The Ecological Validity of Traditional Pen-and-paper, Ecologically Oriented, and Virtual Reality Neuropsychological Test Measures to Cognitive Impairment and Real-World Function after Mild Traumatic Brain Injury

by

Eliyas Jeffay

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Psychology
University of Toronto

The Ecological Validity of Traditional Pen-and-paper, Ecologically Oriented, and Virtual Reality Neuropsychological Test Measures to Cognitive Impairment and Real-World Function after Mild Traumatic Brain Injury

Eliyas Jeffay

Doctor of Philosophy

Department of Psychology
University of Toronto

2019

Abstract

A growing concern in neuropsychology is whether neuropsychological test measures (NTMs) can predict functional outcome (i.e., ecological validity). Ecological validity can be understood in two ways: veridicality (i.e., prediction of functional outcome is independent of how the test reflects the outcome being measured) and verisimilitude (i.e., tests resemble the functional outcome they are predicting). Historically, the veridicality approach has been utilized (i.e., by way of traditional, pen-and-paper NTMs), but there has been a movement to employ verisimilitude approaches (e.g., Behavioural Assessment for Dysexecutive Syndrome; BADS). These approaches were examined in patients experiencing ongoing cognitive complaints in the post-acute period of recovery (> 3 months) following a mild traumatic brain injury (mTBI). This was the principal inclusion criteria for the studies described here. A meta-analytic study was conducted and found that traditional NTMs were not sensitive to persistent cognitive complaints in this population across all domains. Studies 2 and 3 utilized archival data to determine whether NTMs could predict return to work (RTW) status. The BADS predicted employment status by way of medium-to-large effects, while traditional NTMs did not. Overall, these findings suggest the verisimilitude approach is more ecologically valid than the veridicality approach. As such, two virtual reality tests (VRTs) evaluating attention and executive function were developed to investigate their ability to predict RTW. Tests of attention (VR and traditional tests) significantly predicted group membership at 82% accuracy, with 82.6% sensitivity and 81.5% specificity. The attention shift trial of the VRT and to a lesser degree the total speed score of the Ruff 2 & 7 were predictive of employment status. Overall, this research provides empirical evidence for the verisimilitude approach when evaluating RTW status in patients who are in the post-acute period of recovery following mTBI. Moreover, it provides initial evidence for the clinical utility of VRTs to evaluate real-world functioning.

Acknowledgments

A PhD is not completed alone. Thus, I have many people to thank for their support and assistance throughout the process.

First, I am forever indebted to my primary supervisor, mentor, "academic father", and friend, Dr. Konstantine K. Zakzanis. From teaching Abnormal Psychology in undergrad to editing this dissertation, you have fostered my professional development in a way that strikes balance between giving independence and providing assistance. Since day one you openly referred to and treated me as your colleague. You are, inspiringly, the embodiment of the phrase 'hard work truly pays off', which has been the mantra that has motivated me through difficult times. Thank you, truly, for *everything* you have done.

To my committee members, Dr. Anthony Ruocco and Dr. Mary Pat McAndrews, thank you for your time, guidance, thoughtful comments, and patience throughout my graduate years. On more than one occasion, I had reached out to each of you for advice on a wide range of topics. Every single time, you had provided a level of insight that was most helpful. I have been honoured to have been supervised by such a prestigious team and I do hope to be able to pay your teachings forward.

To my lab family, Manu, Anosha, Matt, Steven, Paul, & Elisabeth, thank you for assisting in the pilot projects that were pivotal for the direction of this graduate work. Raymond, your time with us was tragically cut-short, but you will forever be remembered and missed. To my lab sisters, Diana, Kyrsten, Courtney, Sonya, Angie, Krysta, and Emnet thank you for always cheering me up and knowing when I needed help even though I would be too proud to ask for it. Zach, you are the reason I continued with undergrad and the reason for my place in the lab. I will never forget our early philosophical chats and your generosity in sharing your knowledge and experience with me.

To my family, thank you for your honesty, love, and guidance. For many reasons, nothing in my life, including obtaining a PhD from one of the top universities in the world, would be possible without my parents, Abbas and Leila Jeffay. Immigrating to a country, learning a new language, adapting to a new culture, and raising three children has been the set example of the what can be achieved with perseverance and hard work. It is my life goal to pay you back and make you proud. To my two brothers, thank you for always knowing what to say to make me laugh when I was feeling low and for doing my share of the housework when I had to focus on my studies.

To my best friend, Dr. Jeffrey Davis Knight, thank you for leading by example in showing that I can do better. Though it seems that we usually take two steps backwards and one step forward in our adventures, we’ve learned a lot together and have gained memories that I truly cherish. Plain and simple, I look up to you. I hope that our friendship continues until one of us has dementia. Not it.

Lastly, to my dearest fiancé, Monica. How fortuitous that I was able to find the love of my life whilst initiating the work of my life. Thank you for being an endless source of motivation, support, and most of all patience throughout what seemed like a life-time. Despite studying psychology, I learned true patience and compassion by your lead. I love you and am excited to learn more from you as we embark on the next chapter of our lives together.

I cannot possibly thank everyone. If I have left anyone out it was not intentional. Please consider this as a formal thank you to those that had helped me along this journey. You know who you are.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3D = Three Dimensional
ABI = Acquired Brain Injury
ACRM = American Congress of Rehabilitation Medicine
ADHD = Attention Deficit Hyperactivity Disorder
(i)ADL = (Instrumental) Activities of Daily Living
ANOVA = Analysis of Variance
APA = American Psychological Association
BADS = Behavioural Assessment of The Dysexecutive Syndrome
BVMT = Brief Visual Memory Test – Revised
CAD = Canadian Dollars
CBT = Conveyor Belt Task
CDC = Centers for Disease Control
CIHI = Canadian Institute for Health Information
CPU = Central Processing Unit
COWAT = Controlled Oral Word Associated Test
CT = Computed Tomography
CVLT-II = California Verbal Memory Test – Second Edition
DAI = Diffuse Axonal Injury
D-KEFS = Delis-Kaplan Executive Function System
DEX = Dysexecutive Questionnaire
DTI = Diffusion Tensor Imaging
FSIQ = Full Scale Intelligence Quotient
GB = Gigabyte
GCS = Glasgow Coma Scale
ICN = Inconsistency
(FS)IQ = (Full Scale) Intelligence Quotient
INF = Infrequency
LCD = Liquid Crystal Display
LOC = Loss of Consciousness
MRI = Magnetic Resonance Imaging
NAART = North American Reading Test
NART = National Adult Reading Test
NIH = National Institute of Health
NIM = Negative Impression Management
PASAT = Paced Auditory Serial Addition Test
PAI = Personality Assessment Inventory
PC = Personal Computer
PIM = Positive Impression Management
PTA = Post Traumatic Amnesia
RAM = Random Access Memory
RCFT = Rey Complex Figure Test
RSAT = Ruff 2 & 7s Selective Attention Test
RTW = Return to Work
SEM = Standard Error of The Mean
SES = Socioeconomic Status
SDMT = Symbol Digit Modalities Test

SPSS = Statistical Package for The Social Sciences
mTBI = Mild Traumatic Brain Injury
TBI = Traumatic Brain Injury
TEA = Test of Everyday Attention
TMT(A/B) = Trail Making Test (A/B)
TOL = Tower of London
US(D) = United States (Dollars)
VE = Virtual Environment
VR = Virtual Reality
WAIS-R = Wechsler Adult Intelligence Scale - Revised
WASI = Wechsler Abbreviated Scale of Intelligence
WMS-R = Wechsler Memory Scale - Revised
WCST = Wisconsin Card Sorting Test
WHO = World Health Organization

Chapter 1

1 General Introduction

Broadly defined, a traumatic brain injury (TBI) is a craniocerebral trauma associated with decreased level of consciousness, amnesia or other neurologic or neuropsychological abnormalities, skull fracture, intracranial lesions, or death (Faul, Xu, Wald, & Coronado, 2010; Thurman, Sniezek, Johnson, Greenspan, & Smith, 1995). TBIs have been deemed a serious public-health problem by the World Health Organization (WHO), particularly when symptoms are coupled with resultant disability with estimated economic costs ranging in the billions annually (WHO, 2006). Moreover, research in this area has been inconsistent due, in part, to disagreement on what factors constitute a TBI (Cassidy, Carroll, Cote, Hold, & Nygren, 2004; Frost, Farrer, Primosch, & Hedges, 2013; Menon, Schwab, Wright, & Maas, 2010). In addition, the sensitivity of clinical tests measuring functional outcome has been questioned. In the following review, these issues will be elaborated.

Despite the obvious individual and socioeconomic burden that can follow a TBI, empirical research as it pertains to outcome is characterized by inconsistent and often conflicting findings. One possible explanation may stem from variable injury characteristics that have been adopted to diagnose a TBI. Moreover, it may be possible that the use of inadequately sensitive test measures to articulate the breadth and severity of subjective complaints following injury may further account for these inconsistencies and conflicting findings.

In this work, the aim is to review the current body of knowledge to determine the general sensitivity of traditional pen-and-paper neuropsychological test measures in patients with mild TBI (mTBI), determine which traditional pen-and-paper tests may be sensitive to predicting functional outcome, compare the functional outcome sensitivity of tests that are purposed to be high in ecological validity by way of resembling real-world activities, and to compare novel, ecologically-valid, neuropsychological test measures with respect to return to work (RTW) outcome in a sample of patients who had sustained a mTBI.

1.1 Definitions

A fundamental aspect of conducting research is to operationally define a variable and to adhere strictly to this criterion. While TBIs are operationally defined within research studies, the definitions vary and are inconsistent. In other words, different researchers use different variables to define the same phenomenon (Sharp & Jenkins, 2015). To this end, different classification systems have been proposed and utilized by researchers and clinicians (Schretlen & Shapiro, 2003). Clinically, the severity of brain injury lies on a continuum from uncomplicated mTBI, to complicated mTBI, to moderate, to severe and is determined by way of acute injury characteristics and the presence or absence of positive findings on neuroimaging (Bryant, et al., 2010; Kay et al., 1993; Menon et al., 2010). Generally, however, severity has been defined by factors such as the duration of post-traumatic amnesia (PTA), the Glasgow Coma Scale (GCS) score, duration of loss of consciousness (LOC), and absence/presence of neuroimaging findings.

The GCS score is the most commonly employed scale for classifying TBI severity (Menon et al., 2010). This scale was first proposed by Teasdale and Jennett (1974) as an aid to the clinical assessment of post-traumatic (un)consciousness. Typically employed by first responders and during the acute periods of trauma in the hospital, the scale has three components: eye opening, verbal response, and motor response. The eye-opening component score ranges from 1 (none) to 4 (spontaneous); verbal response score ranges from 1 (none) to 5 (oriented conversation); and the motor response score ranges from 1 (none) to 6 (obeys simple commands). The lowest possible score is 3 (deep coma or death) and the highest score is 15 (full wakefulness). It is generally agreed that following some sort of trauma to the head, a GCS score of 3-8 represents a severe TBI, 9-12 reflects a moderate TBI (Bryant et al., 2010) and score ranging from 13-15 would be indicative of a mTBI (Kay et al., 1993).

Beyond the GCS, several neurological organizations have published position statements on the definition of TBI severities (Cifu et al., 2009; Malec et al., 2007; McCrory et al., 2017; Menon et al., 2010) and specifically that of mTBI (American Academy of Neurology, 1997; Holm, Cassidy, Carroll, & Borg, 2005; Cifu et al., 2009; Kay et al., 1993; Ruff et al., 2009). Traditionally, the GCS score described above is used in conjunction with the PTA and/or LOC duration as they have been found to be better predictors of outcome (Dikmen, Machamer, & Temkin, 2009). PTA is defined as the period from injury to the time when memory for day-to-day events return. When PTA duration is less than 60 minutes, the severity falls in the mild category. A duration of 60 minutes to 24 hours is considered moderate, and a duration greater than 24 hours is usually indicative of a severe injury (Bryant et al., 2010). However, reliability of the PTA estimate has been reported to be poor and inconsistent and may be missing altogether in the medical chart. When present, however, these indicators of consciousness may continue to be poor indicators of outcome without considering information regarding injury characteristics, mechanism of injury, and neuroimaging findings (see Shames, Treger, Ring, & Giaquinto, 2007). With respect to the latter, the role of neuroimaging in the clinical diagnosis of a TBI is limited to positive or negative findings. To this end, a positive finding on neuroimaging studies (i.e., contusion, hemorrhage) is commonly regarded as escalating the severity to either moderate or severe depending on the other components (e.g., mechanism of injury, GCS, and PTA) with the exception for a mTBI. The presence of a positive neuroimaging finding in an injury that would otherwise be considered a mTBI is termed a 'complicated' mTBI.

1.2 Inconsistencies with Definitions

Multiple definitions have been proposed by different organizations with resultant inconsistencies with respect to the operational definition of a TBI. In other words, although all organizations report findings using patients who have sustained a TBI, the specific definition that is employed to characterize severity is variable. For example, one study may define a mTBI as a head injury with brief LOC period, whereas another study may define it as a head injury with a period of LOC that is less than 30 minutes, negative neuroimaging findings, a GCS score of 13-15, and a PTA period less than 24 hours (Sharp & Jenkins, 2015). In addition, multiple terms are used interchangeably to refer to traumatic brain injury; head injury, concussion, intracranial injury, acquired brain injury, and brain injury. Such inconsistencies make it difficult for researchers to communicate and limits the progression of the field, as findings may be related to different constructs (Sharp & Jenkins, 2015).

The definition of moderate or severe TBI is less contentious and holds more agreement among researchers and clinicians than that of the definition of mTBI (Belanger & Vanderploeg, 2005; McCrea et al., 2009; Ruff et al., 2009). Diagnosing a mTBI is more challenging because of the apparently rapid resolution of symptoms and lack of objective findings on neuroimaging (Ruff et al., 2009). This is troubling as the majority (up to 90% by some reports) of all reported TBIs are mild (Cancelliere, Coronado, Taylor, & Xu, 2017; Faul et al., 2010; Krause, McArthur, Silverman, & Jayaraman, 1996).

1.3 Prevalence & Incidence

It has been estimated that a head injury occurs approximately once every 15-16 seconds (Signoretti, Vagnozzi, Tavazzi, & Lazzarino, 2010) and that one in every 50 visits to the emergency department is due to a TBI-related injury (Taylor, Bell, Breidling, & Xu, 2017). In order to understand the impact of TBI on society, the incidence (the number of new cases reported per the general population) and prevalence (a measure of all patients affected) rates need to be discussed (Corrigan, Selassie, & Orman, 2010). In the United States, the Centers for Disease Control (CDC) regularly publishes incidence rates for civilian emergency department visits, hospitalization, and deaths as a result of TBI (Faul et al., 2010; Rutland-Brown, Langlois, Thomas, & Xi, 2006; Langlois, Rutland-Brown, & Thomas, 2004; Langlois, Rutland-Brown, & Wald, 2006; Taylor et al., 2017). Throughout the reporting years, there appears to be several trends emerging. From 1995-2001, the number of TBI-related emergency department visits were 1,111,000; this rose to 1,365,000 from 2002-2006; from 2007-2013 this increased again but this time by 80% to 2,460,420. The number of TBI-related hospitalizations was reported to be 235,000 from 1995-2001; 275,000 from 2002-2006; and 281,610 in 2007-2013. Lastly, the number of TBI-related deaths were reported to be approximately 50,000 from 1995-2001; approximately 52,000 from 2002-2006; and finally, 55,927 from 2007-2013 (Faul et al., 2010; Rutland-Brown et al., 2006; Taylor et al., 2017). The overall trend from 1995 to 2013 is an increase in the number of TBI-related emergency department visits, hospitalizations, and deaths.

A reliable finding that has been found across all TBI incidence analyses has been the gender effect. Specifically, males have a higher risk of TBI-related injuries than do females. The WHO has reported that males have 2-3 times greater likelihood of sustaining a TBI-related injury than females (WHO, 2006). Based on the CDC data, American males are 1.5 times more likely to suffer a TBI-related injury than are females (Taylor et al., 2017). Furthermore, analysis of national health records for TBI-related injuries reveal that males regularly have significantly higher rates of risk in Poland (Miekisiak et al., 2016), New Zealand (Ao et al., 2015), France (Masson et al., 2003), Sweden (Kleiven, Peloso, & von Holst, 2003), Germany (Steudel, Cortbus, & Schwerdtfeger, 2005), and all other European countries reported in a systematic review of the incidence rates of TBI (Tagliaferri, Compagnone, Korsic, Servadei, & Kraus, 2006). Canadian estimates are comparable; Fu, Jing, Fu, and Cusimano (2016) found that males consistently had 60-80% higher rates of risk for TBI-related injuries than females over 8 years of data (2002-2009), which is consistent with other Canadian findings (Colantonio et al., 2010). It has been proposed that this gender imbalance is likely attributed to the risk taking behaviours and activities observed in men (Corrigan et al., 2010).

The top three principal mechanisms for head-injury have remained consistent in their relative occurrence over the CDCs reporting years (Faul et al., 2010; Rutland-Brown et al., 2006; Taylor et al., 2017). To this end, the mechanism that has been attributed to the highest number of TBI-related injuries are falls (413.2 per 100,000 population). Falls were the most common mechanism of TBI-related injury and accounted for 47.2% of the combined total of TBI-related emergency department visits, hospitalizations, and deaths in 2013. Motor-vehicle accidents crashes were second most common accounting for 15.4% of the total TBI-related injuries (142.1 per 100,000 population) in 2013. Lastly, TBIs that fell in the category of 'struck by/against' were the third most common mechanism of injury accounting for 13.7% of the total TBI-related injuries (121.7 per 100,000 population). As a point of clarification, struck by/against represents all TBI-related injuries that were the result of being struck by or against an object. The authors of the incidence report explained that the sports-related and recreation-relation activities were likely in the struck by/against category but they cautioned that this could not be verified based on their analysis (Taylor et al., 2017). Moreover, it was reported that the most common principle mechanism of injury from those aged 15-34 was motor-vehicle accidents but the authors noted that the incidence of motor-vehicle accidents had overall decreased as compared to previous reporting years. This decrease was attributed to behavioural and vehicle improvements (e.g., better airbags and electronic stability control).

Prevalence rates have been of little focus in the literature due to logistic difficulties in estimating such data accurately and reliably, since many national health care systems do not routinely gather and report such data (Corrigan et al., 2010; Fleminger & Ponsford, 2005; WHO, 2006). However, there are some national healthcare systems that have access to their entire population health data and collect information on lifetime prevalence of TBI. In the United States it is estimated that 1-2% of the population live with a TBI-related disability (Langlois et al., 2006). Many studies have been published regarding the incidence and prevalence rates of TBI in specific demographic communities (i.e., sex, high/low socioeconomic status [SES], age, race) but few have studied the interaction effects of these variables in a single study (Kisser, Waldstein, Evans, & Zonderman, 2017). Hence, Kisser and colleagues (2017) sought to investigate TBI prevalence rates in a demographically diverse community sample. Specifically, they conducted an archival study on 2,881 participants from the Healthy Aging in Neighborhoods of Diversity across the Life Span study (Evans et al., 2010). Participants self-reported lifetime history of TBI and were stratified across age, sex, socioeconomic status, and race (White vs. African-American). The results from a logistic regression analysis revealed a three-way interaction between age, race, and socioeconomic status. Specifically, they found that White males aged between 30-36 who were considered to have low socioeconomic status and African-American males between the ages of 56-64 who were also considered to have low socioeconomic status were most at risk for TBI. The authors indicated that these results are preliminary in nature and that further studies should be conducted to confirm their findings across other rural dwellings to increase generalizability. Moreover, incidence of TBI was measured via self-report, which has been well-established as an unreliable and inconsistent measure of TBI occurrence (Evans et al., 2010). Lastly, the larger study from which the data was collected, did not recruit patients under the age of 30. This is an important limitation as many studies have indicated that the two age groups that are most at risk for TBIs are the very young (0-4 years) and the elderly (Faul et al., 2010; Rutland-Brown et al., 2006; Langlois et al., 2004; Langlois et al., 2006; Taylor et al., 2017).

Since almost all TBI-related emergency department visits are of the mild severity, a follow-up study to Kisser at al. (2017) was conducted to determine the sociodemographic characteristics of mTBI in a national sample from 2006-2012 (Cancelliere et al. 2017). The study found that the majority of all mTBI-related emergency department visits were in conjunction with other injuries and that the very young (males and females, 0-4 years), young adults (males 15-24 years), the elderly (females 65+ years) had the highest average annual rates of emergency department visits. Moreover, their analysis revealed more subtle risk factors. Namely, that males, seniors, assault injuries, suburban or rural populations, low SES, non-private insurance, weekend (as opposed to a weekday) admission, and visiting a teaching hospital were associated with an admission to the emergency department with a mTBI. The authors noted that the prevalence for mTBI related emergency department visits had increased when compared to previous studies with children under 3 having the largest increase.

To sum, mTBI is not only difficult to diagnose, but it is a growing problem.

1.4 Pathophysiology

TBIs have also been categorized into open and closed head injuries based on the pathophysiological processes that occur as a consequence of the two types of injuries. Open head injuries are classified when something penetrates through the dura mater into the brain, resulting in an open wound and/or skull fracture. Common open head injuries include high velocity objects such as bullet wounds (50%) but can also be caused by low velocity objects such as other weapons (e.g., knives via stab; 10-20%; Ball, 2015). In addition, if skull fractures caused by head trauma can penetrate the brain, then it would be considered an open head injury. Open head injuries result in clinically serious injuries and poorer outcomes as compared to closed head injuries (Ball, 2015). As noted by Ball (2015), open head injuries (i.e., gunshot wounds) that cross the midline are especially fatal, whereas the prognosis for unilateral gunshot wounds are much more favourable. In addition, other foreign objects such as shrapnel, dirt, glass, and other debris may cause additional complications hindering recovery (Ball, 2015).

Closed head injuries, conversely, do not involve penetration of the skull but rather are the result of a blow to the head (Bigler, 2001). A resultant skull fracture where the fragments are not loose and penetrate the dura mater can also be considered a closed head injury. The brain is surrounded by cerebrospinal fluid and is in a state of buoyancy in the cranium. Following a blow to the head, the inertial forces cause the brain to initially move towards the direction of the locus of the blow (coup) and impact the bony protrusions of the interior of the cranial fossa. Immediately following the initial impact, the brain recoils to the opposite direction of the blow (contrecoup) and once again may collide with the bony protrusions of the interior cranial fossa (Bigler, 2001; Genneralli, Thibault, & Graham, 1998). On both occasions, the brain may incur forces that cause it to accelerate and decelerate quickly, rotate, and twist. Events such as car accidents, athletic injury, and falls are common causes of closed head injuries (Bigler, 2001). Closed head injuries constitute the majority of head injuries (Leon-Carrion, Dominguez-Morales, Martin, & Cabezas, 2005; Faul et al., 2010) and thus, will be the main focus of this review and work.

The pathophysiology of TBI has been further classified into primary and secondary injuries in the literature. Primary injuries, also known as direct injury, result from the energy transfer of the impact to the brain (Bigler, 2001; 2008). For example, if a person is hit on the forehead with a baseball bat, the kinetic energy from the baseball bat will transfer to compress the head. The brain inside the skull will move towards the anterior cranial fossa due to inertial forces and possibly impact the surface of the bone. Bone, by definition, is hard, whereas the brain is soft and fragile. The results of the impact may cause stretching, twisting and ultimately degeneration of axons which may occur all simultaneously and either at a focal point or have dispersed effects across the entire brain (Bigler, 2001; 2008).

