(3), 510–514. Retreived from http://journals.lww.com.ezproxy.massey.ac.nz/neurosurgery/Abstract/1996/09000/Neuropsychological_Functioning_and_Recovery_after.14.aspx

Maillard-Wermelinger, A., Yeates, K. O., Taylor, H. G., Rusin, J., Bangert, B., Dietrich, A., … Wright, M. (2009). Mild traumatic brain injury and executive functions in school-aged children. *Developmental Neurorehabilitation*, *12*(5), 330–341. doi:10.3109/17518420903087251

Majerske, C. W., Mihalik, J. P., Ren, D., Collins, M. W., Reddy, C. C., Lovell, M. R., & Wagner, A. K. (2008). Concussion in sports: Postconcussive activity levels,

symptoms, and neurocognitive performance. *Journal of Athletic Training*, *43*(3), 265. doi: 10.4085/1062-6050-43.3.265

Makdissi, M. (2009). Is the simple versus complex classification of concussion a valid and useful differentiation? *British Journal of Sports Medicine*, *43*(1), 23–27. doi:10.1136/bjsm.2009.058206

Makdissi, M., Collie, A., Maruff, P., Darby, D. G., Bush, A., McCrory, P., & Bennell, K. (2001). Computerised cognitive assessment of concussed Australian Rules footballers. *British Journal of Sports Medicine*, *35*(5), 354–360. doi:10.1136/bjsm.35.5.354

Marar, M., McIlvain, N. M., Fields, S. K., & Comstock, R. D. (2012). Epidemiology of concussions among United States high school athletes in 20 sports. *American Journal of Sports Medicine*, *40*(4), 747–755. doi:10.1177/0363546511435626

Maroon, J. C., Lovell, M. R., Norwig, J., Podell, K., Powell, J., & Harlt, R. (2000). Cerebral concussion in athletes: evaluation and neuropsychological testing. *Neurosurgery, 47*, 659–69

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: comparison of simplex models and one-factor models. *Journal of Educational Measurement*, *30*(2), 157–183. doi:10.1111/j.1745-3984.1993.tb01072.x

Marshall, S. W., & Spencer, R. J. (2001). Concussion in rugby: The hidden epidemic. *Journal of Athletic Training*, *36*(3), 334. Retreived from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC155428/

Martin, T. A., Donders, J., & Thompson, E. (2000). Potential of and problems with new measures of psychometric intelligence after traumatic brain injury. *Rehabilitation Psychology*, *45*(4), 402–408. doi:10.1037/0090-5550.45.4.402

Matser, E. T, Kessels, A. G, Lezak, M. D, Jordan, B. D, & Troost, J. (1999). Neuropsychological impairment in amateur soccer players. *Journal of the American Medical Association, 282*(10), 971–973. doi:10.1001/jama.282.10.971

Mayers, L. (2008). Return-to-play criteria after athletic concussion: A need for revision. *Archives of Neurology*, *65*(9), 1158–1161. doi:10.1001/archneur.65.9.1158

Mayers, L. B., & Redick, T. S. (2012a). Authors' reply to "Response to Mayers and Redick: Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues." *Journal of Clinical and Experimental Neuropsychology*, *34*(4), 435–442. doi:10.1080/13803395.2012.667790

Mayers, L. B., & Redick, T. S. (2012b). Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues. *Journal of Clinical and Experimental Neuropsychology*, *34*(3), 235–242. doi:10.1080/13803395.2011.630655

McCaffrey, M. A., Mihalik, J. P., Crowell, D. H., Shields, E. W., & Guskiewicz, K. M. (2007). Measurement of head impacts in collegiate football players: clinical measures of concussion after high- and low-magnitude impacts. *Neurosurgery*, *61*(6), 1236–1243. doi:10.1227/01.neu.0000306102.91506.8b

McClincy, M. P., Lovell, M. R., Pardini, J., Collins, M. W., & Spore, M. K. (2006). Recovery from sports concussion in high school and collegiate athletes. *Brain Injury*, *20*(1), 33–39. doi:10.1080/02699050500309817

McCrea, M, Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., … Kelly, J. P. (2003). Acute effects and recovery time following concussion in collegiate football players: The ncaa concussion study. *Journal of the American Medical Association*, *290*(19), 2556–2563. doi:10.1001/jama.290.19.2556

McCrea, M., Guskiewicz, K., Randolph, C., Barr, W. B., Hammeke, T. A., Marshall, S. W., … Kelly, J. P. (2013). Incidence, clinical course, and predictors of prolonged recovery time following sport-related concussion in high school and college athletes. *Journal of the International Neuropsychological Society*, *19*(1), 22–33. doi:10.1017/S1355617712000872

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. (2004). Unreported concussion in high school football players: implications for prevention. *Clinical Journal of Sport Medicine*, *14*(1), 13–17. Retreived from http://journals.lww.com.ezproxy.massey.ac.nz/cjsportsmed/Abstract/2004/01000/Unreported_Concussion_in_High_School_Football.3.aspx

McCrea, M., Kelly, J. P., Kluge, J., Ackley, B., & Randolph, C. (1997). Standardized assessment of concussion in football players. *Neurology*, *48*(3), 586–588. doi:10.1212/WNL.48.3.586

McCrea, M., Kelly, J. P., Randolph, C., Cisler, R., Berger, L., Marshall, L. F., … Cantu, R. C. (2002). Immediate neurocognitive effects of concussion. *Neurosurgery*, *50*(5), 1032–1042. Retreived from http://www.cdc.gov/ncipc/pub-res/tbi_toolkit/mccrae/discussion.htm

McCrory, P. (2001). Does second impact syndrome exist? *Clinical Journal of Sport Medicine*, *11*(3), 144–149. doi:10.1097/00042752-200107000-00004

McCrory, P., Johnston, K., Meeuwisse, W., Aubry, M., Cantu, R., Dvorak, J., … Schamasch, P. (2005). Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *British Journal of Sports Medicine*, *39*(4), 196–204. doi:10.1136/bjsm.2005.018614

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvořák, J., Echemendia, R. J., … Turner, M. (2013). Consensus statement on concussion in sport: the 4th International Conference on Concussion in Sport held in Zurich, November 2012.

British Journal of Sports Medicine, *47*(5), 250–258. doi:10.1136/bjsports-2013-092313

McCrory, P., Meeuwisse, W., Johnston, K., Dvorak, J., Aubry, M., Molloy, M., & Cantu, R. (2009). Consensus Statement on Concussion in Sport - The third International Conference on Concussion in Sport Held in Zurich, November 2008. *The Physician and Sportsmedicine*, *37*(2), 141–159. doi:10.3810/psm.2009.06.1721

McCrory, P. R., & Berkovic, S. F. (1998). Second impact syndrome. *Neurology*, *50*(3), 677–683. doi:10.1212/WNL.50.3.677

McDonald, J. W., & Johnston, M. V. (1990). Physiological and pathophysiological roles of excitatory amino acids during central nervous system development. *Brain Research Reviews*, *15*(1), 41–70. doi:10.1016/0165-0173(90)90011-C

McDowell, S., Whyte, J., & D'Esposito, M. (1997). Working memory impairments in traumatic brain injury: evidence from a dual-task paradigm. *Neuropsychologia*, *35*(10), 1341–1353. doi:10.1016/S0028-3932(97)00082-1

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–119). New York, NY, US: Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 136–181). New York, NY: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. doi:10.1016/j.intell.2008.08.004

McGrew, K. S., Werder, J., K., & Woodcock, R., W. (1991). *The WJ-R technical manual*. Chicago: Riverside Publishing.

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

Meade, A. W., Johnson, E. C., Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. J*ournal of Applied Psychology, 93*(3), 568-592. doi: 10.1037/0021-9010.93.3.568

Meade, A. W., & Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods*, *9*(3), 369–403. doi:10.1177/1094428105283384

Meehan, W. P., & Bachur, R. G. (2009). Sport-related concussion. *Pediatrics*, *123*(1), 114–123. doi:10.1542/peds.2008-0309

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi:10.1007/BF02294825

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11), 69–77. doi:10.1097/01.mlr.0000245438.73837.89

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. doi:10.1037/0003-066X.50.9.741

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111–130. Retreived from http://mvint.usbmed.edu.co:8002/ojs/index.php/web/article/view/465/469

Miller, E. (1970). Simple and choice reaction time following severe head injury. *Cortex*, *6*(1), 121–127. doi:10.1016/S0010-9452(70)80040-5

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. doi:10.1037/h0043158

Miller, J. R., Adamson, G. J., Pink, M. M., & Sweet, J. C. (2007). Comparison of preseason, midseason, and postseason neurocognitive scores in uninjured collegiate football players. *The American Journal of Sports Medicine*, *35*(8), 1284–1288. doi:10.1177/0363546507300261

Morris, T. (2010). Traumatic brain injury. In C. L. Armstrong & L. Morrow (Eds.), *Handbook of Medical Neuropsychology* (pp. 17–32). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-1364-7_2