Clinically, the ensuing damage and symptomatology is generally presented in the form of secondary injury. Secondary injuries manifest following the initial impact and may progress from primary injuries. It has been reported to include multiple lesions in white matter tracts (e.g., diffuse axonal injury; DAI), cerebral contusions, and focal shear injury (Bigler, 2001). In addition, a disruption to cortical vasculature can cause hemorrhages, which may lead to hematomas. Thus, secondary injuries include but are not limited to contusions, hemorrhage, increased intracranial pressure (e.g., hydrocephalus, cerebral edema), hypoxia, and ischemia. Moreover, the onset of secondary injury can range from minutes to weeks after the date of injury and the clinical duration of can last from days to years (Bigler, 2008). During trauma (e.g., motor vehicle accident) the physical forces can act in a linear, transverse, or rotational direction. The head may undergo rapid acceleration and deceleration on multiple planes of axis, which have been reported to cause cellular disruption (e.g., altered conduction velocities, changes to mitochondrial energy output, oxidative stress, cytoskeleton degeneration, neuroinflammation, and ionic imbalance) that may be undetectable through neuroimaging (Bigler, 2001; 2008; Blenow, Hardy, & Zetterberg, 2012; Genneralli et al., 1982; Genneralli et al.,1986; Gennarelli et al., 1998; Ommaya, Fisch, Mahone, Carrao, & Letcher, 1993; Stritch, 1970).

1.5 Cognitive Sequelae & Typical Outcome

Although there is evidence in the extant literature for cognitive impairment across all domains following a TBI, the diffuse nature of neural damage that occurs following a TBI has been argued to be associated to the hallmarks symptoms of slowed speed of processing with subsequent deficits in executive control and sustained attention (Dikmen et al., 2009; Richard, O’Conner, Dey, Robertson, & Levine, 2018; Mathias & Wheaton, 2007). The same mechanism, to a milder and undetectable degree with current routine clinical neuroimaging techniques, has been proposed to be related to speed of processing, attention, and executive impairments following a mTBI (Bigler, 2008).

A systematic review of the cognitive sequelae following a TBI shows a clear dose-response relationship between TBI severity with the breadth and severity of cognitive impairment (Dikmen et al., 2009). They concluded that there is robust evidence of long-term (greater than six months) cognitive impairments associated with moderate to severe TBI, which were across multiple domains and included attention, episodic memory, speed of processing, and executive functioning. Specifically, within the domain of attention, difficulties with sustained, selective, and divided attention are characteristic following a TBI (Kinsella, 2008; Stuss et al., 1989). Reported symptoms such as difficulties with inhibitory control, being easily prone to distraction, and poor concentration are common following a TBI. Despite inconsistent test measures used to measure speed of processing, impairment in this domain has frequently been found following a TBI (DeLuca, 2008). Moreover, impairments in speed of processing may affect other cognitive domains such as complex attention and executive functioning (Kinsella, 2008). Executive dysfunction such as impaired working memory, planning, organization, problem solving ability, and judgement are also reported in this population and may manifest clinically as symptoms of poor self-control, insight, and self-awareness (DeLuca, 2008). Memory impairment is also prominent following a TBI, which some have argued may be as a result of impaired encoding due to reduced speed of processing and resultant poor working memory (DeLuca, 2008). To this end, difficulties with respect to encoding, storage, and retrieval have been reported (Ponsford & Kinsella, 1992). In more severe TBIs, deficits in language and visuoconstructional abilities have been documented. In the absence of a focal lesion to the language regions (left hemisphere in most individuals), classic aphasia is generally a rare occurrence following a TBI (Sohlberg & Mateer, 1990). More common are expressive difficulties such as word finding problems with resultant circumlocutions, which may be due to impairments in working memory and reduced speed of processing (Dikmen et al., 2009). Similarly, visuoconstructional impairment is thought to be indirectly related to executive dysfunction such as impairments in working memory, planning, and organization and less so due to poor visual acuity (DeLuca, 2008).

Several components of attention have been identified with neuroanatomical associations. For example, three attention networks have been reported by Posner and Peterson (1990; revisited in Peterson & Posner, 2012); the alerting, orienting, and executive networks. The alerting network has been proposed to maintain basic levels of alertness by way of the neurotransmitter norepinephrine through the reticular activating system projecting ventrally to structures such as the locus coeruleus and anterior cingulate cortex towards the right frontal and parietal lobes. The orienting network orients and prioritizes sensory input or visuospatial areas in either top down (internally sourced) or bottom up (external stimuli) manner. The orienting network has been associated with the ventral and dorsal projections to areas of the brain such as the superior parietal, temporal parietal junction, frontal eye fields, and superior colliculus through the modulation of the neurotransmitter acetylcholine. The executive control network monitors and resolves conflicts among thoughts, feelings, and responses and is associated with the midline frontal areas such as the anterior cingulate cortex along with the lateral and ventral prefrontal cortex. However, inconsistent relationships between these cognitive and anatomical networks with clinical test measures of attention have been reported. For example, partial agreement for this relationship was reported by Mirsky, Anthony, Duncan, Ahearn, and Kellan (1991). They conducted a principle components analysis using mixed neuropsychiatric and healthy control sample to identify separate components of attention derived from clinical measures. They concluded to a four-factor model which they labeled (1) focus attention which was related to inferior parietal and superior temporal structures through tests such as the Trail Making Test (TMT); (2) sustained attention which was related to the mesopontine reticular formation and thalamic reticular nucleus and continuous performance tests; (3) shift attention which was related to the dorsolateral frontal and anterior cingulate cortex and the Wisconsin Card Sorting Test (WCST); and (4) numerical encoding which was related to the Digit Span and Arithmetic tests from the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997). Similarly, Shum, McFarland, and Bain (1990) conducted a principle component analysis on tests of attention in samples of patients with closed head injury and controls and found a three-factor model that included visuo-motor scanning (TMT), sustained-selective attention (Stroop, Serial 7s), and attention span (Digit Span). However, multiple factor analyses of 13 measures of attention in TBI samples found limited evidence for clinical test measures being related to distinct components of attention. Rather, the authors concluded that most tests of attention may be related to ‘global attention’ or ‘general neuropsychological impairment’ (Schmidt, Trueblood, Merwin, & Durham, 1994).

Regarding the sequelae following a mTBI, the constellation of symptomatology early after onset can be all encompassing and nonspecific. Following a mTBI the reported physical symptoms can include, but are not limited to, headache, dizziness, sleep disturbance, fatigue, irritability, sensitivity to noise and light and nausea (Alves, Macciocchi, & Barth, 1993; Binder, 1997; Christensen et al., 2008; Gasquoine, 1997). Emotional distress, such as depression, specific anxiety disorders (Hibbard, Uysal, Keplar, Bogdany, & Silver, 1998) and other forms of psychopathology are also common (Mathias & Coats, 1999; Wrightson & Gronwall, 1999). A myriad of neuropsychological impairments have been reported as well and virtually every neuropsychological domain has been documented to be impaired following a mTBI. These include slowed information processing (Johansson, Berglund, & Ronnback, 2009; Madigan, DeLuca, Diamond, Tramontano, & Averill, 2000; Mathias et al., 2004), reduced visuomotor speed (Cremona-Meteyard & Geffen, 1994; Levin et al., 1987), attentional deficits (Chan, 2005; Ziino & Ponsford, 2006), memory difficulties (Levin et al., 2004; McAllister, Flashman, McDonald, & Saykin, 2006), executive dysfunction (Hartlkainen et al., 2010; Nolin, 2006; Ord et al., 2010), reduced expressive fluency (Henry & Crawford, 2004; Zakzanis, McDonald & Troyer, 2011) and reduced awareness (Sherer et al., 1998).

McCrea et al., (2009) proposed a theoretical model of recovery of a single, uncomplicated mTBI. Their model consisted of three periods: acute, sub-acute, and chronic periods. The acute period was defined as the period immediately after the injury up to approximately 5 days post-injury characterized by symptoms and cognitive impairments that were severe enough to disrupt normal daily functioning. The subacute period spanned approximately 5 days to 30 days post-injury with the gradual resolution of symptoms and cognitive and functional impairments. The majority of patients with mTBI achieve full recovery within this period (McCrea et al., 2009). The chronic period begins past the 30 days post-injury marks and extends beyond. Within this period, a small proportion of patients, estimated by the authors at 5%, will continue to report persistent symptoms and cognitive complaints. The authors note that, physiologically, the brain will return to a normal state of function as per functional neuroimaging and that complaints may be influenced either by a misdiagnosis (i.e., more severe TBI) or non-injury characteristics such as mood disturbance, anxiety, post-traumatic stress, pain, or secondary gains.

Within the literature, the expected cognitive recovery following a single uncomplicated mTBI continues to be a topic of debate. There are meta-analyses, mathematical amalgamations of reported effect sizes across studies, that have reported no significant cognitive effects beyond 3 months or termed *post-acute* period of recovery (Belanger, Curtiss, Demery, Lebowitz, & Vanderpleog, 2005; Binder, Rohling, & Larrabee, 1997; Frencham, Fox, & Mayberry, 2005; Rohling et al., 2011). Indeed, Binder and colleagues (1997) reported small effect sizes on neuropsychological test measures, which suggested that the prevalence of persistent neuropsychological deficit after three months post-injury is likely to be little to none. However, there is support from other studies that report persistent symptoms and cognitive deficits on neuropsychological test measures (specifically in the domains of attention and processing speed) in the post-acute period and beyond (Berenstein, 2002; Chan, 2002; Johansson et al., 2009; Potter, Jory, Bassett, Barrett, & Mychalkiw, 2002; Solbakk, Reinvang, Neilson, & Sundet, 1999; Vanderploeg, Curtiss, & Belanger, 2005). Persistent neuropsychological impairments after one-year post-injury have been reported with prevalence rates of 7% (Binder et al., 1997) to more than 40% (Alves et al., 1993). Generally, estimates have ranged from 5% - 20% of patients who continue to experience persistent problems in the longer term (Cancelliere et al., 2014; Cassidy et al., 2004; Iverson, 2005). Full recovery is expected within 1-3 months, yet some patients continue to exhibit persistent difficulties. This incongruence can result in conflicting evidence-based decisions amongst clinicians.

Traditionally, structural and functional abnormalities following a mTBI were not found leading researchers to conclude that mTBI does not cause any brain-related structural impairment. Indeed, the absence of neuroimaging findings is a condition for the diagnosis of mild versus moderate or severe TBI. This absence of findings may be due to absence of brain related pathology or it could be a limitation of the sensitivity of the resolution of clinical neuroimaging techniques. Recent technological advances have allowed researchers to detect functional and structural abnormalities (McAllister et al., 1999; McAllister, Sparling, Flashman, & Saykin, 2001; Solbakk, Reinvang, & Nielsen, 2000). For example, DAI has been visualized through advanced neuroimaging techniques allowing the characterization of the impact of a single mTBI on cerebral perfusion and metabolism (Ruff, 2005). This is important as DAI has been related to aspects of attention and executive functioning such as reduced speed of processing, working memory, attentional/inhibitory control (McAllister et al., 2001). DAI is highly susceptible to the fronto-temporal areas of the brain, which are related to attention and executive functioning (Stuss et al., 1989) and may contributory to those who report persistent symptoms beyond the typical recovery period and to the attention and executive impairments following a mTBI (Paré, Rabin, Fogel, & Pépin, 2019; Solbakk et al., 2000). To this end, impairments on tests of attention and executive functioning may provide a sensitive measure of persistent cognitive impairment (Mangels, Craik, Levine, Schwartz, & Stuss, 2002). However, as noted above, a meta-analysis on the neuropsychological profile of mTBI reported small effect sizes on tests of processing speed, working memory, and attention (Binder et al., 1997). Persistent impairments on neuropsychological tests may be explained by a host of factors, such as pain disorders and headache (Iverson & McCracken, 1997; Radanov, Dvorak, & Valach, 1992), fatigue (Johannson et al., 2009), stress at the time of injury (Alexander, 1995; Bryant, 2008), depression and mood disorders (Garden & Sullivan, 2010; Iverson, 2006; Suhr & Gunstad, 2002), post-traumatic stress (Hoge et al., 2008; Kennedy, Leal, Lewis, Cullen, & Amador, 2010; Nelson, Yoash-Gantz, Pickett, & Campbell, 2009), premorbid personality characteristic (Mittenberg & Strauman, 2000) and involvement in litigation (Binder et al., 1997; Belanger et al., 2005; Greiffenstein & Baker, 2001; Lees-Haley et al., 2001). It has also been argued that these impairments may reflect unremitting neuropathological alterations in mTBI that have yet to be fully understood (Bigler, 2001; 2003; 2004).

1.6 TBI vs mTBI and Return to Work

An additional factor that may explain persistent symptoms in the post-acute period of recovery from a mTBI may be that clinical neuropsychologists lack the necessary tools to measure such persistent impairments as it pertains to real-world functioning (Bigler, 2003; Silverberg & Millis, 2009). In other words, while patients who sustained a mTBI may perform well on pen-and-paper neuropsychological test measures within the context of an examiner directed, office-based assessment, it may be that existing neuropsychological test measures fail to adequately assess real-world neuropsychological dysfunction in this patient population.

Many patients who are being referred to a neuropsychologist for a clinical evaluation are typically sent by third party insurers to determine their prognosis for future functioning, in particular, their ability to return to gainful employment. Since the majority of mTBIs occur in young males (Taylor et al., 2017) who have a lifetime of working lives ahead of them, return to work (RTW) is a salient concern (Ruffalo, Friedland, Dawson, Colantonio, Lindsay, 1999). RTW disability can defined as time spent away work, sick leave, or a decrease in productivity, working with functional limitations (Cancelliere et al., 2014). Following a brain injury, RTW has been argued as a key indicator of real-world functioning as poorer psychosocial adjustment, reported physical symptoms, greater usage of health services, and lower reported quality of life have been associated with poorer RTW (Cancelliere et al., 2014). Moreover, small to moderate effect sizes have been reported between cognitive performance and RTW after moderate to severe TBI (Chaytor & Schmitter-Edgecomb, 2003; Marcotte & Grant, 2010). Impairments in the domains of executive functioning (Wehman, Brieout, Targett., 2017), attention (Crepaeu &

Scherzer, 1993; Wehman et al., 2017), speed of processing (Asikainen, Nybo, Muller, Sarna, & Kaste, 1998; DeLuca, 2008) learning and memory (Cifu et al., 1997; Ip, Dornan, & Schentag, 1995), and language (Guilmette, 2008) have been correlated with RTW status. Particularly, impaired performance on measures of attention and executive functioning have been reported to be good predictors of RTW (Devitt et al., 2006; Spitz, Ponsford, & Schonberger, 2013). For example, one study found that the Tinker Toy Test was predictive of RTW status in a sample of TBI patients who were two years post-injury (Bayless, Varney, & Roberts, 1989). Another study found that the Stroop Task, which measures aspects of divided attention and executive inhibition, was significantly associated with unemployment after a moderate-severe TBI (Asikainen et al., 1998). Moreover, a quantitative review supported the prognostic value of performance on executive test measures and predicting RTW following a moderate-severe TBI (Ownsworth & McKenna, 2004). These impairment manifest behaviourally as slowed thinking, problem with memory, reduced concentration, poor planning and problem solving ability, fatigue, irritability, headaches, depression, and anxiety (Dikmen, Machamer, Winn, & Temkin, 1995).

Unfortunately, the same associations between performance on neuropsychological test measures and RTW is not found when examining mTBI. This is partly due to the lack of high-quality published basic studied. For example, a systematic review of RTW following mTBI by the WHO Collaborating Centre Task Force on mTBI found only four studies that met their inclusion criteria (Cancelliere et al., 2014). The authors concluded that most mTBI patients RTW within 3-6 months and predictors of delayed RTW were less than 11 years of formal education, nausea or vomiting in hospital, extracranial injuries, severe pain early after injury, and job independence or decision-making latitude. Neuropsychological test performance was not predictive of RTW status following mTBI. Similarly, a well-designed longitudinal study that examined factors related to RTW following mTBI consecutively recruited 145 patients with mTBI in Finland upon hospital admission (Wäljas et al., 2014). They found that older age, multiple bodily injuries, and fatigue significantly predicted slower RTW. Once again, performance on neuropsychological test measures did not significantly predict RTW status. In addition, a Canadian longitudinal study that recruited 85 patients with mTBI found that neuropsychological test measures were not related to RTW one-year post-injury (Nolin & Heroux, 2006).

In sum, there appears to be an interaction effect between TBI severity and neuropsychological test performance with RTW. There is some evidence to support the relationship between neuropsychological test performance and RTW in moderate-severe TBI but none in mTBI. Analogous to the early technological limitations in detecting evidence of mTBI in neuroimaging described above, it may be that neuropsychological test measures are not sensitive enough to detect the subtle cognitive impairments that occur following a mTBI. Some have argued the need for sufficiently more demanding tasks to better detect the subtle deficits in mTBI (Cicerone, 1997; Bernstein, 1999; Frencham et al., 2005; Paré et al., 2019).

1.7 Additional Issues with Neuropsychological Assessment

In addition to the insensitivity of tests, there may be other issues surrounding neuropsychological assessment. First, the testing environment is artificial as it takes place in a heavily controlled and contrived setting. The testing environment is often quiet, distraction free, with strict pre-determined rules, and with behavioural prompts given by the clinician (Manchester, Priestly, & Jackson, 2004). The argument has been that such control is required for standardized testing and that if the examinee has difficulty performing in such ideal settings, then it is reasonable to assert that they would exhibit impairment in the real-world. However, the testing environment does not reflect real-life (Chaytor & Schmitter-Edgecombe, 2003). Clinically, we need to be mindful of the plethora of variables that need to be considered when comparing a controlled testing environment with the unpredictable and variant nature of the real-world (Sbordone, 2001).

Second, data are collected about the examinee over a short sampling period and used to predict their behaviours over long periods and over different situations. Such an oversimplification undermines the complexity of the human psyche and does not take into consideration intraindividual vacillations in performance (Zakzanis & Jeffay, 2011) or state and trait characteristics of the individual (Bannon, Gonsalvez, Croft, & Boyce, 2006). Neuropsychological performance represents a snapshot of an individual's level of functioning that is predicated on the interaction between the internal and external circumstances surrounding them during the testing.

Third, there is considerable overlap of cognitive domains across and within tests used. For example, despite a lack of consensus for the term executive functioning, tests have been classified based on the researcher's inclination; the controlled oral word association test (COWAT) has been classified as a test of executive functions or language or both depending which text one follows (Lezak, Howieson, Bigler, & Tranel, 2012). Further, most tests measure a range of abilities that are beyond the target ability (Long, 1996). Compensatory strategies may artificially inflate performance and tests are not sensitive to these strategies.

Fourth, sensitivity and specificity of tests vary based on group characteristics. Most neuropsychologists are well aware of this and carefully select tests that have been shown to be sensitive to their patient's diagnostics. However, some tests are almost universally used (Rabin, Barr, & Burton, 2005; Rabin, Paolillo, & Barr, 2016), such as the Stroop, TMT, and the WCST despite the lack of compelling evidence for their inclusion other than familiarity through training.

Lastly, neuropsychological test measures were never developed with the intention of predicting everyday functioning. The lack of sensitivity and predictive ability of neuropsychological test measures may be due to the evolution of the field of clinical neuropsychology. Neuropsychology is a young field but has undergone several paradigm shifts over the past 70 years (Bilder, 2011). In its infancy (~1950), neuropsychology related behavioural manifestations to brain localization using a flexible battery of cognitive tests. The neuroimaging revolution (~1980) allowed clinicians to locate brain abnormalities with increasing accuracy and sensitivity. However, brain imaging has its limitations, especially in diagnosing conditions where lesions or infarcts are not apparent on imaging studies (i.e., mTBI, early dementia, and psychological/psychiatric disorders). To these ends, neuropsychological assessments held value in determining the nature and severity of behavioural manifestations that may arise. Thus, the neuropsychologist's role shifted from localizing site and size of lesions to characterizing the cognitive strengths and weaknesses of an individual using extensive normative data and standardized psychometric measures. Some evidence for the continued use of traditional tests to estimate strengths and weaknesses was published in addition to new measures that were developed and validated to meet this new need (for an excellent review, see Marcotte & Grant, 2010). Currently, neuropsychology is undergoing another shift: expanding the characterization of cognitive strengths and weaknesses and relating it to functional impairment. Despite all the advances in modern neuroscience, little is known about how the brain allows us to interact with the external world to complete daily tasks, eve simple ones such as cooking (Burgess et al., 2006). Accordingly, researchers have attempted to develop new tools that are grounded in cognition theory and maintain the psychometric rigor that neuropsychologists are accustomed to but are designed to be reflective of real-world performance (Marcotte & Grant, 2010). In other words, devising measures that relate to real-world impairment such as resuming employment, being able to live independently, managing personal finances, preparing meals, driving, and engaging in care-giving activities. Here, the assessing neuropsychologist is asked by the referring party to articulate the veracity, breadth, and severity of a patient's cognitive dysfunction as it pertains to everyday functioning (see Marcotte & Grant, 2010). In fact, the majority of modern-day referrals to neuropsychologists are concerned with a patient's capacity to RTW or school or to perform iADLs such as medication management, shopping, cooking, driving and managing their finances (Rabin et al., 2005). Just prior to this shift, justifications about real-world impairment were predicated on performance on traditional tests and its association to certain activities of daily living based on face validity. The evidence of the link between test performance and functional ability was lacking, which called into question the clinical usefulness of a neuropsychological assessment (Ruff, 2003). To this end, traditional tests such as the WCST (Heaton, Chelune, Talley, Gary, & Curtiss, 1993), the Tower of London Revised (TOL; Culbertson & Zillmer, 2001), and the TMT (Reitan, 1979) that were originally designed to aid in lesion site and location may not be adequately equipped to address clinical questions regarding real-world functioning.

This is troubling since neuropsychological test findings can weigh significantly into the opinions related to the degree to which an individual can be expected to successfully reintegrate into their ADL (Acker & Davis, 1989; Chaytor & Schmitter-Edgecombe, 2003; Lezak et al., 2012; Sherer et al., 2002). These opinions facilitate the planning and implementation of rehabilitation programs to expedite reintegration and maximize the quality of life for persons with cognitive impairment. For these ends, if the measures employed lack adequate predictive validity to support inferences made about real-world outcome and functioning, it has been argued that clinical neuropsychologists evaluating RTW following a single mTBI may be basing their conclusions on weak scientific grounds[1] (Watt & Crowe, 2018). In addition, we may be providing a disservice to our patient's wellbeing. Thus, it is of unequivocal importance that our

[1] This statement is not a general sweeping statement for the practice of clinical neuropsychology as a whole. There are ample lines of evidence for its use and utility in the surgical planning, assessment, and treatment effects of epilepsy along with differential diagnosis of dementias (see Lezak et al., 2012 for a review).

neuropsychological test measures validly predict everyday functioning. In other words, neuropsychological test measures need to be more *ecologically valid*.

1.8 Ecological Validity

Ecological validity was initially coined by Brunswik (1955) to refer to the generalizability of controlled experimental findings to the real-world. It was later adapted to neuropsychology by Sbordone (1996) to refer to "the functional and predictive relationship between the patient's performance on a set of neuropsychological test measures and behavior in real-world settings (e.g., at home, work, school, community, etc.)" (p. 23). In other words, how can the performance in an artificial test environment relate to the real-world functional ability, which can contain numerous extraneous and distracting stimuli.

Ecological validity is generally approached in two ways; veridicality and verisimilitude. First, *veridicality,* is defined as "the extent to which test results reflect or can predict phenomena in the open environment" (Franzen & Wilhelm, 1996, p. 93). This is relevant to traditional pen-and-paper neuropsychological test measures that consociate performance to real-world functioning by way of empirical measurement and statistical analyses (see Franzen & Wilhelm, 1996; Hammond, 1998; Kvavilashivili & Ellis, 2004). Most traditional pen-and-paper neuropsychological test measures measure domains of cognitive functioning. Performance on these measures may predict problems with everyday functioning based on the assumption that the test measures assess the functions and constructs that are significant to the successful completion of real-world tasks. How performance on these traditional pen-and-paper neuropsychological test measures relate or predict real-world functioning, such as RTW, is a measure of their veridicality.

Verisimilitude however, is described as the degree to which a neuropsychological test resembles tasks encountered in everyday life (also see Franzen & Wilhelm, 1996). The more closely a test resembles real-world tasks or behaviours, the smaller the gap between test performance and real-world performance (Spooner & Pachana, 2006). Considering that the definition of *everyday functioning* may differ from person to person (e.g., consider the differences between an investment banker and a homemaker), 100% verisimilitude across all types of patients would not be possible. Moreover, it would not be efficient to construct and validate tests that are reflective of each individual's particular set of everyday functions.

It has been argued that neuropsychological test measures lack verisimilitude (Chaytor & Schmitter-Edgecombe, 2003; Farley, Higginson, Sherman, & MacDougall, 2011; Franzen & Wilhelm, 1996; Marcotte & Grant, 2010). Consequently, the lack of verisimilitude presents a formidable problem particularly in the context of a clinical disorder such as mTBI where the maximum prevalence of persistent neuropsychological impairments on formal testing after three months post-injury is likely to be little to none. Yet, real-world disability in the everyday lives of patients beyond the expected benchmark for recovery are common (see Binder, 1997). To this end, it may be worthwhile to develop new measures of cognitive functioning with better sensitivity and ecological validity, which will ultimately be more congruent with what we are asked to address within the context of a neuropsychological examination (e.g., whether a patient can go back to work).

1.9 Ecologically Oriented Tests

Ecologically oriented tests have been developed to address the lack of verisimilitude in neuropsychological testing but little to no research has been conducted on a sample of patients with mTBI in the post-acute period of recovery. For example, the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996) is a set of neuropsychological test measures with purported ecological validity and is based on the concept of verisimilitude. The BADS is a battery of tests used to evaluate instrumental activities of daily living that may be due to impairments related to dysexecutive syndrome. The six subtests are designed to estimate executive functional aspects such as: planning disorders, inhibitory control, problem solving, temporal judgment and behavioural alterations. The BADS has demonstrated good concurrent validity with other measures of executive dysfunction (Bennett, Ong, & Ponsford, 2005; Norris & Tate, 2000). In addition, it was found to be significantly related to informant-ratings of disability on the Dysexecutive Questionnaire (DEX; Wilson et al, 1996), suggesting good ecological validity (Evans, Chan, McKenna, & Wilson, 1997; Wilson et al., 1996). It has also been found to be superior in its ecological validity in comparison to other neuropsychological test measures such as TMT (Lezak et al., 2012) and WCST, as measured by the clinician-ratings on the Role Functioning Scale (McPheeters, 1984; Norris & Tate, 2000) and DEX (Wilson et al., 1996; Bennett et al., 2005). However, a dearth of supporting research, poor psychometric properties and norms, ceiling effects in patients with subtle cognitive dysfunction, limit its clinical effectiveness and reliability (Marcotte & Grant, 2010; Norris & Tate, 2000). In addition, criticisms regarding the lack of similarity between the subtests and real-world situations have questioned its ecological validity (Acker, 1990; Alderman, Burgess, Knight, & Herman, 2003).

1.10 Naturalistic Assessments

Another creative alternative using the verisimilitude approach involves testing in the real-world using naturalistic assessments. Here, some researchers have developed tests that simulate vocational environments such as situational vocational evaluations (LeBlanc, Hayden, Paulman, 2000) and other situations of daily activities such as the Multiple Errands Test (Alderman et al., 2003). Here, the assessor would directly observe the patient simulating actual work-related activities and then rate them with respect to their visual processing, memory and executive functioning (LeBlanc et al., 2000). Although this method appears to be high in verisimilitude, it is plagued by its non-standard approaches, which question the reliability and validity of the generalizations and comparisons that can be made within and between patients (Marcotte & Grant, 2010; Robertson & Schmitter-Edgecombe, 2017).

1.11 Virtual Reality

Virtual Reality (VR) combines the advantages of naturalistic assessments whilst potentially minimizing the disadvantages such as cost, lack of control over stimulus delivery, and high resource demands. Analogous to the revolution of medical imaging, technological advances such as VR have demonstrated its ability to compensate for the lack of representativeness in traditional pen-and-paper neuropsychological assessment (Banville et al., 2010). VR allows individuals to interact with and become immersed into a 3D computer generated environment (Lengenfelder, Schultheis, Al-Shihabi, Mourant, & Deluca, 2002). The use of VR in a clinical capacity is increasingly common including: psychiatric treatment (McLay et al., 2011), neuropsychological rehabilitation (Trepagnier, 1999; Wilson, Foreman, & Stanton, 1997), neurological rehabilitation (Lange et al., 2011) and surgical training (Seymour et al., 2002). Further, the application of VR to increase ecological validity in neuropsychological assessment is also becoming increasingly prevalent (Campbell, et al., 2009; Kang et al., 2008; Ku et al., 2003; Mraz, et al., 2003; Parsons et al., 2012; Zakzanis, Quintin, Graham, & Mraz, 2009). For example, Virtual Environments (VE) have been constructed to simulate a number of real-world tasks measuring aspects of executive functioning (i.e., planning, multi-tasking, etc.) in variable contexts, such as driving (Barkley, Murphy, O'Connell, Anderson, & Conner, 2006; Lengenfelder et al., 2002; Plancher, Nicolas, & Piolino, 2008) and shopping (Jovanovski, Zakzanis, Campbel, Erb, & Nussbaum, 2012; Raspelli et al., 2010). VEs have also been designed to measure other aspects of cognition in an albeit, virtual, real-world setting such as spatial navigation and memory by way of city mazes (Tippett et al., 2009).

The use of VR offers several advantages over traditional pen-and-paper neuropsychological test measures. One of the most obvious of these is immersion of the patient in a realistic VE which would allow the patient to be assessed in different types of real-world situations without actually physically constructing such environments (Schultheis, Himelstein, & Rizzo, 2002). Additionally, VR allows the clinician or researcher to directly observe an individual's functionality in a realistic situation, while placing strict control on the environment (Schultheis et al., 2002). For example, Rizzo and colleagues (2002) used an office VE to determine the relationship between memory abilities and employment using a functionally and vocationally relevant setting with patients who had sustained a TBI. Within their office VE, prospective, incidental, short- and long-term recall, and recognition were assessed using objects around the office. Distractions were also creatively implemented by way of another office with different objects and through interactions initiated by the office avatars (virtual co-workers; Rizzo, Schultheis, Kerns, & Mateer, 2004). Further, Rizzo et al. (2000, 2002) assessed attention processes in children with Attention Deficit Hyperactivity Disorder (ADHD) using a VR classroom. Here, reaction times were measured while a series of typical classroom distractors (e.g., ambiance noise, peripheral motion, etc.) were systematically presented. While accounting for hyperactive motor movements, the VR program was able to differentiate between groups with 100% accuracy (Rizzo, et al., 2002). Their results suggest that the use of VR test measures within a clinical framework to assess clinical disorders such as ADHD is promising.

A further advantage is that VEs also provide an increased approximation of the natural environment, and the subject is able to use multiple cognitive domains while completing a task as they would in the real-world. In other words, since neuropsychological functioning in the real-world is multi-factorial, VEs can be tailored to parallel these environments and to ultimately increase verisimilitude and in turn the ecological validity. In addition, VEs can be developed to measure those cognitive constructs that have historically been difficult to measure using traditional pen-and-paper tasks (i.e., executive and attentional functions; Lezak et al., 2012).

1.12 Overview of Studies & Hypotheses

In sum, the issues surrounding neuropsychological assessment are (i) a shift in the referral question from the site and size of the potential lesion to functional capacity; (ii) a lack of evidence supporting the ecological validity using traditional pen-and-paper neuropsychological test measures in mTBI to determine vocational status (iii) potential for ecologically oriented tests but limited by way of the lack of studies validating their use in mTBI patient populations; (iv) the need to develop better and ecologically valid tools; and (v) the advent and possible utility of VR as an ecologically valid measure to better predict vocational status. Consequently, the overarching objectives of the current set of studies was to evaluate the ecological validity of neuropsychological test measures as they pertain to mTBI. To address these objectives, a series of four studies were undertaken:

Study 1. An up-to-date meta-analysis of the sensitivity of traditional pen-and-paper neuropsychological test measures in the post-acute period of mTBI using a random-effects model was conducted. Based on the findings from previous studies (e.g., Binder et al., 1997; Rohling et al., 2012; Schretlen & Shapiro, 2003), it was hypothesized that domain level effect sizes on neuropsychological test performance would not be significantly different between controls and patients who were in the post-acute period of recovery following a mTBI.

Study 2. This study examined archival data so to understand veridicality. This study was undertaken to investigate whether traditional pen-and-paper neuropsychological test measures could differentiate between RTW status in a sample of patients who were in the post-acute period of recovery following a mTBI. It was hypothesized that traditional pen-and-paper neuropsychological test measures would not be sensitive to RTW status.

Study 3. In a similar manner to *Study 2,* this study explored the veridicality and verisimilitude approaches in predicting RTW using archival data. Traditional neuropsychological test measures were employed for the veridicality approach and the BADS was employed for the verisimilitude approach. Based on the above literature review, it was hypothesized that BADS would be more sensitive than traditional neuropsychological test measures in their ability to predict RTW, but the effect sizes are expected to be small to moderate (Norris & Tate, 2000).

Study 4. Two novel VR tests of attention and executive functioning were developed to improve ecological validity of neuropsychological assessment in patients who were in the post-acute recovery period following a mTBI. These measures were meant to be representative of the type of activities one would be engaged in as a general labourer (e.g., assembly line worker or a courier). The VR tests were compared with their traditional test counterparts to determine relative sensitivities in RTW. The VR tasks were developed to be high in verisimilitude and were hypothesized to be more sensitive to RTW status than traditional tests.

Chapter 2

2 Study 1: Neuropsychological Outcome Following the Post-Acute Phase of Mild Traumatic Brain Injury: A Meta-Analysis

2.1 Narrative vs Quantitative Reviews

Several narrative reviews have been published describing the impact on cognition following mTBI (see Erlanger et al., 1999; Ewing-Cobbs & Barnes, 2002; Goleburn & Golden, 2001; Maroon et al., 2000; McCrea et al., 2009; McCrea, Kelly, Randolph, Cisler, & Berger, 2002). Narrative reviews are useful in that they document the unfolding story of a particular research theme by summarizing and synthesizing conclusions across many studies with similar aims. In other words, authors of a narrative review interpret words using their own words to draw conclusions on real-world phenomena and to provide a basis for further work in the field (Ellis, 2010). Indeed, such reviews are valuable and are a step in the hierarchy of evidence as noted in the widely acclaimed 6S model of evidence-based practice – a standard for healthcare professionals to find evidence quickly and efficiently (see DiCenso, Bayley, & Haynes, 2009).

In a narrative review, the objectivity of the results are at risk as they are (by their very nature) highly susceptible to review bias. Although there is usually an effort taken to control for the quality of the studies reviewed, they are rarely comprehensive and frequently come to the wrong conclusion mainly due to vote-counting methods (Borenstein, Hedges, Higgins, & Rothstein, 2010). Here, all studies that lie on one or the other side of a subjectively attributed threshold are tallied and the side with the most votes is considered evidence of a general trend. Unfortunately, the probability of detecting an effect falls as the amount of evidence increases (Hedges and Olkin, 1985). Additionally, differences in the precision of estimates are ignored in narrative reviews such that studies with large effects but with low precision are likely to attract more attention than studies with small but precise effects (Borenstein et al., 2010).

A quantitative systematic review, by contrast, offers several advantages over conventional narrative reviews. For instance, a meta-analysis incorporates individual effect size estimates from different, but similar studies to combine them into a mean overall effect size. It is an objective, quantitative audit of the available and often (estimate of) unpublished research, with an emphasis on cumulating data as opposed to potentially subjective conclusions of individual narrative reviews. It is transparent as each study's methods and data are carefully evaluated against specific inclusion and exclusion criteria. Most importantly, a meta-analysis can provide objective answers to questions regarding the nature of an effect even in the presence of conflicting findings.

Accordingly, the main intention of this chapter is to review the current body of knowledge on mTBI and to provide a comprehensive and up-to-date meta-analysis describing the effects of mTBI on neuropsychological test performance in the post-acute injury period (i.e., > 3 months following the time of injury). To date, several similar but different meta-analyses have been conducted in the past (Binder et al., 1997; Zakzanis, Leach, & Kaplan, 1999; Schretlen & Shapiro, 2003; Belanger & Vanderploeg, 2005; Frencham, et al., 2005; Pertab, James, & Bigler, 2009; Rohling et al., 2011), each with their advantages and limitations. A brief chronological review will follow.

2.2 Review of Previous mTBI Meta-Analyses

The first meta-analysis was conducted by Binder and colleagues (1997). They performed a meta-analysis to review the neuropsychological outcome after 'mild head injury' (the term that was the predecessor to mTBI). They hypothesized that effect sizes would be significantly greater than zero both in general and across cognitive domains. They reviewed studies from 1986-1994 based on 'MEDLINE' and 'PsyScan: Neuropsychology' to determine inclusion eligibility. Here, the inclusion criteria consisted of (a) participants must have incurred a mTBI (GCS $\geq$ 13); (b) enough information must be reported to calculate effect sizes; (c) less than a 50% attrition rate; (d) mTBI data was separated from other TBI severities (if applicable); and (e) only data for adults were evaluated. Furthermore, an additional criterion that was not formally mentioned was that of chronicity and the use of an appropriate control group. Binder and colleagues only included studies that examined patients at least three months post-injury. Their search resulted in 11 studies that were eligible, which included a total of 314 patients with mild head injury and 308 control subjects. Importantly, their results showed that the mean sample-weighted effect sizes produced varying results based on the statistical method employed. Using the more liberal effect size statistic, Cohen's *d*, the overall effect size (averaged across all domains) was $d = 0.12$ (SD 0.18) which was statistically significant from zero. Whereas, the more conservative effect size method of Hedge's *g* statistic (with pooled standard deviations) produced an overall effect size that was not statistically significant from zero ($g = 0.07$, $SD = 0.17$).

Across cognitive domains, Binder and colleagues (1997) found that effect sizes in the attention domain were significantly different than zero, indicating that differences in neuropsychological test performance between groups was present. The authors concluded here that tests of attention were most sensitive in differentiating between mild head injury and controls. However, in more practical terms, they illustrated that the overall effect size was equivalent to 2-points on the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) IQ test or on the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987) General Memory scale. Similarly, the effect size found in the attention domain was equivalent to just 3 points on the Attention/Concentration Index of the WMS-R. Both of these estimated point values are smaller than the reported measurement error (Wechsler, 1981; 1987) and are thus, practically undetectable. They also estimated a positive predictive value of only 32% based on a 5% prevalence of brain injury when the sensitivity and specificities were artificially set at 0.90. In other words, even if neuropsychological procedures had an improbably high rate of 0.90 sensitivity and specificity values in determining subtle brain dysfunction, the probability that neuropsychological test performance alone could differentiate patients with subtle brain dysfunction would only be 32% (see Table 3 of Binder et al., 1997).

Their work, while important and novel, was not without limitations. Firstly, the homogeneity of their sample was confounded by the inclusion of sports-related mTBIs which have been documented as having different neuropsychological and recovery outcomes (Belanger et al., 2005; McRory et al., 2018). Moreover, they included studies with patients with mTBI who deviate from the current standards (i.e., ACRM definition of mTBI; Kay et al., 1993) in that they included patients who sustained a PTA period of greater than 24-hours, LOC greater than 30-minutes, and included patients with mTBI that had positive neuroimaging findings. In other cases, the authors included studies where the mTBI samples did not actually sustain an injury but rather were exposed to simulated high altitude oxygen deprivation. In addition, possible practice effects confound their data collection as some of the studies collected data on patients as part of a longitudinal treatment with multiple assessments at different time points with some of them being within the acute post-injury period.

Following this, Zakzanis and colleagues (1999) expanded on Binder's original meta-analysis by examining effect sizes for individual tests in an effort to determine the practical sensitivity of neuropsychological test measures. These authors searched through studies published between

1980 and 1997 in several key journals (see Table 3.1 from Zakzanis et al., 1999) along with PsycINFO and MEDLINE index scans that compared cognitive performance between controls and patients with mTBI. Formal inclusion criteria were as follows: (a) a diagnosis of mTBI based on GCS scores of 13-15, PTA < 24 hours, LOC < 20 minutes, and a normal neurological exam with negative neuroimaging findings; (b) a control group must have been incorporated; and (c) reported statistics that can be converted to effect sizes. Overall, 12 studies were included in the meta-analysis with a total sample of 952 patients with mTBI and 495 controls. Effect sizes were calculated based on specific neuropsychological test measures. Commentary of these results was provided based on their categorization into seven cognitive domains: cognitive flexibility and abstraction ($d = 0.72$, $SD = 0.48$), attention/concentration ($d = 0.63$, $SD = 0.31$), memory acquisition ($d = 0.69$, $SD = 0.42$), delayed recall ($d = 0.71$, $SD = 0.43$), verbal skill ($d = 0.62$, $SD = 0.35$), performance skill ($d = 0.47$, $SD = 0.35$), and manual dexterity ($d = 0.44$, $SD = 0.11$). Their justification for assignment of specific tests to these cognitive domains was based on a combination of factor analytic research, published neuropsychological meta-analyses, and Lezak and colleagues' (2012) theoretical and practice-related *a priori* domains. Their findings indicated that tests of frontal dysfunction (e.g., phonemic and design fluency, WCST perseverative variables) along with encoding and retrieval from declarative memory were most impaired in mTBI sample as compared to controls.

While novel and practically oriented, this study was also not without limitations. Firstly, while effect size estimates based on specific neuropsychological test measures is a useful estimate of test specificity, this method sacrifices sample size. Unlike the minimal assessment of cognitive function in multiple sclerosis (MACFIMS; Benedict et al., 2006), a standard battery for the evaluation of mTBI does not exist. As a result, researchers use and report different measures in the literature when evaluating mTBI. Accordingly, grouping by test versus grouping by domains results in smaller sample sizes per group. This was evident as per table 14.2 from Zakzanis et al., (1999) where sample sizes, by which the mean estimated effect size was calculated from, ranged from $n = 1$ to $n = 3$. Moreover, estimated mean effects were not appropriately weighted and cognitive specific or an overall effect size was not reported. Lastly, the time since injury was not reported and thus, it is not possible to delineate the impact of acute versus post-acute cognitive performance.

Schretlen and Shapiro (2003) asked *what is the size of the effects of mild head injury and moderate-severe TBI on cognitive functioning*? They pursued this question by utilizing a meta-analysis of the literature across the spectrum of TBI from the acute period to the post-acute period. They only included studies of patients that were recruited either prospectively or having a history of TBI. The authors searched through MEDLINE and PsycINFO databases from 1984 through to February 2003 with the following criteria: (a) English; (b) adults (or older adolescents); (c) compared patients with control subjects; (d) effect sizes could be calculated based on the published results; (e) separated severities (mild vs moderate-severe); and (f) no whiplash or penetrating head injuries. Their search resulted in 39 articles with 48 total comparisons, which included 1,716 patients and 1,164 controls. They calculated two versions of the mean effect size estimate; one using the pooled standard deviation estimate (d_{pooled}) and another using a more conservative approach of using the control group's standard deviation as the overall standard deviation estimate (d_{control}). Individual test result comparisons produced 418 estimates of $d_{\text{pooled}} = -0.46$ ($SD = 0.40$) and 409 of $d_{\text{control}} = -0.66$ ($SD = 0.61$), which were both significantly different than zero as per a one-sample t-test ($p < 0.0001$). They parsed their results based on severity. To this end, the mean effect size estimate for mTBI was found to be $d_{\text{pooled}} = -0.24$ and $d_{\text{control}} = -0.31$, which were both significantly different than zero ($p < 0.0001$). They also compared the d_{pooled} and d_{control} effect sizes by periods of time since injury; <7 days, 7-29 days, 30-89 days, and >89 days. For the mTBI group, their results indicated that there were no significant differences between the groups at 30-89 days post-injury. In other words, they found that the overall effect size estimates of cognitive functioning between mTBI and controls were insignificant at 30-89 days post-injury, regardless of using the overall standard pooled variance or the more conservative control values for variance. This finding was in contrast to that of moderate-severe TBI, which was significantly different across all time periods. The authors concluded that patients with mTBI performed at around the 33rd percentile range of their matched controls at six-days post-injury, which increased to 48th percentile of their matched controls at 30-89 days, and up to 55% of their matched controls at the >89 day time period. Though comprehensive, the study also has a few limitations. First, the majority of the studies identified for the meta-analysis were cross-sectional and thus, the consistency of the reported performance could not be evaluated. This is a limitation of the literature more so than a limitation of this particular study as there are few studies that follow patients across the different stages of recovery following a TBI. Additionally, moderate and severe patients with TBI were grouped together, resulting in their findings either relating an under- or over-estimate of the effect sizes of one or the other severity. Lastly, they did not group the effects into cognitive domains, which would have been greatly beneficial to the clinician and researcher.