Moser, R. S., Iverson, G. L., Echemendia, R. J., Lovell, M. R., Schatz, P., Webbe, F. M., … Barth, J. T. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology*, *22*(8), 909–916. doi:10.1016/j.acn.2007.09.004

Moser, R. S., Schatz, P., & Jordan, B. D. (2005). Prolonged effects of concussion in high school athletes. *Neurosurgery*, *57*(2), 300–306. Retrieved from http://www.sccnj.com/articles/MoserSchatzJordan.pdf

Nasser, F., & Takahashi, T. (2003). The effect of using item parcels on ad hoc goodness-of-fit indexes in confirmatory factor analysis: An example using sarason's reactions to tests. *Applied Measurement in Education*, *16*(1), 75–97. doi:10.1207/S15324818AME1601_4

Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement*, *63*(5), 729–757. doi:10.1177/0013164403258228

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice* (Vol. xvii). New Jersey, NJ: Lawrence Erlbaum Associates Publishers.

Nunnally, J. C., & Bernstein, I. H. (1991). *Psychometric theory*. New York, NY: McGraw.

Olvey, S. (2002, July). *Concussion in motorsports*. Paper presented at the New Developments in Sports-related Concussion Conference, Pittsburgh, Pennsylvania.

Pardini, D., Stump, J. E., Lovell, M. R., Collins, M. W., Moritz, K. & Fu, F. (2004). The post-concussion symptom scale (PCSS): a factor analysis. *British Journal of Sports Medicine*, *38*(5), 661–662. Retreived from http://bjsm.bmj.com/content/38/5/654.full

Parker, K. (1983). Factor analysis of the WAIS-R at nine age levels between 16 and 74 years. *Journal of Consulting and Clinical Psychology*, *51*(2), 302–308. doi:10.1037/0022-006X.51.2.302

Patel, D. D. R., Shivdasani, V., & Baker, R. J. (2005). Management of sport-related concussion in young athletes. *Sports Medicine*, *35*(8), 671–684. doi:10.2165/00007256-200535080-00002

Pellman, E. J., Lovell, M. R., Viano, D. C., & Casson, I. R. (2006). Concussion in professional [American] football: recovery of NFL and high school athletes assessed by computerized neuropsychological testing-part 12. *Neurosurgery*, *58*(2), 263–74. doi:10.1227/01.NEU.0000200272.56192.62

Pellman, E. J., Lovell, M. R., Viano, D. C., Casson, I. R., & Tucker, A. M. (2004). Concussion in professional football: Neuropsychological testing. *Neurosurgery, 55*(6) 1290–1305. doi:10.1227/01.NEU.0000149244.97560.91

Pellman, E. J., Viano, D. C., Tucker, A., M., Casson, I. R., & Waeckerle, J., F. (2003). Concussion in professional football: reconstruction of game impacts and injuries. *Neurosurgery*, *53*(4), 799–812.

Pickles, W. (1950). Acute general edema of the brain in children with head injuries. *New England Journal of Medicine*, *242*(16), 607–611. doi:10.1056/NEJM195004202421602

Powell, J. W., & Barber-Foss, K. D. (1999). Injury patterns in selected high school sports: A review of the 1995-1997 seasons. *Journal of Athletic Training*, *34*(3), 277–284. Retrieved from http://www-ncbi-nlm-nih-gov.ezproxy.massey.ac.nz/pmc/articles/PMC1322923/

Puga, P. A. (2011). Unreported concussion in high school football players. (Masters Dissertation, University of Texas). Retreived from http://digitalcommons.utep.edu/dissertations/AAI1494533

Randolph, C., Millis, S., Barr, W. B., McCrea, M., Guskiewicz, K. M., Hammeke, T. A., & Kelly, J. P. (2009). Concussion symptom inventory: an empirically derived scale for

monitoring resolution of symptoms following sport-related concussion. *Archives of Clinical Neuropsychology*, 25, 57-69.

Raykov, T., & Marcoulides, G. A. (2006). A first course in structural equation modeling (2nd ed.,). New Jersey, NJ: Lawrence Erlbaum Associates.