Belanger and colleagues (2005) also aimed to identify the effects of mTBI on neuropsychological test performance with particular focus on performance across cognitive domains and to determine the differences in estimated mean effect sizes between time since injury and sample characteristics (i.e., litigation vs unselected samples vs clinic-based samples). Studies included into the meta-analysis were published between 1970 to March 2004 in Pubmed and PsycINFO indices. Their inclusion and exclusion criteria were as follows: (a) non-sports related studies; (b) exclusion of whiplash studies; (c) mTBI had to be defined; (d) the mTBI group must be compared to a control group; (e) if other severities were compared, mTBI findings must be separated; (f) samples must be compared against neuropsychological test measures (either experimental or standardized); (g) enough statistics must be reported to calculate an effect size; (h) adults and adolescent data only, no children; and (i) time since injury must be reported. Their search returned 133 potential studies from which, 39 studies met eligibility. A total of 41 effect sizes were extracted, with careful control of studies that reported on the same data across publications. These effect sizes were grouped into nine cognitive domains based on highly cited texts (Lezak et al., 2012; Spreen & Strauss, 1998); Global Cognitive Ability, attention, executive functioning, fluency, memory acquisition, delayed memory, language, visuospatial ability, and motor abilities. Overall estimation of the mean effect size across all studies and domains was found to be $d = 0.54$[2], which was significantly different from zero in one-sample t-test ($p < 0.05$). Parsed by cognitive domain, all estimated mean effect sizes resulted in moderate-large effects (Cohen, 1988) and were all significantly different than zero ($p < 0.05$) with the exception of motor abilities, which the authors reasoned was due to having only two samples for that estimation. Their findings indicate that the largest effect sizes were found in the domains of

[2] The authors calculated effect sizes in a way where positive values were indicative of better performance by the control group

fluency (*d* = 0.77) and delayed memory (*d* = 0.69) with the smallest in motor abilities (*d* = 0.16) and executive functions (*d* = 0.21). Moderator analysis revealed that as time since injury increased, the overall effect size decreased with the exception of visuospatial skills which was found to have increased. Furthermore, analysis of the sample selection (i.e., litigation vs clinic-based vs unselected samples) revealed no effect size differences in unselected samples in the post-acute period across all cognitive domains with the exception of the fluency domain. Here, the authors reasoned however, that the results from the fluency domain were largely attributable to an outlier study (Mangels et al., 2002). As a result, they dismissed this as a real finding and concluded that their analysis found evidence for full neuropsychological recovery by 3 months postinjury after a mTBI. Additionally, no significant differences were found when comparing litigation-based samples with unselected samples. However, they found that the unselected samples improved with time, whereas the litigation group did not. Moderator analysis did not reveal any differences in effect size estimates between studies that used patients in litigation (that had passed validity screening measures) vs non-litigating samples. These effect sizes were also comparable in the post-acute period. The authors concluded that cause of ongoing cognitive difficulties found by the litigating sample was not clear and that further investigation was required.

While informative and effectively building on the findings and methods of Schretlen and Shapiro colleagues (2003), there are key limitations to Belanger et al.'s (2005) review. Firstly, the number of studies that were grouped into a cognitive domain were not equal and the final overall population effect size was not weighted. To this end, some studies had a few samples (i.e., fluency and language) but contributed equally to the overall domain mean effect size as compared to domains that had many samples. As such, the overall population effect size estimate was skewed and may be a less than accurate representation of the overall population effect size. On the other hand, domains such as attention garnered a plethora of samples which would likely represent a more stable approximation of the population estimated mean effect size. In addition, coding of moderating variables such as sample selection was limited as it was not always reported by the original studies. For example, it was not always reported if the sample was sourced from patients in litigation or not. Here, the authors simply coded whether or not the samples included some, if any, patients who had a pending legal case as it pertains to their brain injury status and claim. The authors cautioned that the results from the moderator analysis were inaccurate due to inconsistent reporting by the original studies.

In an effort to provide an update to the meta-analysis authored by Binder and colleagues (1997), Frencham et al. (2005) used similar procedures with the inclusion of updated studies from 1995-2005. Whereas Binder and colleagues (1997) only included studies that had reported in the post-acute periods of mTBI, Frencham et al. (2005) included studies with mTBI at all post-injury periods. Moreover, particular focus was given to determine the relationship between cognitive domain and recovery at various instances post-injury. It was hypothesized that overall performance would be significantly poorer for the mTBI group as compared to controls. In addition, they expected the largest effect sizes to be in the domains of speed of processing and attention.

Frencham and colleagues (2005) reviewed PsycINFO for mTBI publications in addition to gathering studies from the reference sections of relevant reviews in order to increase their total sample size of studies. The authors separated their inclusion and exclusion criteria. To this end, their inclusion criteria included: (a) studies published from 1995-2005; (b) written in English; (c) compared mTBI against a control group; and (d) data that had not been originally analyzed by the Binder et al. (1997) study. The exclusion criteria were as follows: (a) child mTBI; (b) mTBI data was not separated from other severities; (c) presence of symptomatology to necessitate mTBI inclusion into the study; (d) GCS scores less than 13; (e) whiplash or non-impact head injuries; and (f) greater than 50% attrition in case of follow-up studies. Their search resulted in 17 studies that were eligible with a total patient sample size of 634 and 485 healthy controls. They calculated mean effect size estimates across all studies and for each cognitive domain. Effect sizes in the form of Hedges' *g* were calculated with the pooled (control and mTBI) standard deviation while correcting for sample size as per Hedges and Olkin (1985). Here, the authors parsed the data based on time since injury; acute and post-acute. Acute effect size estimates across all studies was found to be $g = 0.33$[3] ($SD = 0.24$), which was significantly different from zero ($t(11) = 4.81$, $p < 0.0005$). The post-acute data was parsed further into current

[3] With a positive charge indicative of better performance by the control group

findings, which were effect size estimates based on studies since Binder et al., (1997), and combined with the results found by Binder et al., (1997). Current findings revealed an effect size estimate of g = 0.28 (SD = 0.34) and the combined overall estimated effect size was g = 0.11 (SD =0.30). Neither of the post-acute effect size estimates met statistical significance. The authors also reported an overall (combination of acute and post-acute data) estimated effect size of g = 0.32 (SD =0.26), which was significantly different from zero as per a one-sample t-test ($t(16) = 5.01$, $p < 0.0005$).

Frencham and colleagues (2005) also calculated effect size estimates for cognitive domains with similar domain groupings as per Binder and colleagues (1997) with the exception of combining memory acquisition and delayed memory into one domain; memory. Effect sizes were grouped in the following seven domains: working memory/attention, perceptual organization, verbal comprehension, motor skills, memory, executive functioning, and speed of processing. The calculated effect size estimates for the domains of working memory/attention, memory, executive functioning, and speed of processing were found to be small to moderate (Cohen, 1992) and statistically significant as per a one-sample t-test. The largest of these, speed of processing, yielded and estimated mean effect size of g = 0.47 (SD = 0.25). No significant differences were reported within the domains of perceptual organization, verbal comprehension, and motor skills. Further analysis was completed to investigate the possibility of a moderating effect of time since injury. A Spearman's rank order correlation was conducted and a significant correlation between time since injury and overall effect size was found, with 22% of the variance being accounted for by this moderating variable. This finding was replicated across each neuropsychological domain with a significant correlation found between time since injury and the estimated mean effect size for memory. Moreover, their data analysis revealed an overlap percentage of 78.7%, indicating only a small degree of discriminability between the two groups. From a cognitive domain perspective, the largest degree of non-overlap was found in the speed of processing domain with 33% of the performance by the mTBI not being shared by the controls. Finally, approximately 2.70% of the total variance across all neuropsychological test measures and 5.20% of the speed of processing results were accounted for by group membership (either mTBI or control).

The authors concluded that their combined findings (g = 0.11) were consistent with the findings reported by Binder and colleagues (d = 0.07; 1997) with small but statically insignificant positive effect sizes of mTBI on cognitive performance with a trend towards slightly larger effect sizes. The authors attributed this trend towards the influence of data collected in the very acute period (e.g., < 24 hours since injury). From a cognitive perspective, they also found small and significant effect sizes under the combined domain label of attention and concentration with additional demonstrated effect size differences in the domains of memory and executive functioning. In addition, the authors stated that their results were in line with other authors (Binder et al., 1997; Dikmen, McLean, & Temkin, 1986; Dikmen et al., 1995; Ponsford et al., 2000) in that the majority of the recovery occurred within the first three months following injury.

While the study was important with respect to replication purposes, some limitations were present. Namely, the homogeneity of the sample was somewhat compromised with the inclusion of studies that included sports-related studies (i.e., football players) and the large variability in the definition used for mTBI across studies. Regarding the former, there is evidence that the biomechanics and medical attention seeking behaviour of sports-related mTBIs and motor vehicle accidents are quite different (Belanger et al., 2005; Pertab et al., 2009). With respect to the latter, the definition ranged from "a blow to the head causing individual to stop what they were doing due to a loss of consciousness" (Frencham et al., 2005, p. 338) to the definition as per the ACRM (Kay et al., 1993). Although the authors separated their effects by acute and post-acute periods, they did not, however stratify further by cognitive domain. This was especially relevant since their results of the overall neuropsychological findings indicated that there was a significant difference between the groups in the acute period but not in the post-acute period.

Pertab and colleagues (2009) highlighted the limitations of meta-analysis techniques used to derive meaning in mTBI recovery and that the generalizability of the previous meta-analyses was called into question. Accordingly, they attempted to reanalyze the data from Binder and colleagues (1997) and Frencham et al. (2005) using a fixed-effects model[4] parsed by mechanism of injury, diagnostic criteria employed, type of neuropsychological test measures employed, and

---

[4] A fixed-effects model assumes that the effect sizes were derived from the same sample/population.

symptomatic vs non-symptomatic patients with mTBI. They hypothesized that these factors may account for the unaccounted variability reported in previous meta-analyses. Although no formal inclusion/exclusion criteria were listed, they included all the studies that both Binder et al. (1997) and Frencham et al. (2005) included within their studies. A total of 25 studies were collected and effect size estimates (*g*) were calculated for each neuropsychological variable reported by the individual studies. In addition, the mechanism of injury, diagnostic classification used for the injury severity, type of assessment tool, along with whether or not the data was separated by ‘symptomatic’ versus ‘non-symptomatic’ was coded. Commercially available standardized neuropsychological test measures were grouped based on procedural similarity. For example, there are many variations of list-learning memory tests that follow a very similar procedure of practiced learning, followed by a distraction, followed up by an immediate and delayed free and cued recall.

Pertab et al.’s (2009) results indicated that 18 of the 25 studies provided adequate data regarding the variables of interest in order to calculate an estimated effect size. The authors reported effect size estimates for neuropsychological test measures based on time since injury (acute and post-acute period). To this end, in the post-acute period, the effect size estimates ranged from $g = -0.52$[5] (verbal paired memory) to $g = 0.11$ (figure memory). The effect size estimates were larger for the acute-period, ranging from $g = -0.89$ to $g = -0.06$. The authors found that only four individual studies reported whether or not patients with mTBI were symptomatic or not but that only one study provided means and standard deviation data. As a result, the authors abandoned quantitative analysis of this factor and proceeded with a narrative review of these four studies. As for injury mechanism, the authors found that seven individual studies separated the studies based on injury mechanism but concluded that this was insufficient to explore and draw conclusions from, and thus abandoned this factor as well. Based on diagnostic criteria, the authors found that seven studies met the ACRM (Kay et al., 1993) criteria, whereas the others

[5] Here, the authors arbitrarily set the negative charge on the effect size estimate as indicative of poorer performance by the mTBI group.

were variable. Once again, the authors concluded that there was not enough data in order to proceed with quantitative analysis and abandoned this factor.

In summary, Pertab and colleagues (2009) concluded that effect size estimates of procedurally similar neuropsychological test measures yielded effect size estimates from small to large (Cohen, 1988), with the largest and statistically significant in the areas of verbal paired memory (g = -0.52), coding tasks (g = -0.33), and digit span (g = 0.31). As the authors were unable to quantitatively analyze factors such as mechanism of injury, symptomatic vs asymptomatic patients with mTBI, or diagnostic criteria, they cautioned clinicians from taking the results of prior meta-analysis without consideration of these important but omitted factors from the individual studies that the meta-analysis was based on. Moreover, the authors proposed that future studies in mTBI aim to collect and separate data based on these factors so that they can eventually be quantitatively analyzed in future meta-analyses. Limitations include failing to analyze factors due to insufficient sample sizes, factors which the authors argued were important and being largely ignored by previous meta-analysis. Moreover, since Pertab and colleagues (2009) expressed interest in maintaining homogeneity of the sample, it was unclear why they did not report the statistical correlate to verify their assumptions and report on the magnitude (i.e., Q, τ^2, or I^2; Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Other limitations include failing to take into consideration newer studies that may meet inclusion/exclusion criteria. Interestingly, this very same argument was stated by the authors when they stated the following:

> “the National Library of Medicine lists over 100 mTBI articles published since the Frencham et al. (2005) study that examine neuropsychological outcome in mTBI. By including past as well as more contemporary studies there should now be sufficient sample size to utilize meta-analytic techniques to better answer the residual neurobehavioral effects of having sustained a mTBI by controlling for and examining the variables identified in this critique” (Pertab et al., 2009, p. 505).

It is unclear, however, why they did not pursue their own suggestion.

Although this study was quite strict with respect to the inclusion criteria in an effort to maintain homogeneity of the sample, this came at a sacrifice of being unable to analyze data from three of the four proposed factors of interest. Accordingly, Pertab and colleagues’ (2009) study offered little novel knowledge but rather raised the important point that future basic studies should include additional factors in an effort to determine whether or not these can account for the remaining variability. Most importantly, the study employed a fixed-model approach with respect to the calculation of the estimated effect size. Since there are potentially many variables that may confound the homogeneity of the sample across all studies included mTBI meta-analyses (e.g., subject recruitment process, mechanism of injury, and specific measures used to name a few), this assumption would surely be violated. A random-effects model treats each effect size more as a sample from an overall population and recommended to be used in studies where there is evidence of heterogeneity (Lipsey & Wilson, 2001).

Indeed, Rohling and colleagues (2011) noted the limitations of the Pertab et al. (2009) study and attempted to reanalyze their data based on a random-effects model with additional analysis in to the effect-size estimates at different time-points to document differences in recovery. In an effort to provide a direct comparison to the findings from Pertab et al. (2009), the authors included the same studies, which was a combination of the studies from Binder et al., (1997) and Frencham and colleagues (2005). Effect sizes were calculated based on a random-effects model (Lipsey & Wilson, 2001), were sample-weighted with corrections for small sample sizes (i.e., n =< 20) as per Hedges (1985), and the pooled standard deviation was used. Moreover, effect size estimates were grouped by periods of < 7 days, 8-30 days, 31-92 days, and > 93 days based on time since injury. These effect size estimates were also grouped into the following cognitive domains: verbal memory, visual memory, working memory, executive functioning, processing speed, verbal comprehension and perceptual reasoning. The authors analyzed 25 studies, with 48 samples, totaling 2,828 patients with mTBI and 2,053 controls. Across all studies and time periods, the mean effect size estimate was $d = -0.28$, indicating that patients with mTBI performed worse than controls on the order of a quarter of a standard deviation. Across all time periods, only the results for the < 7 days and 8-30 days groups produced a significant ($p < 0.05$) estimated effect size of $d = -0.39$ and $d = -0.32$, respectively. Estimations for the 30-92 days ($d = -0.14$) and >93 days' ($d = -0.07$) time periods resulted in small but insignificant effect sizes. Across cognitive domains, all domains with the exception of executive functioning and perceptual reasoning produced small to moderate and statistically significant estimated mean effect sizes. All effect size estimates approach 0 over the course of recovery with the largest residual domain effect found in working memory ($d = -0.19$) at >93 days. The authors concluded that their re-analysis using a random-effects model resulted in similar findings to Binder et al.

(1997) and Frencham et al. (2005), in that the neuropsychological outcome of a mTBI after 90 days is little to none with an overall percent overlap of 79%, which would be clinically undetectable (Zakzanis, 2001). They also concluded that a small but highly impaired subgroup of patients with mTBI is unlikely to exist, citing congruence across prior meta-analyses (with the exception of Pertab et al., 2009) and the greater statistical likelihood that a meta-analysis would be able to find the true population differences than individual studies. Practically, the authors reported that their findings can be used to understand that verbal and visual memory would be preferentially impaired during the early recovery period with a recovery of all domains by the start of the third month. Limitations of this study are similar to those mentioned for prior meta-analyses. Some are inherent to the use of a meta-analysis (i.e., heterogeneity of the sample based on different samples, neuropsychological test measures used with resultant differences in sensitivities and specificities, differences in mechanism of injury, differences in time since injury, and diagnostic criterion used to define mTBI, etc.), whereas, other limitations include the lack of inclusion of newly published research. Here, the authors in 2011 re-analyzed the 25 studies that were used in the most recent meta-analysis in 2005 by Frencham and colleagues. As argued by Rohling et al. (2011), increasing the sample size (i.e., the number of studies included) increases the power and decreases the influence of a Type I or II error to skew the results.

In sum and broadly speaking, several meta-analyses have been conducted on the neuropsychological outcome following mTBI. Three are novel and original whereas, three are re-analyses with focus on different moderating factors using different statistical approaches. The majority, five of six, found support for the claim that little to no neuropsychological differences exist in and beyond the post-acute period of recovery following a mTBI. However, one study argued that a significant amount of variability was unaccounted for by previous approaches due to problematic methodology leading to heterogeneous samples (Pertab et al., 2009). Here, the authors related that this variability was evidence for the lasting negative effects of mTBI as measured by neuropsychological test measures. As discussed above, each meta-analysis contributed meaningful information to the knowledge base whilst having limitations. Arguably, the most ideal meta-analysis would combine all of the advantages that each above-reviewed meta-analysis contributed whilst minimizing for the limitations in an effort to accurately estimate the true population effect. Such characteristics would include minimizing heterogeneity of the sample (i.e., definition of mTBI, documented sustained injury, co-morbidities, etc.), controlling for practice effects for repeated measures studies, measuring across the clinically relevant recovery periods, using a random-effects model, and most importantly, including recent studies as the most recent meta-analyses has analyzed data from studies that were published over a decade ago.

To date, there have been multiple reviews citing support for the full recovery of neuropsychological effects following a mTBI in the post-acute period (Belanger et al., 2005; Binder et al., 1997; Schretlen & Shapiro, 2003), whereas other reviews have concluded that there may be specific domains that continue to show impairment in the post-acute period (Frencham et al., 2005; Pertab et al., 2009; Ruff & Jamora, 2009). Still, there are reviews citing that there may be a small population of patients that continue to have persistent cognitive impairment well into the post-acute period (termed *'miserable minority'* as per Rohling et al., 2012). Most consistently, deficits in attention (Chan, 2002; Potter et al., 2002; Solbakk et al., 1999; Vanderploeg et al., 2005) and speed of information processing (Bernstein, 2002; Johansson et al., 2009; Potter et al., 2002) have been reported. Though useful, these reviews are limited (as elaborated above). For example, many of them have serious flaws in heterogeneity by way of inclusion/exclusion criteria of the sample, neuropsychological test/domain classification, mechanism of injury (i.e., in the context of athletes incurring injury during play), combining adult and child mTBI, and or inclusion of samples that meet different definitions of mTBI. Moreover, these studies have limitations regarding the statistics employed. Although it may seem simple to gather effect sizes from each individual study and to average them, a comprehensive meta-analysis is more nuanced than that. To this end, sample weighted means, better estimates of the variance, heterogeneity of the sample of effect sizes and tests (i.e., fixed versus random effects model; whether an *a priori* assumption of or through post-hoc validation), along with useful moderating variables need to be thoughtfully utilized and analyzed in light of the interpretation of the findings. Moreover, despite the increasing number of publications on the topic of mTBI, the nature and cognitive course of the post-acute period following a mTBI remains an area of intense controversy – particularly in the medico-legal arena (Belanger et al., 2005; Rohling et al., 2012).

2.3 Present Study

Accordingly, the aim of this study was to provide a comprehensive and up to date meta-analysis on the neuropsychological profile of patients with mTBI in the post-acute stages of recovery using a random effects model. Specifically, we sought to calculate the combined (across all studies) and per cognitive domain mean effect sizes. Given that the majority of the meta-analyses reviewed found no neuropsychological differences between controls and patients with mTBI, we hypothesized that the overall effect size would also be insignificant in the post-acute period of mTBI. Regarding specific cognitive domains, the above review revealed subtle but significant effects on attention and psychomotor domains (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005). To this end, we hypothesized that attention and psychomotor speed domains would produce a statistical effect that would be significantly different than zero.

2.4 Methods

2.4.1 Meta-analysis

We employed standard meta-analytic techniques (Cooper & Hedges, 1994; Hedges & Olken, 1985; Rosenthal, 1991; 1995). As noted above, the advantage of a meta-analysis over a traditional narrative review is the ability to articulate the magnitude of the effect across primary studies. Magnitude is calculated using the effect size estimate, *d*, which reflects the degree to which the dependent variable is present in the sample group or the degree to which the null hypothesis is false (Cohen, 1988).

Mathematically, *d* is calculated similar to the *z* standardized variable. That is, the difference between the two means is divided by a standard deviation. Specifically, the mTBI group mean from a neuropsychological test was subtracted from the control group mean and divided by the pooled standard deviation of both samples. In the absence of a reported mean and standard deviation, effect size estimates can also be derived from different statistics such as correlation coefficients, binary data, and other commonly reported statistics (see Borenstein et al., 2010; Zakzanis, 1999). A negative value for *d* represents better performance by the control group and a positive value represents better performance by the mTBI group. A pooled standard deviation (Cohen, 1988) was utilized for this study as mathematically, it is the average standard deviation of the two groups. However, it should be noted that there are other approaches. To this end, the use of the control standard deviation has been argued to be a better theoretical correlation to the norms comparison that a clinician would utilize when interpreting performance from patients (Schretlen & Shapiro, 2003). However, as shown by Schretlen and Shapiro (2003), the difference between using a pooled standard deviation or the control standard deviation was minimal and statistically insignificant. Accordingly, and in an effort to survey the true population standard deviation without adding any statistical premonitions, the current study employed the pooled standard deviation when calculating effect sizes.

It has been shown that effect size calculations using smaller samples leads to an inflation of the effect (Frencham et al., 2005; Zakzanis, 1999). Hedges and Olkin (1985) suggested the utilization of a sample-size "bias correction factor" (p. 80), which they labeled as *g*. It was shown that the absolute value of the *g* statistic was slightly lower than the uncorrected *d* statistic (used in previous meta-analyses: Binder et al., 1997). Weighting effect size estimates based on sample size increases accuracy by correcting for an over- or under-estimate due to unequal sample sizes between groups. As such, sample sizes were corrected using the *g* effect size for the current study. Weighting was conducted in accordance to the methods described by Hedges and Olkin (1985), where the product of the experimental and control sample sizes is divided by the sum of the group sizes. The result is then divided by the sum of the ratios for all studies, which in turn "sum to unity" (Hedges & Olkin, 1985 p. 10). The summed sample-weighted effect sizes provide the weighted mean estimate.

A brief note should be mentioned regarding the controversy of utilizing a meta-analysis to derive clinical direction (Hunter & Schmidt, 1990). As noted by Iverson (2010), the product of a meta-analysis is only as good as the literature search that contributed to it and can obscure subgroups or individual effects within the population. To this end, the term "garbage in, garbage out" (p. 380) has been used to illustrate how poor selection criteria and without thoughtful consideration of the quality of the study based on the variables of interest may result in an overall estimate mean effect size that actually has less power than an individual study (Borenstein et al., 2010). One method to control the "waste management" (p. 380) is to set strict inclusion and exclusion criterion to uphold the homogeneity of the studied variables and to only include high quality studies (Borenstein et al., 2010). Moreover, one advantage of using a meta-analytic approach is the ability to statistically investigate the variation of the characteristics of the studies based on the effect size (Borenstein et al., 2010).