Rees, R. J., & Bellon, M. L. (2007). Post concussion syndrome ebb and flow: longitudinal effects and management. *NeuroRehabilitation*, *22*(3), 229–242. Retreived from http://web.a.ebscohost.com.ezproxy.massey.ac.nz/ehost/pdfviewer/pdfviewer?sid=c 415dbf2-3e40-4bf2-b169-460a44469e94%40sessionmgr4002&vid=2&hid=4209

Reeves, D. L., Winter, K. P., Bleiberg, J., & Kane, R. L. (2007). ANAM genogram: historical perspectives, description, and current endeavors. *Archives of Clinical Neuropsychology*, *22*(1), 15–37. doi:10.1016/j.acn.2006.10.013

Reitan, R. M., & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review*, *4*(3), 161–198. doi:10.1007/BF01874891

Rogers, S. A., Kang, C. H., & Miller, K. J. (2007). Cognitive profiles of aging and aging-related conditions. *Aging Health*, *3*(4), 457–470. doi:10.2217/1745509X.3.4.457

Roid, G., H. (2003). Standford-Binet Inelligence Scale (5th ed.,). Itasca, IL: Riverside Publishing.

Roman, D. D., Edwall, G. E., Buchanan, R. J., & Patton, J. H. (1991). Extended norms for the paced auditory serial addition task. *Clinical Neuropsychologist*, *5*(1), 33–40. doi:10.1080/13854049108401840

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. doi:10.1016/j.neuroimage.2011.11.050

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in. *Personality and Social Psychology Bulletin*, *28*(12), 1629–1646. doi:10.1177/014616702237645

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. doi:10.1037/0033-295X.103.3.403

Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the differential ability scales and the Woodcock-Johnson Tests of Cognitive Abilities–Third Edition. *Psychology in the Schools*, *44*(2), 119–138. doi:10.1002/pits.20211

Saunders RL, & Harbaugh RE. (1984). The second impact in catastrophic contact-sports head trauma. *Journal of American Medical Association*, *252*(4), 538–539. doi:10.1001/jama.1984.03350040068030

Savalei, V., & Bentler, P. M. (2010). Structural equation modeling. In I. B. Weiner, & W. Edward Craighead, *The Corsini Encyclopedia of Psychology* (Vol. II). New Jersey, NJ: John Wiley & Sons, Inc. doi: 10.1002/9780470479216.corpsy0953

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

Schatz, P., Kontos, A., & Elbin, R. (2012). Response to Mayers and Redick: "Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: Psychometric issues." *Journal of Clinical and Experimental Neuropsychology*, *34*(4), 428–434. doi:10.1080/13803395.2012.667789

Schatz, P., & Maerlender, A. (2013). A two-factor theory for concussion assessment using ImPACT: memory and speed. *Archives of Clinical Neuropsychology*, *28*(8), 791–797. doi:10.1093/arclin/act077

Schatz, P., Pardini, J. E., Lovell, M. R., Collins, M. W., & Podell, K. (2006). Sensitivity and specificity of the ImPACT test battery for concussion in athletes. *Archives of Clinical Neuropsychology*, *21*(1), 91–99. doi:10.1016/j.acn.2005.08.001

Schatz, P., & Putz, B. O. (2006). Cross-validation of measures used for computer-based assessment of concussion. *Applied Neuropsychology*, *13*(3), 151–159. doi:10.1207/s15324826an1303_2

Schatz, P., & Zillmer, E. A. (2003). Computer-based assessment of sports-related concussion. *Applied Neuropsychology*, *10*(1), 42–47. doi:10.1207/S15324826AN1001_6

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. doi:10.1016/j.hrmr.2008.03.003

Schneider, R., C. (1973). *Head and neck injuries in football: meachanisms, treatment, and prevention.* Baltimore: Williams & Wilkins.

Schneider, W., J., & McGrew, K., S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, J. Genshaft, & P. L. Harrison, *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford Press.

Schoenberg, M. R., & Scott, J. G. (2011). *The Little Black Book of Neuropsychology: A Syndrome-Based Approach*. New York, NY: Springer.

Scott, J. G. (2011). Attention/concentration: The distractible patient. In M. R. Schoenberg & J. G. Scott (Eds.), *The Little Black Book of Neuropsychology: A Syndrome-Based Approach* (pp. 149-158). New York, NY: Springer.

Scott, J. G., & Schoenberg, M. R. (2011). Memory and learning: The forgetful patient. In M. R. Schoenberg & J. G. Scott (Eds.), *The Little Black Book of Neuropsychology* (pp. 179–200). New York, NY: Springer.