2.4.2 Search strategy

The search strategy and inclusion criteria in the current study was set at a strict level to preserve the homogeneity of the sample. A literature search using OVID Medline and PsychINFO databases was conducted on all mTBI based neuropsychological studies from 1970[6] to June 2017. The following search syntax was used (adult NOT child) AND (neuropsychology OR cognition) AND (mild OR minor) AND (head injury OR brain injury OR traumatic brain injury OR concussion). The de-duplicate feature was utilized in an effort to reduce duplicate results across databases. Furthermore, the following additional parameters were set to refine the search: studies that were in the English language, empirical studies and peer reviewed studies only. Reference sections of reviews were also combed through to find studies that were not populated in the literature search. The search returned 647 results which were individually and manually reviewed to determine if they met the study selection criteria: (1) no studies studying athletes[7]; (2) exclusion of studies that did not report any brain injury (i.e., whiplash); (3) mTBI had to be defined somehow – ideally aligned with the ACRM guidelines (Kay et al., 1993); (4) mTBI must be compared to a control group (either healthy or non-neurological); (5) if other severities were included, mTBI findings must be separated; (6) samples must be compared on neuropsychological test measures (clinically validated or experimental); (7) appropriate statistics must be reported to calculate an effect size; (8) no children as their progress following a mTBI has been reported to be different–adults and adolescents only (Borg et al., 2004; Capruso & Levin, 1992); (9) time since injury must be reported; and (10) exclusion of samples with positive neuroimaging findings. With respect to the combined control group, some studies used healthy participants as controls whereas others used 'non-neurological' or patients with orthopaedic injury as control, which is increasingly common (Frencham et al., 2005). It is argued, however, that cognitive functioning can be affected in non-neurological patients by way of other means (e.g., indirectly through pain). Hence, the mean of the overall control comparison would be an aggregate of the potential cognitive effects of non-neurological patients and healthy controls. However, any effect size differences between the mTBI and control groups would be helpful in

[6] Modeled after Belanger et al., (2005) in an effort to review any eligible studies.

[7] As per Belanger et al., (2005), studies that recruited sports-related patients with mTBI prospectively were excluded due to the potential differences in medical attention seeking.

assessing the specificity of mTBI. Moreover, unlike prior meta-analyses (i.e., Binder et al., 1997; Belanger et al., 2005) which included an mTBI sample with positive neuroimaging findings, we chose to restrict this as an exclusion criterion. The rationale for this was that the inclusion of mTBI samples with positive neuroimaging would be reflective of a mixed uncomplicated mTBI and complicated mTBI sample. Since the recovery outcome for patients who have sustained an uncomplicated mTBI have been shown to be different than those who have sustained a complicated mTBI (see Bigler, 2003; Pertab et al., 2009), the inclusion of a mixed sample would threaten homogeneity of the sample and thus, these studies were excluded from the analysis.

Following a three-level filtration process to determine eligibility, the final sample of studies were 31 with a total of 1469 patients with mTBI and 4281 controls. Patients recruited from a clinic-based or hospital-based setting were combined with patients recruited from a litigating-based sample that had passed all validity tests. This was justified based on the insignificant differences between mTBI litigation-based samples and mTBI clinic samples on estimated mean effect sizes on cognitive performance (Belanger et al., 2005).

There was considerable variability with respect to the presentation of the data across the 31 studies. Most studies reported raw means and standard deviations for the mTBI and control groups. Few studies reported means and standard error of the mean (SEM). Here, the SEM was converted into a *SD* manually to calculate the effect size (Zakzanis, 2001). Some studies, however, reported standardized scores (i.e., T and/or Z-scores), which were only included if the standardized score was consistently calculated using the same normative group, was calculated for the same measure (and not a combination of measures or domains), and was not an outlier as per the effect-size calculations. To this end, the study conducted by Clarke, Genat, and Anderson (2012) was excluded as they reported mean standardized scores from multiple measures. Maruta, Palacios, Zimmerman, Ghajar, & Mukherjee (2016a), Maruta et al., (2016b) and Losoi et al. (2016) reported z-score means, which was included as they reported the means per study and consistently used the same normative database as a comparison.

Unfortunately, some studies that appeared to be high quality were excluded as the mTBI data was combined with other severities or because an effect size calculation was not possible due to a lack of reported statistics. Some studies that initially met the inclusion criteria were removed upon closer inspection. To this end, the study by Clarke and colleagues (2012) was of high quality but could not be included as they used a mixed sample which included patients with TBI with comorbid psychiatric and substance abuse disorders. Moreover, raw scores were not reported but rather averages of z-scores per domains were given, reducing the degrees of freedom of the estimate. In addition, the authors grouped the tests as per cognitive domain in a manner that was not consistent with the groupings of the cognitive domains of the current study (e.g., including the COWAT in the executive functions domain rather than the fluency domain). For these reasons, this study was excluded for the analysis.

Some authors used the same samples across publications. In an effort to reduce overlapping data entry, each of these cases were reviewed and more recent and relevant data were extracted for the meta-analysis. Dikmen, Machamer, and Temkin (2001) used the same control sample as Dikmen, Machamer, and Temkin (2017), but each study included measures that were not found in the other. Accordingly, all unique comparisons were extracted with preference for the more recent study. Similarly, Maruta and colleagues (2016a) used the same participants as Maruta et al. (2016b). Data from the most updated study (Maruta et al., 2016b) was used to calculate effect sizes except for the CVLT-II results which were only available on the initial publication (Maruta et al., 2016a). DeMonte, Geffen, May, & MacFarland (2004) separated the data for their control and patient samples by gender and did not provide combined data. Accordingly, a weighted average of the means and standard deviations were calculated and included within the data analysis. Likewise, Sterr, Herron, Hayward, and Montaldi (2006) reported data from patients with mTBI separately as those with and without post-concussion symptoms. Accordingly, the groups were combined and weighted based on their respective sample sizes. Effect sizes were calculated from z-score means and standard deviations from Losoi et al. (2016) as raw scores were not reported. Only the “mean reaction time all” (MacFlynn et al., 1984, p. 1328) from the experimental design from MacFlynn and colleagues was extracted to calculate the effect size.

A between groups design was employed with group association (mTBI or control) as the independent variable and performance on neuropsychological test measures as the dependent variables, which were analyzed as continuous variables. As per the work of Hedges and Vevea (1998), fixed and random effects models may be employed to approach a meta-analysis. There are assumptions and limitations to the use of each particular method, however. A fixed-effects model assumes that the effect sizes were derived from the same sample/population with random error being the only contributing error. In contrast, a random-effects model treats each calculated effect size as a sample from an overall population and is recommended to be used in studies where there is evidence of heterogeneity (Hedges & Vevea, 1998). Previous meta-analyses in the area of mTBI outcome have used fixed effects models (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005; Pertab et al., 2009; Schretlen & Shapiro, 2003) with the exception of one, which used a random-effects model (Rohling et al., 2011). Since there are potentially many variables that may confound the heterogeneity of the sample across all studies including this meta-analysis (e.g., subject recruitment process, mechanism of injury, and specific measures used in the studies), this assumption of the fixed-effects model would surely be violated in the current study. Accordingly, a random-effects model was used in this current study.

The *Q* statistic was also calculated to examine homogeneity of effect sizes across the studies (Hunter & Schmidt, 1990). This statistic, which is calculated by dividing the variance of the sample-weighted *d* by the sampling error variance multiplied by the number of data samples, assesses whether there is true heterogeneity in the meta-analysis. Failing to meet significance would be indicative of a single population effect with sampling error being the only difference between the studies (i.e., fixed-effects model). However, the *Q* statistic is limited to only describing whether or not heterogeneity exists within the samples, it cannot determine the extent of the true heterogeneity (Huedo-Medina et al., 2006). Assuming random-effects model to be the case, another statistic, τ^2 (tau-squared), reflects how much the true population effect size estimate across the studies of a meta-analysis differ by (Heudo-Medina et al., 2006) but is unable to provide practical estimates of the magnitude of variance. As such, the I^2 proposed by Higgins, Thompson, Deeks, and Altman (2003), measures true heterogeneity by dividing the difference between *Q* with the degrees of freedom (k -1) by the *Q* statistic multiplied by 100. Here, the I^2 index can directly be interpreted as the percentage of the total variability in a set of effect sizes due to true heterogeneity (between-studies; Higgins et al., 2003; Huedo-Medina et al., 2006). This index ranges from 0 to 100 (i.e., 0% or 100%) with 0 indicative of no heterogeneity between studies (i.e., only sampling error) and 100 indicative of complete heterogeneity between studies. It has been proposed that a value of 25, 50, and 75 would be indicative of low, medium, and high heterogeneity, respectively (Higgins and Thompson, 2002). Accordingly, the I^2 was also calculated in an effort to determine the overall practical true population heterogeneity.

In an effort to determine the neuropsychological profile of a patient who sustained a mTBI after the typical 3-month recovery period, effect sizes were calculated and classified based on cognitive domains. The inclusion of a particular measure into a cognitive domain was based on a combination and congruence from factor analysis research (as per Binder et al., 1997; Dikmen et al., 1995; Frencham et al., 2005) and based on neuropsychological reference literature (e.g., Lezak et al., 2012; Strauss, Sherman, & Spreen, 2006). This was done because of the limitations of using factor analysis to define tests into "pure" cognitive domains and face validity bias (Binder et al., 1997; Dikmen et al., 1995; Frencham et al., 2005). Specifically, eight cognitive domains were used in this meta-analysis: Global Cognitive Ability, Attention & Psychomotor Speed, Executive Functions, Fluency, Memory Acquisition, Delayed Memory, Language, and Visuospatial Ability. Examples of the tests that were categorized into these domains can be found on Table 1. Verbal fluency has been reported as a test of language and also executive function and used by neuropsychologist for both purposes in clinical practice (see Rabin et al., 2016). Whiteside et al., (2016) provided and excellent literature review citing evidence of verbal fluency performance being related to structure associated with executive functions and language functions through focal lesion and imaging studies. These authors used factor analysis to explore the underlying cognitive structure of verbal fluency and it's relation to either language or executive functions domains. The findings from their study suggested that the FAS and Animals tests loaded onto language factors exclusively and not onto executive functions. However, they cautioned that replication studies with larger and broader samples would be required to confirm their results. To this end, verbal fluency tests appear to be in-between the domains of language and executive functions. The current study treated verbal fluency as a separate domain for these reasons. This also allowed to maintain groupings from previous meta-analyses for replications purposes and to provide a comparative discussion.

Experimental tasks that did not have any literature justifying their factor loadings were grouped into cognitive domains based on the original authors classification. Across all domains, negative effect size values indicated worse performance of the mTBI group as compared to the control group. For tests where a smaller value was indicative of better performance (i.e., reaction time, completion time), the polarity of the calculated effect size was reversed so that positive effect size values would consistently represent better performance and negative effect size values represent worse performance. Due to differences in sample sizes across studies, effect sizes were sample weighted as per Hedges and Olkin (1985).

Table 1.

Examples of the Tests that Represented Within Each Cognitive Domain.

Cognitive Domain	Examples of Tests
Global Cognitive Ability	Wechsler Scales of Intelligence NAART
Attention & Psychomotor Speed	Trail Making Test - Trial A Digit Span Symbol Digit Modalities Test PASAT Finger Tapping Test Grooved Pegboard Continuous Performance Test
Executive Functions	Wisconsin Card Sorting Test Trail Making Test - Trial B Stroop Matrix Reasoning Test Category Test
Fluency	COWAT Animal Fluency
Memory Acquisition	California Verbal Learning Test - Immediate Recall Logical Memory - Immediate Recall Rey-Osterrieth Complex Figure Task - Immediate Recall
Delayed Memory	California Verbal Learning Test - Delayed Recall Logical Memory Immediate - Delayed Recall Rey-Osterrieth Complex Figure Task - Delayed Recall Brief Visual Memory Test - Delayed Recall
Language	Boston Naming Test Vocabulary Similarities Token Test
Visuospatial Abilities	Block Design Judgment of Line Orientation Picture Completion

Rey-Osterrieth Complex Figure Task - Copy Trial

Note: Procedural variations of the same test were grouped accordingly.

Abbreviations: COWAT = Controlled Oral Word Associated Test; NAART = North American Reading Test; PASAT = Paced Auditory Serial Addition Test.

2.4.3 Statistical Analysis

All statistical analyses related to the effect size calculations, sample weighting, Q and τ^2 heterogeneity statistics, and Fail-Safe N calculations were computed using Comprehensive Meta-Analysis (Version 3.3.070). Additional heterogeneity statistics (I^2) were computed using Microsoft Excel Spreadsheet (Microsoft Office 365 Home and Student, 2016) using the formulae described by Higgins and colleagues (2003). One-sample *t*-tests to compare effect size estimates against zero were conducted using the Statistical Program for the Social Sciences (SPSS) Version 22.0 for Windows (IBM Corp, 2013).

2.5 Results

2.5.1 Sample characteristics

Overall, 316 effect sizes were calculated from extracted data across 31 studies (see Table 2). The mean age of patients with mTBI was 33.5 years (SD = 10.03; 62.4% male) and the mean age for controls was 34.1 years (SD = 10.32; 59.3% male), which was not statistically different ($t(54) = 0.50$, $p = 0.62$). Education of the samples was not reported across all studies and when it was, it was not consistently reported using the same metric. From the data available, however, the average education of the mTBI sample was 13.1 years (SD = 2.11) and for controls was 14.2 years (SD = 2.09), which was not statistically different ($t(50) = 1.77$, $p = 0.08$). Time since injury was also inconsistently reported with some researchers reporting the mean and standard deviation of the sample while others reported the minimum time since injury as part of their recruitment criteria. Here, time since injury ranged from three-months to 16 years. To this end, 14 of the 31 studies reported a time since injury that was 12 months or greater, whereas the majority of the sample reported a time since injury of less than 12 months.

Table 2.

Details of Neuropsychological Studies of Mild Traumatic Brain Injury Included in Meta-Analysis.

First author	Year	Cognitive ability domain(s) examined	n MTBI	n Controls	Total n	Time since injury (min OR Mean and SD)	No of effect sizes extracted
Barwood	2013	G, FL, L	16	16	32	6 months	6
Bohnen	1992	A	11	11	22	22.9 months	2
Cicerone	1996	A	15	9	24	6 months	3
Demery	2010	G, A, EX	20	24	44	62.9 months (11.4)	10
Dikmen	2001	DM	63	109	172	12 months	1
Dikmen	2017	G, A, EX, DM	114	109	223	12 months	9
Gosselin	2012	A	44	40	84	7.6 months (8.4)	2
Harman-Smith	2013	G, A, EX, L, V	83	94	177	150.7 days (27.8)	14
Konrad	2011	A, EX, FL, AQ, DM	33	33	66	4.75 years	21
Kraus	2007	G, A, EX, AQ, DM	20	18	38	107 months	20
Losoi	2016	A, EX, FL, AQ, DM	65	40	105	6 months	13
MacFlynn	1984	A	28	40	68	6 months	1
Mangels	2002	G, A, EX, FL, AQ, DM, L	11	10	21	1.5 years	19
Maruta	2016a	G, DM	31	32	63	4 months	4
Maruta	2016b	A	33	140	173	4 months	9
Mathias	2007	G	21	21	42	3 months	1
Meyer	2004	G, A, EX, FL, AQ, DM, L, V	57	30	87	7.59 months (10.99)	41
Michael	2015	A, EX, FL, AQ, DM	37	58	95	5 months	6
Ord	2010	EX	67	20	87	12 months	7
Ponsford	2011	A, DM	90	80	170	3 months	4
Rabinowitz	2015	A, AQ, DM	66	38	104	3 months	5
Raskin	1997	G, A, EX, AQ, DM, V	10	10	20	24.89 months (10.0)	30
Raskin	1996	F	17	22	39	38.87 months (22.23)	2

Ruffolo	2000	A, EX	62	49	111	21.2 months (22.9)	2
Shandrea-Ochsner	2013	G, A, EX, FL, AQ, DM	20	21	41	44.62 months (30.26)	15
Sterr	2006	G	38	38	76	12 months	1
Storzbach	2015	A, EX, AQ, DM, V	49	40	89	52.64 months (31.72)	19
Swan	2015	G, A, EX, FL, AQ, DM	31	33	64	3 months	20
Tiersky	1998	A, EX, AQ, DM, V	33	20	53	3 months	13
Vanderploeg	2005	G, A, EX, FL, AQ, DM, V	254	3057	3311	16 years	14
Veeramuthu	2016	A, EX, DM, L, V	30	19	49	6 months	5

Abbreviations: G = Global Cognitive Ability, A = Attention & Psychomotor Speed, EX = Executive Functions, FL = Fluency, AQ = Acquisition Memory, DM = Delayed Memory, L = Language, V = Visuospatial Ability

The majority of the control groups used across all studies were healthy controls (23/31 studies) and were matched by age and education. Eight studies used alternative controls groups; three studies utilized trauma controls, two studies used non-neurological (undisclosed) controls, two used orthopaedic controls, and one study used combat controls. The utilization of trauma controls has been argued to be ideal as it closely approximates the experience of a mTBI patient whilst controlling for brain injury (Rohling et al., 2012), however, studies that used healthy controls were included in the current analysis in an effort in increase samples sizes to perform inferential statistics with reasonable power.

As for diagnostic criteria, the majority (19/31) of the studies included in the current analysis defined mTBI using the ACRM criteria (Kay et al., 1993) or other guidelines which were identical such as the Department of Defense (Cifu et al., 2009) mTBI definition and the WHO mTBI definition (Holm et al., 2005). There were some studies that did not use the full ACRM guidelines as part of their recruitment criteria. Three studies included mTBI participants if they had a documented GCS score of 13-15 with negative neuroimaging findings but did not comment on the PTA or LOC duration. The loosest definition of mTBI was that of Vanderploeg and colleagues (2005) which defined mTBI as “head injury with altered consciousness” (p. 230).

In addition, Maruta et al. (2006a; 2006b) defined their mTBI sample as "persistent problems believed to result from an isolated concussive head injury that occurred between 90 days and 5 years prior to the date of neurocognitive testing"; "documented medical attention at the time of injury"; "PTA at the time of injury"; "a complete BISQ"; "and if LOC occurred, it did not exceed 24h in the period following the injury" (Maruta et al., 2016a, p. 2). No study in the current analysis included patients with mTBI with positive neuroimaging findings. In addition, all studies in the current analysis had an mTBI group which had GCS scores of 13 or higher.

One study, Vanderploeg et al., (2005)., contributed approximately 71% of the total sample in the mTBI group and 17% of the total sample in the control group. In an effort to minimize the potential skewed effects of this study, all analysis (overall and across cognitive domains) were conducted twice with (g) and without (g_{adj}) the inclusion of Vanderploeg and colleagues' (2005) study and compared statistically using a one-sample t-test.

Estimated mean and sample-weighted effect sizes comparing mTBI and control groups were calculated for each study. An overall mean estimated effect size was calculated in addition to estimates for each cognitive domain in an effort to illustrate the cognitive profile of mTBI in the post-acute period of recovery. All effect size estimates are displayed in Table 3.

Table 3.

Weighted Effect Sizes (g) for Overall Estimates and for Each Neuropsychological Domain and Heterogeneity Statistics.

	n mTBI	*n* Control	*k*	*g* (*SD*)	*t*	*Q*	τ^2	I^2	95% CI
Overall	1469	4281	316	-0.35 (0.156)	12.46***	1368.1***	0.125	61.26%	42.77% - 73.77%
Domain									
Global Cognitive Ability	992.0	3683.5	23	-0.42 (0.351)	2.64*	182.8***	0.281	87.86%	83.09% - 91.28%
Attention & Psychomotor Speed	1211.6	4026.5	116	-0.30 (0.207)	6.62***	487.2***	0.148	73.67%	64.96% - 80.21%
Executive Functions	535.5	3319.0	47	-0.23 (0.230)	4.20***	129.0***	0.066	63.46%	50.09% - 73.25%
Fluency	797.0	3517.0	19	-0.61 (0.336)	6.77***	110.8***	0.170	83.70%	75.74% - 89.05%
Acquisition Memory	897.7	3647.0	36	-0.42 (0.285)	5.15***	147.3***	0.162	77.28%	68.89% - 83.40%
Delayed Memory	897.7	3647.0	48	-0.32 (0.239)	5.11***	173.1***	0.093	73.85%	65.33% - 80.28%
Language	193.3	168.7	12	-0.73 (0.396)	4.31**	44.4***	0.204	76.75%	59.47% - 86.66%
Visuospatial Ability	512.5	3270.0	13	-0.22 (0.293)	2.20*	43.4***	0.056	73.71%	54.43% - 84.84%

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimates of samples sizes are noted in decimals due to averaging of unequal sample sizes of comparisons.

2.5.2 Overall effect size

The overall estimated mean effect sizes was small to moderate (g = -0.35, SD = 0.156) in keeping with Cohen's (1988) heuristic terminology, but significantly different from zero (t(315), $p < 0.001$). The exclusion of the data from Vanderploeg et al. (2005) did not statistically affect the overall estimate mean effect size (g_{adj} = -0.37, SD = 0.170). The percent overlap comparison for the overall effect ranged from 72.6% to 78.7% as per Zakzanis (2001).

Heterogeneity statistics were calculated on the overall analysis which revealed significant heterogeneity (Q = 1368.1, $p < 0.0001$; τ^2 = 0.125; I^2 = 61.26%, CI (95%) = 42.77% - 73.77%). In other words, a significant Q statistic confirmed beyond sampling error heterogeneity and the correct use of a random-effects model. Moreover, the I^2 statistic commented on the magnitude of the heterogeneity as 61.26%, with a 95% confidence interval in the range of 42.77% - 73.77%. The magnitude of the heterogeneity would be considered large as per Higgins and Thompson's (2002) proposal, suggesting the potential that additional variables may moderate the relationship. Moderating variables such as those proposed by Pertab et al., (2009) were initially explored, however, few studies reported these variables. Moreover, inconsistency of the method of reporting adding to the complication. As with Pertab and colleagues (2009), the collection and analysis of potential moderating variables was abandoned in the current study due to the high risk of misrepresentation and misinterpretation.

Null results are typically not reported in the literature and as such, the studies used to calculate effect sizes in a meta-analysis are prone to publication bias of 'positive' results (Rosenthal, 1979). In an effort to determine the robustness of the findings, a Fail-Safe N analysis was conducted to provide a measure of how many unpublished studies would theoretically be needed to usurp the current estimated overall mean effect size findings to an insignificantly low effect size (d = 0.1). Using Orwin's (1983) Fail Safe N formula, approximately 790 unpublished studies with an overall average effect size of 0.1 would be needed to eliminate the current findings.