Segall, M. H. (1986). Culture and behavior: Psychology in global perspective. *Annual Review of Psychology*, *37*(1), 523–564. doi:10.1146/annurev.ps.37.020186.002515

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross- cultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 93–119). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Sherman, E. M. S., Strauss, E., & Spellacy, F. (1997). Validity of the paced auditory serial addition test (pasat) in adults referred for neuropsychological assessment after head injury. *The Clinical Neuropsychologist*, *11*(1), 34–45. doi:10.1080/13854049708407027

Shuttleworth-Edwards, A. B., Whitefield-Alexander, V. J., Radloff, S. E., Taylor, A. M., & Lovell, M. R. (2009). Computerized neuropsychological profiles of South African versus US athletes: a basis for commentary on cross-cultural norming issues in the sports concussion arena. *The Physician and Sportsmedicine*, *37*(4), 45–52. doi:10.3810/psm.2009.12.1741

Shuttleworth-Edwards, A. B., & Radloff, S. E. (2008). Compromised visuomotor processing speed in players of Rugby Union from school through to the national adult level. *Archives of Clinical Neuropsychology*, *23*(5), 511–520. doi:10.1016/j.acn.2008.05.002

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. doi:10.1007/s11336-008-9101-0

Sousa, K. H., & Chen, F. F. (2003). Latent mean structure analysis: a technique to compare group differences in a second-order quality of life model. *Journal of Nursing Measurement*, *11*(3), 253–266. doi: http://dx.doi.org/10.1891/jnum.11.3.253.61272

Spearman, C. (1927). *The abilities of man* (Vol. xxiii). Oxford, England: Macmillan.

Spikman, J. M., van Zomeren, A. H., & Deelman, B. G. (1996). Deficits of attention after closed-head injury: slowness only? *Journal of Clinical and Experimental Neuropsychology*, *18*(5), 755–767. doi:10.1080/01688639608408298

Sport New Zealand. (2001). *SPARC facts 1997-2001*. Retreived from http://www.activenzsurvey.org.nz/Documents/sparc-facts-1997-2001.pdf

SPSS Inc. (2011). IBM SPSS Statistics for Windows, Version 20.0. New York, NY: IBM Corp.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross‑national consumer research. *Journal of Consumer Research*, *25*(1), 78–107. doi:10.1086/209528

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (5th ed.,). New York, NY; Taylor & Francis.

Stone, B. J. (1992). Joint confirmatory factor analyses of the DAS and WISC-R. *Journal of School Psychology*, *30*(2), 185–195. doi:10.1016/0022-4405(92)90030-9

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

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. doi:10.1207/S15327752JPA8001_18

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi:10.1037/h0054651

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.,). New Jersey, NJ: Pearson Education, Limited.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock, & R. O. Mueller, *Structural equation modelling: A second course* (pp. 119–169). Carolina, NC: Information Age Publishing Inc.

Thurman, D. J., Branche, C. M., & Sniezek, J. E. (1998). The epidemiology of sports-related traumatic brain injuries in the United States: recent developments. *The Journal of Head Trauma Rehabilitation*, *13*(2), 1–8. doi:10.1097/00001199-199804000-00003

Tsushima, W. T., Oshiro, R., & Zimbra, D. (2008). Neuropsychological test performance of Hawai'i high school athletes: Hawai'i ImPACT normative data. *Hawaii Medical Journal*, *67*(4), 93–95. Retreived from http://www.hjmph.org/HMJ_Apr08.pdf#page=5

Turkstra, L. S., & Byom, L. J. (2010). Executive functions and communication in adolescents. Retrieved from http://www.asha.org/publications/leader/2010/101221/executive-functions-and-communication-in-adolescents.htm

Ullman, J. B., & Bentler, P. M. (2012). Structural equation modeling. In I. B. Weiner, J. A. Schinka, & W. F. Velicer, *Handbook of Psychology* (2nd ed., pp. 661-690). New Jersey, NJ: John Wiley & Sons, Inc. doi: 10.1002/9781118133880

Van Kampen, D. A., Lovell, M. R., Pardini, J. E., Collins, M. W., & Fu, F. H. (2006). The "value added" of neurocognitive testing after sports-related concussion. *The*

American Journal of Sports Medicine, *34*(10), 1630–1635. doi:10.1177/0363546506288677

Van Zomeren, A. H., & Brouwer, W. H. (1994). *Clinical neuropsychology of attention* (Vol. x). New York, NY, US: Oxford University Press.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, *5*(2), 139–158. doi:10.1177/1094428102005002001

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. doi:10.1177/109442810031002

Wall, S. E., Williams, W. H., Cartwright-Hatton, S., Kelly, T. P., Murray, J., Murray, M., … Turner, M. (2006). Neuropsychological dysfunction following repeat concussions in jockeys. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(4), 518–520. doi:10.1136/jnnp.2004.061044

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale-III*. San Antonio, TX: The Psychological Corporation

Wechsler, D. (2003). *WISC-IV technical and interpertive manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *WAIS-IV administration and scoring manual*. San Antonio, TX: Psychological Coporation.

Wechsler D. (2009). Wechsler Memory Scale–Fourth Edition. San Antonio, TX: Pearson.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Californa, CA: Sage Publications, Inc.

White, P. E., Newton, J. D., Makdissi, M., Sullivan, S. J., Davis, G., McCrory, P., … Finch, C. F. (2013). Knowledge about sports-related concussion: is the message getting through to coaches and trainers? *British Journal of Sports Medicine*, *0*, 1-7. doi:10.1136/bjsports-2013-092785

Whitefield, V., J. (2006). *Glory is temporary, brain injury may be forever: a neuropsychological study on the cumulative effects of sports-related concussive brain injury amongst Grade 12 school boy athlete*. (Doctoral dissertation, Rhodes University). Retreived from http://contentpro.seals.ac.za/iii/cpro/DigitalItemViewPage.external?sp=1004471

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Wilk, C. M., Gold, J., M., Bartko, J., J., Dickerson, F., Fenton, W., S., Knable, M., … Buchanan, W. (2002). Test-retest stability of the repeatable battery for the assessment of neuropsychological status in schizophrenia. *American Journal of Psychiatry*, *159*(5), 838–844. doi:10.1176/appi.ajp.159.5.838

Wilkins, J. C., Valovich McLeod, T. C., Perrin, D. H., & Gansneder, B. M. (2004). Performance on the Balance Error Scoring System decreases after fatigue. *Journal of Athletic Training*, *39*(2), 156–161. doi: 10.1177/1941738112467755

Woodcock, R. W., McGrew, K., & Mather, N. (2001). Woodcock-Johnson Tests of Cognitive Abilities and Tests of Achievement (3rd ed.,). Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K., Mather, N., & Schrank, F. (2003). Diagnostic Supplement to the WJ III Tests of Cognitive Abilities. Rolling Meadows, IL: Riverside.

Appendix A

MASSEY UNIVERSITY
CENTRE FOR PSYCHOLOGY

Dear [Name of Principal]

Re: Request to conduct research at [Name of school]

My name is Shannon Martin and I am completing my Doctorate of Clinical Psychology at Massey University. My research dissertation focuses on the assessment and management of sport concussion among New Zealand high school athletes. I'm writing to your school to see if you would be willing to support this research by participating in this study. Please take five minutes to read the attached document outlining what I propose to do and how it will benefit adolescent athletes.

Concussion is defined as a disturbance of the brain after a blow to, or violent shaking of, the head. A concussion typically results in cognitive, physical, or emotional symptoms. Contrary to popular belief loss of consciousness is not the hallmark symptom of concussion in fact it occurs in only 10% of cases. More common symptoms include confusion, headaches, nausea, and amnesia. Usually these symptoms will resolve spontaneously within 14 days. However in some cases long-term effects are observed, especially following multiple concussions.

Sport concussion is a serious injury which has ended professional sporting careers and in several cases has resulted in death. For instance, Steve Devine an Auckland Blues player retired following serious impairments as a result of multiple concussions. He suffered severe migraines, extreme fatigue, and lack of concentration, for many years. Concussions do not only affect professional players, in fact children and adolescents, whose brains are still developing, are at higher risk of concussion than adults. Recently a Northland teenager, Darryl Sabin was left in a critical condition following multiple concussions. Prior to the final concussion his father was so worried he took out a court injunction to stop his son from playing. However, due to the clubs lack of concussion management procedures Darryl convinced the club to allow him to play. Had they adhered to appropriate concussion protocols this life threatening incident could have been avoided.

The Immediate Post-Concussion Assessment and Cognitive Testing battery (ImPACT), developed at the University of Pittsburgh, is a tool commonly used to assess concussion. ImPACT is computerised and uses neuropsychological tests to measure the cognitive effects of concussion. It is currently used by both the National Football League and National Hockey League in American as well as the Super 15 rugby teams in New Zealand. However, the applicability of ImPACT among New Zealand adolescents is yet to be tested. This is important as due to cultural differences we cannot assume that ImPACT will be a reliable and valid measure when used within New Zealand setting. Consequently, the aim of this research is to compare the performance of New Zealand and US adolescent's on ImPACT so as to establish whether or not ImPACT is suitable to use with New Zealand adolescents. What this study will involve if you are willing to support this research:

- Participation from students who participate in contact sports (i.e., rugby, soccer, and hockey). Participant's data will be excluded if they have a history of two or more

concussions, have sustained a concussion in the past two weeks, or have been diagnosed with a neurological, learning, or psychiatric disorder.