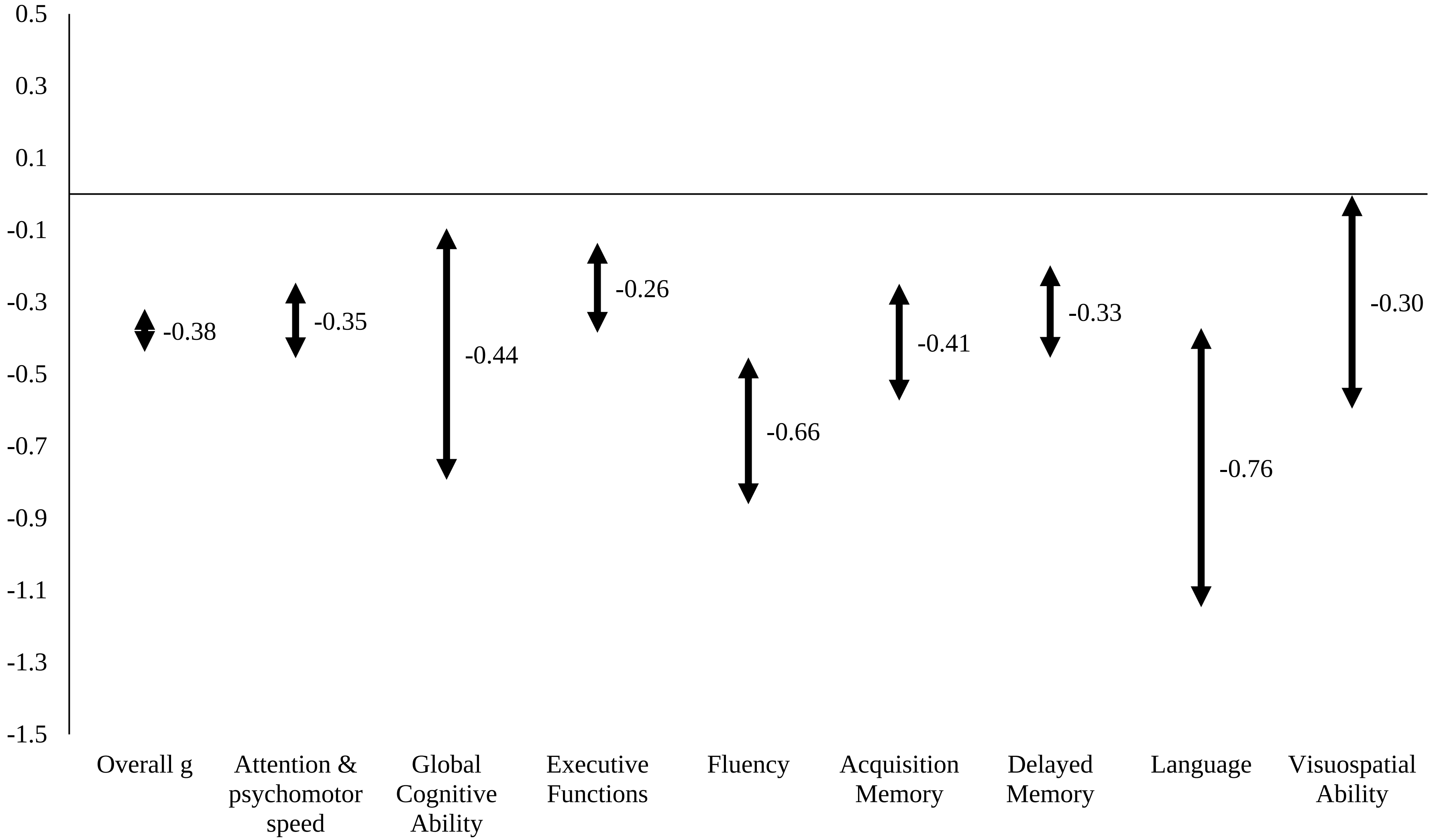
2.5.3 Overall effect size by cognitive domain

In terms of cognitive domain, the mean effect size estimates ranged from small to moderate (Cohen, 1988) and are presented in Table 3 and graphically in Figure 1. Specifically, the effect sizes across domains were as follows: Global Cognitive Ability (g = -0.42, SD = 0.351; g_{adj} = -0.44, SD = 0.379); Attention & Psychomotor Speed (g = -0.30, SD = 0.207; g_{adj} = -0.31, SD = 0.217); Executive Functions (g = -0.23, SD = 0.230; g_{adj} = -0.26, SD = 0.257); Fluency (g = -0.61, SD = 0.336; g_{adj} = -0.70, SD = 0.300); Acquisition memory (g = -0.42, SD = 0.285; g_{adj} = -0.44, SD = 0.285); Delayed Memory (g = -0.32, SD = 0.239; g_{adj} = -0.35, SD = 0.253); Language[8] (g = -0.73, SD = 0.396); and Visuospatial Ability (g = -0.22, SD = 0.293; g_{adj} = -0.31, SD = 0.339). Excluding the data from Vanderploeg et al. (2005) from the effect size calculations for each cognitive domain did not produce statistically significant effect size differences. Thus, although the study contributed a significant proportion of the sample to the mTBI group, it did not influence the estimated mean effect sizes in a meaningful way.

The largest effects were found in the domains of Language and Fluency. Measures categorized into the Executive Functions domain were calculated to have the smallest effect size. Estimates of the mean effect size across all eight domains were statistically different from zero (p < 0.05). Heterogeneity statistics confirmed the use of a random effects model (Q = 1,368.1, p < 0.001; τ^2 = 0.125) with I^2 values (heterogeneity magnitude) ranging from 63.46% (Executive Functions) to 87.86% (Global Cognitive Ability), once again suggesting the potential for moderating variables to account for the variability.

[8] Vanderploeg et al. (2005) did not contribute any effect size estimates to the domain of language

Figure 1. Mean *g* Difference and Confidence Interval of Neuropsychological Domain.



2.6 Discussion

The aim of the current study was to provide a comprehensive and up to date meta-analysis of the neuropsychological profile following mTBI in the post-acute stages after injury using a random-effects model. In contrast to our hypothesis which was that the overall estimated mean effect size would be insignificant, the overall estimated mean effect size across all studies was small to moderate and was found to be statistically different than zero. Moreover, and again in contrast to our hypothesis, all effect size estimates per cognitive domain were statistically different than zero and were considered to be small to medium as per Cohen's heuristics (1988). The largest effects were found in the domains of Language (g = -0.73, *SD* = 0.396) and Fluency (g = -0.61, *SD* = 0.336). Belanger and colleagues (2005) also found Fluency to have the largest effect size estimate across domains. The smallest effect size estimates in the current study were that of Visuospatial Abilities (g = -0.22, *SD* = 0.293) and Executive Functions (g = -0.23, *SD* = 0.230). Regarding Visuospatial Ability, are our estimate is relatively similar to that of Rohling et al. (g = -0.16; 2011) and Frencham et al. (g = -0.25; 2005). In contrast, Belanger et al., (2005) found an estimate g = -0.57 for 'visuospatial skills'. This difference may be due to methodological differences used to categorize specific neuropsychological test measures into domains.

With respect to Executive Functions, this estimate appears to be robust and consistent across most previous studies. The results of the current study found the mean effect size estimate to be g = -0.23 (*SD* = 0.230). This estimate is comparable to the estimates reported by Belanger et al. (2005) and Rohling et al., (2011) at g = -0.21. Moreover, this estimate falls within the range of what Frencham et al., (2005) found at g = -0.30. Taken together, these findings are partially in contrast to estimates reported by Zakzanis et al. (1999), who reported large effect size estimates in tests of Executive Functions. However, the results of Zakzanis et al. (1999) cannot be directly compared as the estimates were not separated by cognitive domain, but rather were calculated per neuropsychological test.

There is a dearth of studies that have explored language-based deficits in the post-acute phase of recovery following a mTBI (Belanger et al., 2005). This is also reflected in the limited number of studies that reported performance on tests of language in the current meta-analysis. The few studies that have published on this topic report evidence of lasting language deficits following a mTBI, with specific impaired performance on test of naming and definitions (Barwood &

Murdoch, 2013; Whelon & Murdoch, 2013). However, the reliability of these studies is limited by low power as a function of small sample sizes, which ranged from 5 to 16. Mangels and colleagues (2002) found that patients who were in the post-acute phase of recovery following mTBI generated significantly fewer words on phonemic fluency trials of the COWAT as compared to controls. The authors suggested that this was related to a deficit in executive control. Clearly, more research is necessary to comment on the robustness of the current findings in relation to the extant literature.

Five out of the previous six meta-analyses regarding mTBI have concluded that there was no neuropsychological difference between controls and patients with mTBI beyond the post-acute recovery period. The findings of the current study suggest that this is an overly simplistic view of the actual data. To this end, the results may be interpreted quite differently depending on the perspective of the reader. Statistically, the current results indicate that the mTBI group is a different group than healthy controls across all cognitive domains studied. From this perspective, one would be inclined to agree with the findings from Pertab et al., (2009) in that there are subtle but significant differences in cognitive performance from patients with mTBI in the post-acute period of recovery. Clinically, however, a marker should be able to discriminate nearly all the patients from healthy controls on the variable of interest (Zakzanis et al., 1999). For example, an effect size of 3.0 has been proposed as a clinically relevant heuristic that would have an overlap of 5% between the patient and healthy control group (Zakzanis et al., 1999). In the current study, the largest difference was seen in the domain of Language with an overlap of approximately 52.6% - 57%. In other words, 43% - 47.4% of patients with mTBI obtained language scores that were not found among healthy controls. Additionally, the probability of superiority metric indicates that there is an approximate 69% chance that an individual chosen at random from the patient group will have a better score than a random individual from the control group. From a practical and clinical perspective, performance on neuropsychological test measures are not sensitive to the persistent cognitive complaints expressed by this population due to the high overlap of scores.

In addition, the results of the current study indicated significant unaccounted variability in the sample. The Q-statistic along with the τ^2 value was statistically significant indicating the presence of more than just sampling error in the sample. The magnitude of the heterogeneity was measured by the I^2 statistic which was approximately 61.3% (95% CI 42.77% - 73.77%). Across cognitive domains, these estimates ranged from 63.46% (95% CI 50.09% - 73.25%; Executive Functions) to 87.86% (95% CI 83.09% - 91.28%; Global Cognitive Ability). Clearly, this indicates the presence of moderating variables that may contribute to the variability that has not been taken into account by the current model. In the early stages of the current study, it was planned to extract data relating to mood, pain, and fatigue in an effort to determine the extent that these variables would moderate the findings. These variables were chosen due to their well-known effects on neuropsychological test performance (see Zakzanis et al., 1999 for a review). Unfortunately, analyzing these moderating variables was abandoned due to inconsistent reporting of these variables across the studies. Previous meta-analyses have produced moderating variable analysis (e.g., Belanger et al., 2005; Frencham et al., 2005). However, these authors noted inconsistent reporting across studies and as a result simply coded the variables in a binary fashion; the presence or absence of the variable of interest (i.e., acute or post-acute period). This was not done in the current study as the risks of dichotomizing continuous variables are too large (see Altman & Royston, 2006). The advantage is that dichotomizing would greatly simplify the analysis and would lead to useful interpretation of the results and variability therein. This is the assumption at least. However, dichotomizing continuous variables would lead to a loss of information, reduced statistical power, underestimates the variability of a binary variable, and limits any interpretation to linear relationships only. Most notably, it would lead to a false assumption that the variable in question may actually be dichotomous in the population when there is no good reason to assume so (Altman & Royston, 2006).

Another factor that may account for the large variability that was unaccounted for in the current model may be test insensitivity to the subtle cognitive effects of mTBI. There have been many studies that have criticized the insufficient sensitivity of neuropsychological test measures, as opposed to other measures, to detect the purported subtle cognitive impairment following mTBI (see McInnis, Friesen, MacKenzie, Westwood, & Boe, 2017). Indeed, this sensitivity and specificity is calculated and reported in the manuals of most neuropsychological test measures. For example, the Symbol Digit Modalities Test (SDMT; Smith, 1993) reports that approximately 36% of the score obtained on the SDMT can be attributable to random error. Although neuropsychological test measures are “currently the best that science has to offer” (Barr, 2001, p. 299) with respect to evaluating the breadth and severity of cognitive impairment, they are not error-free and do not measure complex constructs such as ‘cognitive functioning’ directly and purely. Thus, it may be that our tests are not sensitive to the subtle cognitive impairments that occur at different periods following a mTBI. This is more a systemic limitation of the field of neuropsychology than one that is specific to this meta-analysis. Newer tests are being developed with the goal of improving sensitivity and specificity whilst reflecting real-world impairment (see Burgess et al., 2006; Marcotte & Grant, 2010).

Other limitations of the current study are also inherent to the methodology of a meta-analysis along with the standards of reporting in the field. Firstly, we were only able to code for variables that were reported by the authors of the studies reviewed. It is likely that some studies that were accepted in the final analysis failed to mention a key characteristic that would ultimately exclude them from the current analysis (e.g., mixed sample of sports and non-sports related injuries). The effect sizes calculated may be artificially influenced as a result. Although sample heterogeneity was of particular importance in the current study, the combined sample is far from a pure and homogenous sample due to individual recruitment procedures, differing mechanisms of injury, differences in the definition of mTBI, and other sampling characteristics. In addition, there is no standardized test battery for patients with mTBI. Thus, clinicians and researchers use tests with differing sensitivity and specificity values. For the purposes of this study, tests that were less sensitive and/or specific were grouped in the same cognitive domain as tests that were more. Ideally, the effect size estimates that result should be weighted by their reported sensitivity and specificity values. However, due to inconsistent reporting and estimates of these values, this was not undertaken in the current study. Moreover, such studies were included in an effort to increase power through increasing sample size there by decreasing the influence of a Type I or II error over the results. Future studies are encouraged to find creative statistical approaches that can calculate effect size estimates that are both sample-weighted and weighted based on sensitivity and specificity values of the test measures.

The lack of analysis of moderating variables limits our ability to explain, with reasonable correlational certainty, which potential factors influenced differences between the two groups. Collection of moderating variables was initially planned for in the study but was abandoned as there was considerable variability of reporting methods. Moreover, the lack of standardization in the field for reporting moderating variables has been a long-standing issue, with many researchers promoting standards to but few have adopted them (for a review, see Maas et al., 2017). In addition, the largest estimated cognitive mean effect size was calculated from the domain with the least number of comparisons contributing to it (Language with 12 comparison).

This leads to an unequal and lack of sample size of effect sizes across domains which skews the overall population effect size estimate. Given how large the effect size was estimated in the current study, further basic research is necessary to clarify the stability of this finding.

Lastly but quite importantly, the majority of the studies included came from cross sectional studies with time since injury ranging from 90 days to 16 years. Prior studies have indicated a negative logarithmic trend regarding cognitive impairment and time since injury (Schretlen & Shapiro, 2003). It may be that the studies with samples that are far beyond the 90 days threshold may contribute much smaller effect sizes overall. This was not accounted for or analyzed in the current study as although all studies reported time since injury, they were inconsistent with the way they reported it. Some reported a mean and standard deviation for all their patients, others reported minimum time since injury as inclusion criteria for their study, and still others reported individual time since injury estimates for their entire sample.

2.6.1 Conclusion & Limitations

In conclusion, the current results indicate that there is an overall effect between patients with mTBI and controls across neuropsychological test performance in the post-acute period of recovery, but that this effect is small and the overlap between the two distributions is too large to be clinically useful at this time. Ideally, future basic studies should report consistent statistics and moderating data (Maas et al., 2017) with more emphasis on longitudinal rather than cross-sectional research.

Chapter 3

3 Study 2: Ecological Validity of Traditional Pen-and-paper Neuropsychological Test Measures in mTBI

3.1 Disability

The WHO estimates that "traumatic brain injury is the leading cause of disability in people under 40 years of age" (WHO, 2006, p. 167). It is important to understand the outcomes of patients who sustain a TBI in an effort to better inform policy makers to direct funding to efficient and evidence-based rehabilitation programs which aim to address real-world needs. In the literature, disability has been defined and measured in many ways. For example, disability following a TBI may be defined as an inability to return to productivity, school, work, play, service, along with other premorbid activities and may be measured by self-report, direct assessment, observation, or by way of a questionnaire (for a review, see Saltychev, Eskola, Tenovuo, & Laimi, 2013). Moreover, a formal definition of disability has been defined by the International Classification of Functioning, Disability and Health, which was endorsed by the WHO to be used as a standard definition in TBI research, but few studies have utilized this guideline (Johansson & Bernsprang, 2001).

Following a TBI, physical, cognitive, affective, and behavioural symptoms manifest. The breadth and severity of these symptoms vary from person to person. These symptoms interact in a complex manner and can have impacts on multiples aspects of the patient's life. One of these aspects are their ability to RTW. Employment is considered a source of personal satisfaction, social recognition, and overall life satisfaction (O'Neill et al., 1998). RTW following brain injury has been related to a better sense of well-being, better overall health, increased social and community integration, less utilization of health care services, and better overall quality of life (Corrigan, Bogner, Mysiw, Clinchot, & Fugate, 1997; Steadman-Pare, Colantonio, Ratcliff, Chase, & Vernich, 2001; Wehman, Targett, West, & Kregel, 2005). Thus, a commonly used outcome metric for real-world functioning is RTW status (Ownsworth & McKenna, 2004). Moreover, it is not uncommon for the clinician to be asked by a referring party to estimate when a patient who has sustained a TBI can return to gainful employment. Accordingly, a brief review of RTW following TBI will be discussed.

3.2 Return to Work and TBI

Unfortunately, the results of RTW status following TBI are varied and at times inconsistent. Based on some estimates, 38-54% of patients with a TBI-related injury of any severity return to some sort of gainful employment within one-year after their injury (Corrigan et al., 1997; Doctor et al., 2005; Jourdan et al., 2013; Ketchum et al., 2012); 28% RTW after two-years (Ponsford et al., 2014); 54% RTW after three years (Grauwmeijer, Heijenbrok-Kal, Haitsma, & Ribbers, 2012); 49-66% RTW after five years (Airey, Chell, Ribgy, Tennant, & Connelly, 2001; Mackenzie et al., 2006; Overgaard, Hoyer, & Christensen, 2011; Redmill, McIlwee, McNicholl, & Templeton, 2006; Soberg, Roise, Bautz-Holter, & Finset, 2011); and 92% after ten years (Dahm & Ponsford, 2015). The rich dataset analyzed by the CDC indicate that 56.7% of patients RTW within one year following a TBI-related injury, whereas the 43.3% remaining continued to be disabled (Selassie et al., 2008). To this end, disability was defined broadly and included severe difficulty or inability to perform their activities of daily living, developing post-injury symptoms that prevented the individual from pursuing the activities they wished to, and poor performance on cognitive and mental health measures.

Thurman, Alverson, Dunn, Guerrero, & Sniezek (1999) estimated the prevalence of long-term disability related to TBI was approximately 2% (5.3 million) of the American population. This has since been estimated to decrease to approximately 1% (3.2 million) by Zaloshnja, Miller, Langlois, and Selassie (2008). However, the differences have been argued to be more attributable to methodological differences in inclusion criteria and data availability than a natural decrease in the actual prevalence rates (Corrigan et al., 2010). Despite this apparent decrease it continues to likely be an underestimate of the actual prevalence rate as they do not consider TBIs that were treated in other healthcare settings such as military hospitals or those that did not seek medical attention (Corrigan et al., 2010; Faul et al., 2010; Langlois et al., 2006; Taylor et al., 2017).

Multiple studies have found unstable employment rates following TBI (Arango-Lasprilla, et al., 2009; Odgaard, Johnsen, Pedersen, & Nielsen, 2016; Ponsford & Spitz, 2015). In a nationwide study (Denmark) to determine the rates of RTW after severe TBI, Odgaard and colleagues (2016) analyzed heath care data and found that approximately 11% returned to stable employment one-year post-injury. This was comparatively lower than the 23% in the United States as reported by Ketchum and colleagues (2012). However, two-years post-injury, 30% of the population had attempted to RTW and only 16% were able to secure employment (Odgaard et al., 2016).

Overall, a review of the literature indicates that RTW rates range from 11%-97% (Arango-Lasprilla et al., 2009; Brooks, McKinlay, Symington, Beattie, & Campsie, 1987; Corrigan et al., 2007; Nakase-Richardson, Yablon, & Sherer, 2007; Shames et al., 2007; Wäljas et al. 2014) and that the general trend appears to be an increasing percentage of patients return to some sort of gainful employment over time following their TBI but a considerable variance exists. This large variance has been attributed to factors such as age (Forslund et al., 2017; Jourdan et al., 2013; Willemse-van Son, Ribbers, Verhagen, & Stam, 2007), gender (Ratcliff, Colantonio, Escobar, Chase, & Vernich, 2005; Willemse-van Son et al., 2007), education (Connelly, Chell, Tennant, Rigby, & Airey, 2006; Ponsford, Draper, & Schonberger, 2008; Sigurdardottir, Andelic, Roe, & Schanke, 2009), prior employment (Andelic, Stevens, Sigurdardottir, Arango-Lasprilla, & Roe, 2013; Connelly et al., 2006; Forslund et al., 2017), and ethnicity (Arango-Lasprilla et al., 2009; Shafi et al., 2007). Moreover, injury severity has also been shown to be a predictor of RTW outcome following TBI (Yasuda, Wehman, Targett, Cifu, & West, 2001). Ruffalo and colleagues (1999) examined patients who had sustained a mTBI through a motor vehicle accident 6-9 months following their injury and found that 12% had returned to work at full capacity and 30% had returned but with modified duties. Importantly, 58% had not returned to any gainful employments since their accident. In contrast, however, Roa and colleagues (1990), found that the majority (66%) of a mixed-severity TBI sample had returned to work within a two-year period. However, other studies have found contrasting results. To this end, one seminal study which examined approximately 2,962 TBI patients found that at one-year post-injury, 47% of the mTBI, 45% of the moderate TBI, and 48% of the severe TBI patients continued to report disability (Thornhill, Teasdale, Murray, McEwen, & Roy, 2000). In other words, nearly half of the patients who had sustained TBI of any severity were disabled after one-year. With such variability in the RTW literature, it is difficult for the clinician or policymaker to make sense or use of this information.

3.3 Economic burden

TBI-related disability has economic impacts on the individual, their family, and society at large. Young people (16-24) have the highest incidence rates for TBI, which is unfortunate considering the greater loss of potential productivity. The global costs of TBI-related disability have not been estimated, however, the WHO commented that the costs associated with TBI that is most often reported in the literature appears to be related to the costs associated to hospitalization (WHO, 2006). Estimations of the loss of productivity in terms of dollars have been difficult to study for multiple reasons but the National Institute of Health (NIH) has estimated that the national and annual costs to health care is approximately $9-10 billion (in 1998 dollars; NIH, 1998). Per person, the average lifetime cost of health care following a severe TBI has been estimated at $600,000 to $1,875,000 (NIH, 1999). However, even the NIH panel that provided this estimate cautioned that this

> "may grossly underestimate the economic burden of TBI to family and society because they do not include lost earnings, costs to social services systems, and the value of the time and foregone earnings of the family members who care for persons with TBI" (NIH, 1999, p. 977).

Accordingly, economic costs need to be measured across multiple stratified factors to better reflect the real-world costs associated with sustaining a TBI. Typically, estimates only include hospital costs associated to the acute TBI. Some researchers have pointed out that this is not an accurate reflection of the costs associated to a TBI as the acute costs make up only a fraction of the direct costs and the indirect costs are completely omitted. Direct costs include goods and services associated in the acute and rehabilitation periods of treatment such as ambulance, emergency department, inpatient and outpatient hospital services, along with costs associated with professional health care services and physical and vocational rehabilitation (Schulman, Sacks, & Provenzano, 2002). Moreover, medication, medical supplies, devices, and the patient's time costs fall under the category of direct costs. Indirect costs can be defined as loss of productivity due to mortality and morbidity (Humphreys, Wood, Phillips, & Macey, 2013). More specifically, the loss of potential earnings due to disability for the patient who has sustained a TBI and for their family members who take on the burden of care. Estimations of the direct costs associated with TBI range from $98 million to $302 million (Runge, 1993; Schulman et al., 2002) with per patient costs of $67,504 to $114,231 (Ashley, Schultz, Bryan, Krych, & Hays, 1997). Indirect costs estimates have varied considerably in the literature with estimations ranging from $521 million to $2.8 billion (Kayani, Homan, Yun, & Zhu, 2009; Schulman et al., 2002). These variances can be attributed to regional data, the degree of inclusion of variables, and the specificity of those variables (Humphrey et al., 2013). On a national level, the annual direct costs (i.e., acute medical care and rehabilitation) have been estimated to be roughly $6 billion USD with indirect costs (i.e., loss of productivity, ongoing health maintenance, and long-term care) at $22 billion USD (Yasuda et al., 2001). Thus, the total estimated economic burden of TBI in the United States is approximately $28 billion USD.