- Participation involves the once-off completion of the ImPACT test online. This will take approximately 50 minutes. Data collection would ideally occur in small groups (10 students) in a school computer lab with broadband access at a time that is most suitable to the school.
- Ideally I would like to test 200 consenting students, however, any number of participants you are willing and able to make available would be greatly appreciated.
- As an ethically approved project, all data will be de-identified (rendered anonymous) prior to any analysis or publication.
- Collected data will be stored securely electronically on the US ImPACT database. Access to the collected data will only be possible following approval by the Database Committee and all distributed data will be de-identified.

If you have any questions regarding the research project, or if you would like me to discuss it further with you in person then please feel free to contact me, Shannon Martin at shannon.martin@windowslive.com or 021 144 2977. This project will be supervised by Clifford van Ommen, a senior lecturer at Massey University and a registered Clinical Psychologist. He can be contacted on 09 441-8175 or c.vanommen@massey.ac.nz

Sincerely

[Signature]

Shannon Martin

Appendix B

MASSEY UNIVERSITY
CENTRE FOR PSYCHOLOGY

Dear Student

My name is Shannon Martin and I am a Doctoral student at Massey University. I am currently conducting research investigating the cognitive effects of sport concussion among adolescents and invite you to participate. The project is supervised by Dr Clifford van Ommen, a registered clinical psychologist and the Director of Massey University's Centre for Psychology.

Should you agree to this then it is understood that:

1. This data will be used to investigate whether the US developed ImPACT test is applicable to New Zealand adolescents.
2. Participation involves the completion on-line of the **ImPACT** (Immediate Post Concussion Assessment and Cognitive Testing) programme. This will take 50 minutes to complete and involves a symptom questionnaire and several cognitive measures assessing memory, attention, and reaction time.
3. If you score significantly low in any of the cognitive domains measured (e.g. memory, attention) you will be contacted by Clifford van Ommen.
4. This data will be securely stored electronically on the USA ImPACT database. Access to the collected data will only be possible following approval by the Database Committee and all distributed data will be de-identified.
5. Participation in the research is completely voluntary and you have the right to withdraw from the study up to two weeks following data collection.
6. Information collected from this project will be used anonymously for thesis and publication purposes.
7. Your data will be excluded from the study if you have a history of two or more concussions, have sustained a concussion in the past two weeks, or been diagnosed with a neurological, learning, or psychiatric disorder.

If you are willing to participate in this research then please return this form with the appropriate signature. If you have any questions please feel free to contact me, Shannon Martin, 021 144 2977 or shannon.martin@windowslive.com or my supervisor Dr Clifford van Ommen, c.vanommen@massey.ac.nz or 09 414 0800 extn 41241.

This project has been reviewed and approved by the Massey University Human Ethics Committee: Northern, Application 12/072. If you have any concerns about the conduct of this research, please contact Dr Ralph Bathurst, Chair, Massey University Human Ethics Committee: Northern, telephone 09 414 0800 x 43404 emailhumanethicsnorth@massey.ac.nz

Appendix C

CONSENT FORM

RE: Concussion in Sport: The incremental validity of neuropsychological testing

I have read the Information Sheet and understand the details of the study. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time. I understand this data will be stored securely on the ImPACT database and only be distributed in a de-identified form.

I, _____ agree to participate in the ImPACT Concussion Research Project under the conditions set out in the Information Sheet.

Signature of Student   _____

Date _____

Appendix D

Table 3

Independent t-tests for ImPACT items between New Zealand samples

Item	M(Group1) N = 60	M(Group2) N = 48	t	df	p
Word-memory	93.52	91.35	1.9	106	0.06
Design-memory	74.37	73.83	0.2	106	0.84
XO-memory	68.89	62.67	1.75	106	0.08
XO-correct-interference	26.19	25.95	0.45	106	0.65
XO-reaction-time	0.55	0.56	-0.53	106	0.6
XO-total-incorrect	8.38	7.69	0.57	106	0.57
SM-reaction-time	0.53	0.58	-1.97	106	0.06
SM-memory	65	59.49	1.42	106	0.16
CM-reaction-time	0.8	0.85	-1.71	106	0.09
CM-total-commission	0.53	0.38	1.21	106	0.23
TL-memory	41.66	40.75	0.33	106	0.74
TL-correct-interference	85.57	89.73	-0.14	106	0.17

Note. XO = X and O's; SM = Symbol Match; CM = Colour Match; TL = Three Letters.