Fewer studies have been published on the costs associated with TBI in Canada with most of these focused on the province of Ontario. To this end, in the Canadian Institute for Health Information report on the burden of neurological diseases, disorders and injuries in Canada (CIHI, 2007), the Public Health Agency of Canada estimated that the total direct costs for head injury in 2000-2001 was approximately $151.7 million. However, a report to the Ontario Neurotrauma Foundation conservatively calculated the direct costs to be approximately $120.7 million (Colantonio, Parsons, Vander Laan, & Zagorski, 2009). Although these two figures have been estimated by two different teams of researchers with differences in methodology and inclusion/exclusion criteria, a simple naive comparison shows that Ontario's direct costs associated with TBI is roughly 79.5% of the Canadian total, despite Ontario only having 38% of Canada's patient population (Chen et al., 2012). This incongruence highlights both the need for consistency across economic studies and the substantial direct costs associated with TBI.

In an effort to determine the real-world direct costs of TBI up to three-years post-injury from multiple Ontario administrative healthcare databases, Chen and colleagues (2012) expanded their search to include emergency department care, acute inpatient care, inpatient rehabilitation, and complex continuing care. They found that the direct medical costs were $446 million CAD (2009 dollars) in the first three years of post-injury. Moreover, they calculated the average first-year direct costs were approximately $32,132 decreasing to $2,580 for the second year and $2,234 for the third year. Although the authors sought to determine the costs of TBI in the first three years, there were a few limitations. First, they included most but not all, of the health care services that were paid for by the government; costs such as acute diagnostic services, medication, and medical supplies were not included in their calculations. The study also excluded all private insurance or out-of-pocket expenses which limits the generalizability of their findings especially to those claiming disability benefits through third parties. Moreover, the study only looked at direct costs and therefore did not estimate any loss of income of the patient or their caregiver was not calculated. To this end, the authors acknowledged that the indirect costs likely outweigh direct ones due to lingering impairments and disability that hinders their full income potential.

Additionally, Chen and colleagues (2012) did not directly measure incidence or costs in the emergency department settings but rather used multiple databases. The use of multiple databases has obvious benefits in increasing overall sample size, but a glaring drawback is the potential lack of consistency of the measurement of the variable of interest across databases. Accordingly, Fu and colleagues (2016) conducted a follow-up study to describe the epidemiological patterns and lifetime costs associated to TBI-related EDs in Ontario from April 2009 to March 2010. According to the Workplace Safety and Insurance Board of Ontario, the average number of days off work following a TBI is 38.1 for males and 51.7 for females (Colantonio et al., 2010). Using this, Fu and colleagues (2016) quantified the direct costs and the loss of productivity while considering age, gender, daily average income obtained from Statistics Canada along with the probability of lost workdays following a TBI. Using the incidence data from the National Ambulatory Care Resource System database, they found that in 2009, there were 133,952 TBI-related emergency department visits and the annual lifetime costs were calculated to be $945 million. Of this amount, $292 million represented direct costs (31%) and $653 million (69%) as indirect costs. They also reported their findings using various inclusion criteria, which resulted in lifetime cost estimations ranging from $279 million to $1.22 billion.

A direct comparison between the $446 million estimated over the first three years in the Chen et al. (2012) study and the Fu et al., (2016) annual findings cannot be made, but the discrepancies in the cost estimates can be attributed to in part to the study period, data sources, inclusion criteria, and cost methodology (Fu et al., 2016). Fu and colleagues (2016) also adjusted their calculations to compare against the CIHI's findings in 2007. They used the same age and gender incidence weights to their truncated data, which included all hospital, physician, and drug costs to reflect the CIHI's methods. They found that their national cost estimates were $754 million as compared to $183 million by the CIHI (correcting for inflation up to 2009). The authors attributed these differences to lack of inclusion of long-term data along with an increase in the rate of TBIs and severity. In addition, the authors identified specific demographic factors that echo the incidence literature at large. Namely, that the male gender, youths to young-adults along with the very old were most at risk for TBI. In addition, they found a high incidence of TBIs caused by motor vehicle accidents with a disproportionately high cost. By mechanism of injury, they found that approximately 10% of their population sustained TBIs through a motor vehicle accident but this accounted for 17% of the total lifetime costs. Furthermore, their results indicated that approximately 12% of studied data incurred a TBI through sports-and-bicycle-related injuries. Overall, the authors documented that TBI-related injuries represent a substantial health and economic burden for the province of Ontario.

Given the costs associated with failing to RTW, predicting employment capacity in an effort to refer to the appropriate rehabilitation program should of highest priority for clinicians. As per the results from the first study, an overall effect between mTBI and control across neuropsychological test performance in the post-acute period was found. Moreover, significant effect sizes were found between all eight domains and post-acute period patients with mTBI. The largest effect sizes were found in tests of Language (g = -0.73) and Verbal Fluency (g = -0.61). Despite the relatively large overlap, these findings showed support for at some sensitivity of traditional pen-and-paper neuropsychological test measures in distinguishing between controls and patients with post-acute mTBI.

Some researchers have shown that the expected recovery time from an mTBI is approximately three-months with no residual neuropsychological deficits (Belanger et al., 2005; Binder et al., 1997; McCrea et al., 2009; Schretlen & Shapiro, 2003; van Zomeren & Brouwer, 1994). In other words, these researchers report that patients with mTBI should be indistinguishable from healthy controls after three months based on neuropsychological test performance. To this end, Rohling and colleagues (2011) found very small effect sizes differentiating between mTBI and controls that suggested that “the clinician is more likely to be incorrect than correct by inferring persistent deficit in someone who has suffered a single, uncomplicated MTBI (i.e., positive predictive values will be less than 0.50)” (p. 205). Rohling et al., (2011), however, acknowledged that there are some that have persistent cognitive symptoms with subsequent functional disability well beyond this recovery period—the *miserable minority*.

Accordingly, the literature describes two distinct populations: those who have fully recovered from a mTBI and should be cognitively indistinguishable from healthy controls and those who have not recovered from a mTBI (miserable minority) and continue to demonstrate cognitive impairment. Should this be true, then as per the findings of the updated meta-analysis in Study 1, neuropsychological test performance between these two populations may be able to differentiate between the populations. Specifically, performance on tests of language and verbal fluency should be most sensitive as this is what was found from the results of Study 1. This is important as there is a shift in the neuropsychologist’s role from determining site and size of a lesion to determining and predicting the breadth and severity of cognitive dysfunction as it relates to everyday functioning (Marcotte & Grant, 2010; Zakzanis & Grimes, 2016).

The ability to RTW is a decidedly important outcome following mTBI, considering its value to personal autonomy and its significance within Western society as a whole (Machamer, Temkin, Fraser, Doctor, & Dikman, 2005; Prigatano, 1999). It is the variable of focus in much TBI outcome research, as it can also be more reliably examined quantitatively than many other outcome variables (Sherer, Madison, & Hannay, 2000). A number of studies to date have, in some regard, examined the predictive ability of neuropsychological test measures with respect to post-injury vocational status. Few, however, have examined this relationship beyond the three-month duration in which impairment is expected to resolve, and even fewer empirical studies have addressed the issue of ecological validity directly (Sherer et al., 2002). For example, studies of this nature have often focused on the efficacy of rehabilitation programs, while neuropsychological test performance was a peripheral matter of investigation (Lam, Priddy, & Johnson 1991; Malec, Smigielski, DePompolo, & Thompson, 1993; O'Connell, 2000; Ryan, Sautter, Capps, Meneese, & Barth, 1992). Moreover, many of these studies are limited in terms of their non-specific TBI criteria (Bayless et al., 1989; Hanks et al., 2008; Ip et al., 1995; Melamed, Stern, Rahmani, Groswasser, & Najenson, 1985; Ryan et al., 1992; Wood & Rutterford, 2006), small number and narrow scope of tests employed (Bayless et al., 1989; Burke, Wesolowski, & Guth, 1988; Melamed et al., 1985; O'Connell, 2000; Wedell, Oddy, & Jenkins, 1980), and lack of independent analysis of performance on specific tests (Bowman, 1996; Ezrachi, Ben-Yishav, Kay, Diller, & Rattok, 1991; Hanlon, Demery, Martinovich, & Kelly, 1999; Ryan et al., 1992; Wood & Rutterford, 2006).

3.4 Present Study

In keeping with the above noted gaps in the literature and the importance of being able to predict RTW reliably, the current study sought to examine if traditional pen-and-paper neuropsychological test measures performance could differentiate employment status in patients who were in the post-acute period of recovery following mTBI. Specifically, the current study sought to address the following questions:

1. Can traditional pen-and-paper neuropsychological test measures differentiate between employment status in patients who are in the post-acute period of recovery following mTBI by way of test performance?

    a. If so, which specific neuropsychological test measures are most sensitive to employment status?

Based on the results from the first study, it is hypothesized that performance on tests of language and verbal fluency should differentiate the two employment status groups.

3.5 Methods

3.5.1 Participants

Archival data were randomly gathered from a convenience sample database of 186 patients with mTBI who were involved in litigation and were referred for neuropsychological assessment due to ongoing complaints of cognitive impairment. The ACRM guidelines (Kay et al., 1993) were used to diagnose mTBI: (1) GCS score of 13-15 at time of injury, (2) duration of LOC under 20 minutes, and (3) PTA less than 24 hours. This data was obtained from hospital records and ambulance call reports. Further, patients suffering intracranial injuries were excluded based on the presence of lesions in neuroimaging studies. All patients were without history of neurological or psychiatric disorder, including substance abuse and previous head injury. English fluency was confirmed using the Wide Range Achievement Test – Fourth Edition Reading (WRAT4-Reading; Wilkinson & Robertson, 2006); a minimum Grade 8 reading level was required.

A number of steps were taken to prepare the database for the purposes of the current study. Several performance and embedded validity measures were employed to ensure credibility of responses, such as the Test of Memory Malingering (Tombaugh, 1996), Victoria Symptom Validity Test (Slick, 1997) and the Rey Dot Counting Test (Rey, 1941). Criteria for non-credible test results included the endorsement of one or more unusual symptoms (e.g., itchy fingernails, painful hair tips) and failing one (or more) of the aforementioned measures of performance validity. From the database, 80 patients met the criteria for non-credible test findings and were excluded from the study. Additionally, 7 patients were excluded from the analysis due to insufficient information to determine current RTW status. In order to RTW, one must be employed before their injury. As such, 9 patients were removed from the analysis because they were unemployed at the time of their injury. This left 88 patients in the database, however, 42 participants were further excluded due to missing scores[9]. The final sample consisted of 46 patients and were grouped based on employment status at the time of the assessment (see Table 4). Chi-square (i.e., handedness and gender) and t-tests were conducted to determine any differences between the demographic characteristics. No differences were found between the groups with respect to any of the demographic variables listed in Table 4.

Table 4.

Demographic characteristics of the samples (Study 2).

	Employed (n = 13)		Unemployed (n = 33)	
	Mean	SD	Mean	SD
Age (years)	34.8	12.54	37.9	12.92
Gender (% female)	23.08%		45.46%	
Education (years)	13.8	2.27	13.0	2.70
GCS	14.8	0.71	14.6	0.50
Days since injury	1156.5	775.56	727.8	664.46
Handedness (% Right)	92.31%		87.88%	

Note: No significant differences across groups on all demographic characteristics.
Abbreviations: GCS = Glasgow Coma Scale.

At the time of the assessment, patients were informed about the nature and purpose of the assessment and provided written consent to release the results of the examination for research purposes.

3.5.2 Materials

The following tests were administered:

[9] A fixed and flexible battery approach was employed, as is common practice by most neuropsychologists (Rabin et al., 2016; Sweet, Benson, Nelson, & Moberg, 2015). As a result, some measures were not administered to some participants in an effort to curb patient fatigue and extensive testing duration.

Ruff 2 & 7 Selective Attention Test (RSAT; Ruff & Allen, 1996)

The RSAT served as a measure of visual attention, both sustained and selective (Lezak et al., 2012). The test consists of 20 trials administered in consecutive 15-second intervals. Each trial consists of three rows of mixed numbers and capital letters spanning the width of the page. During the trials, the participant is instructed to cross out all 2's and 7's that appear in the trial from left to right, starting from the top row. At the end of each time interval, they are instructed by the examiner to move onto the next trial. This is repeated until completion of all 20 trials. The duration of the RSAT is 5 minutes long. The total speed score as calculated by the RSAT manual was used in this analysis (Ruff & Allen, 1996) and a lower value is indicative of better performance.

Vocabulary (Wechsler, 1999)

The Vocabulary is a language subtest from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) that measures word knowledge and verbal concept formation. Participants are shown 42 words of increasing difficulty and asked to give brief definitions of each word. Each response was compared against the manual to determine a score of fully correct (2 points), partially correct (1 point), or incorrect (0 points). The total raw score on this subtest was recorded and used in the analysis. Higher scores represent better performance.

Trail Making Test (TMT; Lezak et al., 2012)

The TMT is a commonly administered measure of attention, psychomotor speed, cognitive flexibility, and divided attention and is also in the public domain (Strauss et al., 2006). This test initially appeared on the Army Individual Test Battery in 1944 and was later modified for the Halstead Battery (Reitan, 1955). The test consists of two parts, part A and part B. Part A, also referred to as TMT A, consists of a paper with encircled numbers, from 1 to 25, arranged on the page seemingly at random. The participant is instructed to draw a continuous line to connect all the circles in sequential order. Part B, also referred to as TMT B, consists of a paper with encircled numbers (1 to 13) and letters (A to L) arranged on the page seemingly at random. Here, the participant is instructed to once again draw a continuous line to connect all the circled numbers and letters in alternating order (i.e., 1, A, 2, B, 3, C…etc.). In both parts, a sample trial was presented prior to test administration to insure adequacy of the instructions. The participant is instructed to complete the task as quickly as they can without lifting their pencil from the paper. When an error was made, the participant was notified and instructed to correct their mistake(s) and to continue until the test was completed (Strauss et al., 2006). Performance was based on time to completion and lower values indicate better performance (Reitan, 1979).

Digit Span (Wechsler, 1997)

The digit span is a subtest of the Working Memory component of the WAIS-III. This test examined attention, concentration, and working memory performance. The examinee was presented with sequences of numbers progressively increasing in the number of digits. In the first half of the test, the examinee was required to repeat the sequence to the examiner as it was presented. In the second half of the test, the examinee was required to repeat the sequence to the examiner backwards. Performance was based on the maximum digits forward and the maximum digits backwards (Wechsler, 1997). Higher scores represented better performance.

California Verbal Learning Test – II (CVLT-II; Delis et al., 2000)

This test examined verbal learning and memory. A list of words (List A) was read to the examinee five times; after each reading, they were asked to recall as many words as possible. List A consisted of 16 words—four different categories of words and four words from each category. Another list of words (List B) was read to the examinee once to function as interference. The examinee was then tested on List A for short-delay free recall and cued recall, long-delay free recall and cued recall (after a 20-minute delay), and recognition. Performance was based on the total number of words freely recalled (i.e., without a cue) after 20 minutes (Delis et al., 2000). Higher scores represented better performance.

Rey-Osterrieth Complex Figure Test (RCFT; Meyers & Meyers, 1995)

This test examined a variety of cognitive abilities, such as visuospatial ability, memory, attention and planning. Examinees were presented with a complicated line drawing, and asked to reproduce it, first by copying and then from memory. Performance was based on delayed recall with higher scores representing better performance (Osterrieth, 1944).

Brief Visuospatial Memory Test – Revised (BVMT; Benedict, 1997)

This test examined visual learning and memory. The examinee was presented with a card for 10 seconds, on which there were six simple geometric designs in a 2 x 3 matrix. After the 10 second period, the examinee was asked to reproduce as many of the designs as possible on a blank piece of paper. Two more identical trials were performed. After a 25-minute distraction period, the examinee was asked to perform a delayed recall trial (asked to reproduce the designs again), and a recognition trial (shown 12 cards with a different design on each and required to identify the designs from the original matrix). Performance was based on delayed recall (Benedict, 1997).

Boston Naming test (BNT; Kaplan, Goodglass, & Weintraub, 1983)

The BNT is a language test that was designed to measure confrontational word-retrieval. Participants are shown 60 line-drawings of objects and are asked to name them. Performance was based on the total correct score with higher scores relating to better performance.

Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1989; Spreen & Strauss, 1998)

This test was designed to measure phonemic and semantic fluency. As per the instructions (Spreen & Strauss, 1998), participants were asked to think of as many words as possible in 60 seconds following a phonemic (F, A, and S) or semantic (animals) cue. Exclusions for phonemic responses were proper nouns (such as Bob, Boston or Buick), numbers (such as four, three) and the repetition of a word with a different suffix (such as eat and eating). Two scores were extracted for the purposes of the current study; total raw phonemic responses on “F”, “A” and “S” trials along with total raw semantic responses from the animal trial. Higher scores represent better performance.

3.5.3 Procedure

Participant demographics such as days since injury, GCS scores, and age were derived from the patient’s self-report within the context of a comprehensive clinical history and substantiated by way of file review, both undertaken by a licensed clinical neuropsychologist. Participants were administered the above noted tests by a psychometrist under the supervision of a registered clinical neuropsychologist. The examinees were grouped according to employment status (employed or unemployed) as determined by the examinees’ self-report on vocational status and, when available, corroborated by way of a collateral interview with the patient’s significant other (e.g., spouse, child, or parent). Employment status was also substantiated based on whether participants were receiving income replacement benefits.

3.5.4 Statistical Analysis

Since employment status was a binary outcome as the person either returned to work or did not, this violated the normality assumption required for traditional regression analysis. Thus, a logistic regression with a logit link function was used to model employment status. All statistical calculations were completed using the IBM Statistical Package for the Social Sciences (SPSS) for Windows Version 22.0 (IBM Corp, 2013).

3.6 Results

Independent sample, two-tailed t-tests revealed no significant differences in GCS scores (p = 0.57), days since injury (p = 0.07), or age (p = 0.45) between the two groups. Chi-square analysis revealed no significant differences in handedness ($\chi^2(1) = 0.281$, $p = 0.596$) and gender ($\chi^2(1) = 1.96$, $p = 0.161$) between the two employment groups.

Employment was modeled as a function of neuropsychological test performance controlling for age and days since injury. A total of 46 cases were analyzed and a test of the full model versus a model with intercept only was statistically significant (Omnibus $\chi^2(14) = 37.30$, $p < 0.001$) indicating that the full model significantly predicted employment status. The model accounted for between 55.5% to 79.8% of the variance in employment status, with 93.9% of those who were unemployed successfully predicted. Moreover, 84.6% of those who were employed were accurately predicted. Overall, 91.3% of predictions were accurate. Table 5 gives coefficients, Wald statistic, probability values, and odds ratio for each of the predictor variables. This shows that performance on none of the neuropsychological tests significantly predicted employment status. Age, however, approached statistical significance (p = 0.058) suggesting a trend where an increase of one year of age is associated with a decrease in the odds of employment by a factor of 0.71 (95% *CI* 0.50 and 1.01).

Table 5.

Logistic Regression Predicting Employment Status from Age, Days Since Injury, and Neuropsychological Test Performance.

Predictor	*B*	Wald χ^2	*p*	Odds Ratio
Age	-0.34	3.59	0.058	0.71
Days Since Injury	0.00	0.98	0.322	1.00
WASI Vocabulary	0.06	0.20	0.652	1.06
RSAT	0.19	2.71	0.099	1.21
TMT A	-0.03	0.03	0.873	0.97
TMT B	0.06	1.95	0.162	1.06
Digit Span Forward	-1.20	2.75	0.097	0.30
Digit Span Backward	-0.65	0.68	0.409	0.52
CVLT-II	-0.63	1.20	0.274	0.53
RCFT	-0.02	0.01	0.929	0.98
BVMT	-0.64	0.62	0.432	0.53
BNT	-0.10	0.34	0.559	0.91
Phonemic Fluency	0.26	1.97	0.161	1.29
Semantic Fluency	0.95	2.73	0.098	2.58

Note: Negative values in the *B* coefficient column indicate that the odds of returning to work decrease as the predictor value increases.

3.7 Discussion

The present study sought to determine if traditional pen-and-paper neuropsychological test measures could differentiate between employment status in patients who were in the post-acute period of recovery following a mTBI, and if so, which ones would be better predictors. The probability of the semantic fluency test predicting employment status approached statistical significance ($p = 0.098$) with an odds-ratio that indicated that a point increase in semantic fluency was associated with return to employment by a factor of 2.58 (95% *CI* 0.84 and 7.95).

Although limited by the small sample size of the employed group, performance between the two employment groups approach statistical significance on the semantic fluency test. This result is comparable with study 1 in that performance on verbal fluency measures were found to be differentiating between the groups. Whereas verbal fluency performance differentiated between patients and controls in Study 1, the current findings suggest that result can be extended to potentially differentiate within patients' employment status. Moreover, Raskin & Rearick (1996) found that patients who had sustained a mTBI produced significantly fewer words as compared to controls. Another study found that verbal fluency performance was found to be predictive of employment status 3-15 months after sustaining a mTBI (Drake, Gray, Yoder, Pramuka, & Llewellyn, 2000). These authors interpreted this to reflect of subtle changes in self-monitoring, planning, and initiation that were relevant to RTW following mTBI. Similar findings were observed in severe TBI. Henry and Crawford (2004) have shown through a meta-analytic review of verbal fluency following focal cortical lesions that phonemic fluency is related to frontal regions of the brain, whereas semantic fluency is associated with frontal and temporal regions (but more so temporal). Zakzanis, McDonald, & Troyer (2013) found that severe TBI patients generated significantly fewer words than control participants. This difference was attributed to a deficit in their ability to switch from one semantic category to another. The authors suggested that this switching deficit may be related to a frontal brain dysfunction with increased reliance on the temporal lobes. In sum, performance on semantic fluency has the potential to differentiate between employment status following mTBI but replication with larger sample sizes and increased variety of verbal fluency tests are necessary to provide reliable evidence of clinical utility.

Based on the results from Study 1, language and verbal fluency tests were hypothesized to be predictive of employment status. The BNT and COWAT, the latter of which is comprised of the phonemic and semantic fluency tests, are the top two most frequently used tests in a recent survey of neuropsychologists (see Table 17 of Rabin et al., 2016). Moreover, the Vocabulary subtest from the Weschler series of tests is also listed as one of the most frequently used tests to measure language function (Rabin et at al., 2016). Interestingly, the BNT and COWAT were ranked as 19th and 30th in a previous survey where respondents were asked to list tests that they used to help predict the ability to RTW (see Table 16 from Rabin et al., 2005). The findings of the current study suggest only partial support that performance on semantic fluency test from the COWAT may be able to differentiate between employment status in patients who are in the post-acute recovery period after a mTBI. Cross-sectional replication studies with larger sample sizes are needed to confirm the reliability of the current findings.