Appendix E

Table 8

Skew and Kurtosis Values for Individual Items for Sample 2012A and 2012B

Item	South Africa (N = 264)		South Africa (N = 255)	
	Skew	Kurtosis	Skew	Kurtosis
Word-memory	-1.91	4.18	-1.69	3.31
Design-memory	-0.67	-0.22	-0.65	-0.34
XO-memory	-0.56	-0.38	-0.64	-0.01
XO-correct-interference	-0.69	0.58	-0.86	1.36
XO-reaction-time	0.88	0.33	1.34	2.62
XO-total-incorrect	1.67	3.82	1.91	6.21
SM-reaction-time	1.61	4.07	1.68	4.02
SM-total-memory	-0.26	-0.93	-0.55	-0.47
CM-reaction-time	0.59	5.00	-0.12	4.74
CM-total-commission	1.39	1.35	2.04	4.55
TL-memory	-1.65	3.62	-1.27	1.51
TL-correct-interference	-0.53	-5.3	-0.02	-0.28

Note. XO = X and O's; SM = Symbol Match; CM = Colour Match; TL = Three Letters.

Appendix F

Table 9

Correlation Matrix of the Twelve ImPACT Items

	WM	DM	XO-M	XO-CI	XO-RT	XO-TI	SM-RT	SM-M	CM-RT	CM-C	TL-M	TL-CI
WM	1.00											
DM	.42**	1.00										
XO-M	.20**	.32**	1.00									
XO-TI	.20**	.33**	.26**	1.00								
XO-RT	-.16*	-.26**	-.17**	-.84**	1.00							
XO-TI	-.05	-.13*	-.19**	-.26**	-.22**	1.00						
SM-RT	.02	-.04	.06	-.35**	.38**	-.09	1.00					
SM-M	.30**	.43**	.20**	.26**	-.18**	-.16**	.09	1.00				
CM-RT	-.21**	-.28**	-.07	-.45**	.46**	-.04	.27	-.21**	1.00			
CM-C	-.07	.02	.02	-.03	-.09	.18**	-.08	-.06	-.11	1.00		
TL-M	.28**	.30**	.18**	.18**	-.17**	-.07	-.03	.21**	-.12	-.01	1.00	
TL-CI	.08	.30**	.16**	.56**	-.56**	-.04	-.38**	.27**	-.36**	-.01	.16**	1.00

Appendix G

Table 12
Reliability Values of the Memory Scale

Item	Corrected item-total correlation	Cronbach's alpha if item deleted
Word-memory	.47	.59
Design-memory	.56	.52
XO-memory	.33	.62
SM-memory	.43	.58
TL-memory	.37	.59

Note. XO = X and O's; SM = Symbol Match; TL = Three Letters.

Table 14
Reliability Values of the Reaction Time Scale

Item	Corrected item-total correlation	Cronbach's alpha if item deleted
XO-reaction-time	.47	.40
SM- reaction-time	.32	.49
CM-reaction-time	.40	.44

Note. XO = X and O's; SM = Symbol Match; CM = Colour Match.

Appendix H

Table 16

Skew and Kurtosis Values forEeac Item Over Time, South African Sample

	2011 (116)		2012 (116)		2013 (116)	
	Skew	Kurtosis	Skew	Kurtosis	Skew	Kurtosis
Word-memory	-1.54	1.90	-1.52	2.32	-2.17	4.77
Design-memory	-1.27	1.22	-0.82	-0.02	-0.90	-1.31
XO-memory	-0.43	-0.52	-0.88	0.38	-0.86	0.55
XO-total-interference	-0.75	0.46	-0.30	-0.28	-0.63	0.72
XO-reaction-time	0.93	1.51	0.83	0.55	1.78	8.08
XO-total-incorrect	1.90	5.13	1.35	2.73	2.34	6.89
SM-reaction-time	1.53	3.61	1.17	1.41	1.13	1.71
SM-memory	-0.40	-0.97	-0.64	0.49	-0.85	-0.16
CM-reaction-time	-0.36	6.16	1.27	2.13	1.00	2.42
CM-commissions	1.38	2.07	2.10	5.40	3.01	12.12
TL-memory	-2.15	6.74	-1.74	4.14	-1.80	4.14
TL-correct-counted	-0.14	0.69	-0.27	0.31	-0.06	-1.04

Note. WM = Word Memory; DM = Design Memory; XO = X and O's; SM = Symbol Match; CM = Colour Match; TL = Three Letters.