The results of the current study demonstrate that the Digit Span was not able to predict employment status in an mTBI sample but approached statistical significance (p = 0.097). This was in contrast to other studies which have reported that the Digit Span was part of a ‘Working Memory’ factor, or grouping of tests, that was significantly predictive of employment outcome (Wood & Rutterford, 2006). Further, it was included in a regression equation that significantly predicted degree of occupational functioning on a 9-point scale (Bowman, 1996). In another study, Hart and colleagues (2003) found that performance of Digit Span was significantly related to post-TBI level of caregiver supervision. All of these studies utilized mixed samples of TBI severities and did not stratify based on severity. To this end, performance on working memory tasks may be spared following a mTBI and not significant predictor of employment status. Since the current study used only one measure of working memory, replication with additional measures of working memory would be required in order to support this claim.

The finding that performance across all tests was not predictive of employment status may be explained by several factors. First, there is concern of a significant disconnect between the test environment and the natural world (Levine, Dawson, Schwartz, Boutet, & Stuss, 2000; Norris & Tate, 2000; Sbordone, 1996). For example, Lezak (1976) recommended that testing be conducted in a sterile environment—free of distracting objects, colours and noises—in order to minimize confounds on test performance. This is reflective of the testing environment in which neuropsychological assessments are all undertaken, which includes how the results of the current study were collected. Such an approach neglects the myriad of external stimuli that can impinge on functioning in the work place and makes it difficult to infer real-world functionality from test results (Sbordone, 1996).

Second, neuropsychological test measures tend to be framed by a firm set of structures and rules, with the examiner often prompting responses from the participants. Through this contrived structure, many neuropsychological test measures are limited in terms of addressing only a small sample of the number of behaviours that could be exhibited (Chaytor & Schmitter-Edgecombe, 2003; Levine et al., 2000). These tests may even prevent the exhibition of pathognomic behaviours as a result of their rigidity and the sterile environment in which they are administered (Sbordone, 1996). Furthermore, the majority of neuropsychological test measures may not be detailed enough for a multifactorial outcome variable such as RTW, only tapping into mere aspects of complex processes involved in functional abilities, rather than the entire process (Sbordone, 1996; Wilson, 1993). For example, the Digit Span test and the TMT may be administered to evaluate a patient’s deficits in attention and concentration, a domain in which deficits may be reflective of workplace disability. These two tests may not be sufficient to encapsulate the complex process that is attention and concentration, which can be broken down into such factors as alertness, stimulus selectivity, vigilance, flexibility, resistance to fatigue, and processing speed to name but a few (Posner & Peterson, 1990; Sbordone, 1996). Indeed, while the constituent factors making up complex processes may appear to be intact, the initiation and synthesis of these factors required to carry out a sophisticated task may be hindered (Burgess & Alderman, 1990).

Third, and in extension of the previous point, many of the neuropsychological test measures that are now being administered for clinical purposes were originally developed for research purposes (Holt et al., 2011; Manchester et al., 2004). There is an inherent problem in this, as the requirements that must be met for tests developed for research differ from those that are necessary for the purposes of clinical assessments (Burgess et al., 2006). This is not to say that where and how a test was developed has any bearing on its clinical utility. Indeed, adequate standardization along with studies reliably validating the construct and discriminant validity are necessary. Moreover, ecological validity can be determined via the veridicality approach but there is currently a dearth of such studies that are related to patients who are in the post-acute period following a mTBI.

Finally, patients with TBIs may employ the use of compensatory strategies in testing environments whereas they would not do so in the real-world (Chaytor & Schmitter-Edgecombe, 2003; Sbordone, 1996). Chaytor and colleagues (2006) reported that patients with TBI demonstrated a conscious effort to keep track of failed attempts and made alterations at each attempt in order to complete tests. Even the clinician may engage in behaviours that compensate for the patient's functional deficits, such as modifying test instructions or providing extra cues and prompts to facilitate test completion (Sbordone, 1996). Compensatory behaviors were not measured in the archived database that was used for the current study. Thus, it is not clear if such strategies may have impacted the findings. Clinicians should be conscious of this issue and strive to minimize its impact (Sbordone, 1996).

3.7.1 Limitations

There are a number of limitations to the present study, all of which are suggestions of future directions that could be taken. Firstly, the number of patients in the employed group (n = 13) was small. Accordingly, there is a possibility of insufficient power to overcome the effects of outliers within this group. Replication with a larger sample size would be needed to confirm the pattern of results found in the current study. Second, occupation was not stratified based on type. Such stratification would control for cognitive demands across occupation type (Guilmette & Kastner, 1996; Humphrey & Oddy, 1980). For example, the occupational cognitive demands of a coal miner may differ than a lawyer. To this end, an inability to RTW due to cognitive impairment may be relative. However, such stratification would require a larger overall sample size to statistically control for cognitive demands using inferential statistics. Third, the present study only examined the ecological validity of neuropsychological test measures as it pertained to the ability to RTW. Participants that returned to work in a limited capacity (part-time or with decreased responsibility) were not differentiated, nor were other realms of everyday living addressed, such as cooking and home maintenance. Future replications of the current study are encouraged to segregate RTW based on full-duty or limited-duty (Drake et al., 2000). Fourth, psycho-emotional measures were not included in this study. There exists a category of patients with mTBI colloquially referred to as the miserable minority, which is a subgroup defined by persisting negative emotional symptoms (e.g., depression) as well as a plethora of cognitive and physical complaints (see Rohling et al., 2012). Since patients exhibiting the characteristics of this subgroup have been associated with poorer outcomes, psycho-emotional measures may be more sensitive to the identification of such patients and, thereby, more predictive of disability (Ruff, 2005; Ruff et al., 1996). Such studies are crucial in fully understanding the clinical utility of traditional pen-and-paper neuropsychological test measures when attempting to predict disability experienced in everyday life in patients with mTBI.

3.7.2 Conclusion

The ability to RTW is undoubtedly an essential part of being able to resume pre-injury way of life (Machamer et al., 2005) and neuropsychological testing plays a large role in the assessment of patients who have sustained a mTBI (Lezak et al., 2012). The results of such testing can have far reaching consequences in terms of access to rehabilitative and financial assistance (Acker & Davis, 1989) therefore the predictive validity of neuropsychological test measures with respect to disability is of great importance (Chaytor & Schmitter-Edgecombe, 2003; Sherer et al., 2002).

The current study found that performance on traditional neuropsychological test measures did not predict employment status in a sample of patients who were in the post-acute recovery period following an mTBI. This finding suggests low veridicality of traditional pen-and-paper neuropsychological test measures in a mTBI population. However, the current findings do suggest a potential for performance on semantic fluency to be predictive of employment status. Replication studies with larger sample sizes in the employment group are required to determine the reliability of this claim. These findings may highlight the importance of other facets of a complete neuropsychological assessment, such as the clinical interview and behavioural observations (Wilson, 1993). More importantly, the results of the current study underscore the necessity of developing new and improved techniques. The assessment of patients in a variety of settings, in particular those in which they experience difficulties, may be warranted (Burgess et al., 1998). It may also be beneficial for neuropsychologists to cooperate with other health-care professionals such as occupational therapists when planning for interventions for patients in the recovery stages of mTBI.

Chapter 4

4 Study 3: Ecological Validity of the BADS in mTBI

For patients who sustain a mTBI, it has been argued that standardized neuropsychological test measures such as the Wisconsin Card Sorting Test (WCST; Kongs, Thompson, Iverson, & Heaton, 2000) are not sensitive in determining disability status (Campbell et al., 2009). Certainly, it would be a stretch to assume that one's ability to sort cards is predictive of one's ability to RTW, school, or play. Evidence for this is weak at best and the use of traditional pen-and-paper based neuropsychological test measures have been questioned in terms of their ecological validity. A review of the relationship between neuropsychological tests and outcome found that the majority of tests were either weakly or completely unrelated to measures of outcome (Chaytor & Schmitter-Edgecombe, 2003). In addition, small to moderate effect sizes have been reported in the relationship between neuropsychological test performance and real-world tasks such as driving (Wither et al., 2000), RTW (Kibby et al., 1998), financial management (Okonkwo et al., 2006), and iADLs (Farias et al., 2003).

4.1 Functional Outcome

With respect to RTW, a review of the relationships between neuropsychological assessment variables and vocational status following a mTBI found that performance on neuropsychological test measures 12-36 months after injury were not associated with vocational status (Nolin & Heroux, 2006). Similar results have been reported for neuropsychological performance in the acute periods after injury. Ruffolo, Friedland, Dawson, Colantonio, and Lindsay (1999) examined a sample of 50 patients who sustained a mTBI following a motor vehicle accident one-month post-injury. Their results indicated that cognitive performance was not associated with RTW. In a more recent study, Wäljas and colleagues (2014), who examined various factors relating to RTW in a sample of 109 patients with mTBI, reported that performance on neuropsychological test measures at three-four weeks post-injury were not related to time off work.

Indeed, a robust correlation between performance on neuropsychological test measures and RTW in a sample of patients with mTBI has yet to be demonstrated (Friedland & Dawson, 2001; Mooney, Speed, & Sheppard, 2005; Shames et al., 2007; Temkin, Holubkov, Machamer, Winn, & Dikmen, 1995; Van der Naalt, Zomeren, Sluiter, & Minderhoud, 1999; Vanderploeg, Curtiss, Duchnick, & Luis, 2003). To this end, the results from Study 2 concluded that traditional pen-and-paper neuropsychological test measures were insensitive to employment status in mTBI in the post-acute period. In contrast, however, performance on neuropsychological test measures has been associated with functional outcome as measured by the Functional Independence Measure (Keith, Granger, Hamilton, & Sherman, 1987) for moderate-severe TBI (Smith-Knapp, Corrigan, & Arnett, 1996). Similar relationships have been reported for moderate to severe TBI but not for mTBI (Cattelani, Tansi, Lombardi, & Mazzucchi, 2002; Johansson & Bernspang, 2001; Ryan et al., 1992). Thus, it seems that traditional pen-and-paper neuropsychological test measures may be sensitive to moderate to severe TBI but are insensitive in determining functional outcome for patients who have sustained a mTBI.

4.2 Executive Dysfunction

As detailed in Chapter 1, there may be multiple factors to account for the disconnect between neuropsychological performance and functional outcome such as insensitivity to subtle cognitive changes, the artificial test environment, short sampling window, domain overlap of tests, premorbid functioning, and underutilization of compensatory strategies. In addition, and historically speaking, experimental tests developed by cognitive neuroscience have been adapted in the clinic (for a review, see Burgess et al., 2006; Marcotte & Grant, 2010; Stuss & Levine, 2002). The justification for this was based on the construct, operation, and function model used in cognitive neuroscience (see Burgess et al., 2006). Since the *construct* (e.g., attention) and *operation* (e.g., speed of processing) had high face validity with respect to cognitive characteristics that would be important for normal functioning in one's day to day, the *function* (e.g., scores on a reaction time measure) was used as a proxy for functional impairment as clinicians would have a *sense* of normal performance (Stuss & Levine, 2002). This has been especially difficult for the domain of executive functions (Burgess et al., 2006; Lezak et al., 2012; Stuss & Levine, 2002). Ruff (2003) has stated,

> "for certain cognitive functions such as memory and language, ecological validity is more easily extrapolated…domains such as visuospatial reasoning and especially for executive functioning, it is more difficult to draw conclusions…the extent to which the WCST is able to predict daily problem solving is left up to the clinician's imagination" (p. 852).

Indeed, Stuss and Levine (2002) argue that whilst other domains such as memory or language have clinical measures that have been derived from recent detailed empirical analysis, tests of executive functions have not. Moreover, they go on to state that,

> “standard [executive] tests have come into use through years of clinical practice and form the basis of the “frontal” part of neuropsychological assessment batteries. Some were developed in context of focal lesion research, but most were classified based upon their face validity as executive measures” (Stuss & Levine, 2002, p. 409).

This is troubling as the goals of experimental psychology differ than clinical psychology. To this end, the goals of experimental psychology are to observe, describe, and measure behaviour for the pursuit of knowledge, whereas, the goal of clinical neuropsychology is to assess behaviour, provide strengths and weaknesses in a diagnostic framework, and to recommend appropriate treatment based to optimize outcome. As noted in the general introduction, outcome has shifted from site and size of the lesion to functional ability. This is especially true for neuropsychologists who primarily assess patients with TBI. Whilst the question has changed, the tools that neuropsychologists employ have not. The lack of ecological validity of traditional pen-and-paper neuropsychological test measures has been generally cited (Marcotte & Grant, 2010; Sbordone & Purisch, 1996) and many have opined that clinical neuropsychology is at a developmental stage that is ready to adopt new tests that relate test performance to real-world functioning (Burgess et al., 2006; Ruff, 2003).

4.3 Ecologically Oriented Tests

More recently, there has been a shift towards the development of ecologically oriented assessments due to the growing limitations of traditional pen-and-paper neuropsychological test measures (Bate, Mathias, & Crawford, 2001). For example, the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1996) was developed to measure executive functioning with the concept of verisimilitude in mind (tests that resemble real-world situations). It consists of six sub-tests that were designed by the authors to mimic real-life activities that would cause difficulties for some patients with dysexecutive functioning. The authors of the battery argue that the successful performance of everyday activities consists of being able to plan and set priorities for long periods of time in light of multiple competing stimuli, whereas these parameters are rare to find in traditional pen-and-paper neuropsychological test measures.

There is evidence to support that the BADS has good concurrent validity with other measures of executive dysfunction (Bennett et al., 2005; Norris & Tate, 2000). In addition, a study concluded that the BADS was more sensitive to executive dysfunction as compared to traditional tests (Bennett et al., 2005). It was found to be significantly related to informant-ratings of disability on the DEX, suggesting good ecological validity (Evans et al, 1997; Wilson et al., 1996). However, although it was developed in an effort to be more ecologically valid, the BADS appear to be limited in its predictive ability with respect to functional outcome in patients with brain injury (Norris & Tate, 2000; Wood & Liossi, 2006) and individuals with higher educational achievement (Sohlberg & Mateer, 2001). Moreover, McGeorge & colleagues (2001) compared patients who were reported to have problems planning by their care staff with controls on the BADS. Their results indicated that the patient's performance on the BADS were within normal limits as per the BADS normative values. Similarly, medium-sized correlations were found between only two subtests of the BADS (Zoo-Map and the Six Elements Test) when comparing a sample of brain injured patients to an actual real-life test (see below) and the BADS (Alderman et al., 2003).

Clinical support for the usefulness of the BADS has been shown in diverse neurological patient populations. For example, significant differences in performance on the BADS total score were found between healthy controls and patients with suspected Alzheimer's disease (Armentano, Porto, Brucki, & Nitrini, 2009). Moreover, differences in performance on the BADS were also able to differentiate between Alzheimer's and amnestic mild cognitive impairment (Canali, Brucki, & Bueno, 2007; Armentano et al., 2009). The BADS was also found to be more sensitive to executive dysfunction in patients with non-dementing Parkinson's disease than traditional tests of executive functions (Perfetti et al., 2010). Similar results have been found with respect to TBI. In an effort to determine the sensitivity of the BADS to real-world executive dysfunction, Bennett et al. (2005) administered the battery along with other traditional tests of executive functions (i.e., TMT, WCST, Porteus Maze Test, Controlled Oral Word Association Test, Revised Tinker Toy Test, and the Modified Cognitive Estimation Test) on a sample 64 mixed severity (GCS 3/15 to 15/15) patients with TBI. Ecological validity was operationalized via the results on the Dysexecutive Questionnaire (DEX) ratings from an Occupational Therapist and Neuropsychologist. A step-wise regression analysis using all neuropsychological test measures as predictors and DEX ratings as the outcome measure found that the Modified Six Elements test from the BADS and the Porteus Maze Test accounted for 39% of the variance. Moreover, the analysis was repeated using the DEX ratings from the Occupational Therapist as the outcome predictor, which resulted in only the Action Program Test from the BADS contributed a statistical and unique prediction of the outcome. The authors concluded that the BADS contains useful measures that are more sensitive to executive dysfunctional and ecological validity than traditional tests. Encouragingly, similar results have been reported in other mixed-severity TBI populations with specific emphasis on the Modified Six Elements from the BADS as being able to correctly discriminating group membership between patients and controls (Boelen et al., 2009; Ghawami, Sadeghi, Raghibi, & Rahimi-Movaghar, 2017; Norris & Tate, 2000; Rochat, Ammann, Mayer, Annoni, & Linden, 2009).

While the above briefly summarized the utility of the BADS in moderate-to-severe TBI, less is known in an mTBI sample. To our knowledge, there has only been one investigation using the BADS in patients who have sustained a mTBI. Erez, Rothschild, Katz, Tuchner, and Hartman-Maeir (2009) primarily investigated the relationship between self-awareness and executive functioning with participation in daily life in a small sample of patients who were in the acute and post-acute periods of mTBI (*n* = 13). They found that the Rule Shift Cards, Modified Six Elements, and Zoo Map subtests from the BADS significantly differentiated their mTBI sample from Israeli normative data (Dvir et al., 2003). Moreover, significant and moderate relationships were found between these BADS subtests and outcomes such as money management and employment. The authors concluded that the BADS was shown to have clinical utility and ecological validity in an mTBI sample. Limitations of this study, however, include the small sample size and the heterogeneity of recovery stages in the sample.

4.4 Present Study

Based on the above review, the BADS is a promising tool that may be ecologically valid but further investigation into its sensitivity to employment status in a mTBI sample is needed. Accordingly, in keeping with the ecological validity of the BADS, and in keeping with real-world disability in the everyday lives reported by patients with mTBI beyond the expected benchmark for neuropsychological recovery, the present study was undertaken to address the following research questions:

1. Is the BADS better (i.e., more ecologically valid) at differentiating between employment status compared to traditional pen-and-paper neuropsychological test measures in a sample of patients in the post-acute period of recovery after mTBI?

2. As a follow-up, which test is best at being able to differentiate between the two groups.

Based on the above review and preliminary findings by Erez et al. (2009), it is hypothesized that performance on the BADS subtests will be better predictive of employment outcome than traditional pen-and-paper neuropsychological test measures.

4.5 Methods

4.5.1 Participants

Similar to Study 2, archival data were gathered from a random sample of litigating patients who incurred a mTBI and were referred for a neuropsychological examination due to subjective complaints related to cognitive function. In addition, all patients were without history of neurological and psychiatric disorders, including substance abuse and previous head injury. English fluency confirmation along with malingering and diagnostic criteria for mTBI were similar to what was reported in Study 2.

The final sample consisted of 102 participants, of which 45 overlapped with Study 2. The sample was grouped based on employment status; 30 had returned to some form of gainful employment (i.e., full-time, part-time, modified duties/hours) and 72 who had not. Demographic information along with GCS scores, years of education, pre-morbid FSIQ estimates, days since injury, and handedness of the groups are presented in Table 6. Chi-square (i.e., handedness and gender) and t-tests were conducted to determine any differences between the demographic characteristics. No differences were found between the groups with respect to any of the demographic variables listed in Table 6.

At the time of examination, all examinees were informed about the purpose of assessment and that the resulting outcomes would be forwarded to the referring party. To this end, all examinees understood the nature and purpose to the examination and provided written consent to release the results of the examination. In addition, all examinees provided written consent to the use of outcomes of this study for research purposes. Ethics approval for archival research was permitted by the University of Toronto Research Ethics Board.

Table 6.

Demographic Characteristics of the Samples (Study 3).

	Employed (*n* = 30)		Unemployed (*n* = 72)	
	Mean	SD	Mean	SD
Age (years)	34.5	14.19	44	14.34
Gender (% female)	33.30%		50.00%	
Education (years)	13.7	2.07	12.8	3.36
FSIQ Estimate*	100.2	14.53	94.71	13.55
GCS	14.8	0.5	14.9	0.4
Days since injury	925.9	697.86	734.5	551.83
Handedness (% Right)	86.70%		88.90%	

* Based on available data from *n* = 22 employed and *n* = 52 unemployed
Note: No significant differences across groups on all demographic characteristics.
Abbreviations: FSIQ (Full Scale Intelligence Quotient); GCS (Glasgow Coma Scale).

4.5.2 Materials

The Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1996)

The BADS (Wilson et al., 1996) contains six subtests tapping into different domains of executive functioning such as prospective planning, problem solving and mental flexibility (Wilson et al., 1996). For the purposes of the current study and due the limited data available for the Temporal Judgment subtest, only five of the six subtests were examined. A brief description of each of the subtests follows.

Rule Shift Cards. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

The Rule Shift Cards is a subtest from the BADS battery that assesses cognitive flexibility and set shifting. There are two trials in this subtest, both utilizing 20 playing cards presented on a flip book. In the first, which is considered the practice trial, the participant is instructed to respond “yes” to seeing red playing cards and “no” to black cards. On the second, which is considered the scored trial, the rules are switched to “yes” if the card was shown was the same colour as the preceding card, and “no” if it was different. For both trials, the rules were typed out and placed in view of the examinee throughout the test. The scores were based on the time taken to respond to trial 2, the number of errors made, time violations (if overall time was >67 seconds), and the profile score as per the manual (Wilson et al., 1996). The examinee was scored on the time taken to respond and the number of errors made (Wilson et al., 1996).

Action Program. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

This subtest examined ability to solve novel problems by planning a course of action. The examinee was required to retrieve a cork placed inside a tube using various tools provided, while staying within few rules they could not break. The solution required five simple steps not beyond an examinee's abilities. There was no time limit for this task, though it is worth noting that a prompt would be given if the examinee were unable to get past any one of the five steps. Examinee was scored based on the number of steps independently solved (Wilson et al., 1996).

Key Search. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

This subtest was one of three subtests within the BADS that examined the ability to plan ahead and regulate behaviour. The subject was given a piece of paper and told to imagine that it represented "a large field in which they have lost their keys" (Wilson et al, 1996, p18). They were to draw on the paper, the trail they would walk on the field in order to retrieve the lost key. Examinees were scored on various attributes of the drawn pattern (e.g. the spot from which they left and entered the field, the continuity of the trail, parallel and horizontal pattern of the trail). This subtest was timed– the examinee lost one point if they take longer than 95-seconds to complete the task (Wilson et al., 1996).

Zoo Map. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

This subtest also examined the ability to plan an effective course of action and behavioural regulation. Examinees were required to visit certain attractions on the map while adhering to specific rules. The construction of the test was such that there were only four strategies that could be followed in order to succeed. Two versions, – both with the same objectives – were given. The high demand version, consisting of more strict guidelines, was administered first, and the low demand version was administered afterwards. The examinee was scored on his or her ability to adhere to the rules as well as the continuity of their planned path (Wilson et al., 1996).

Modified Six Elements Test. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, et al., 1996)

The Modified Six Elements is a subtest from the BADS battery that assesses planning and multitasking abilities. Here, participants are to complete three different tasks within 10-minutes. The first task is to dictate an event, the second task is arithmetic problems, and the third task involves naming line drawn pictures. Each task has two trials labeled 'A' and 'B'. As per the instructions, the examinees are told that they are to attempt to do at least some of all six subtasks within the 10-minute timeframe. However, they are also given one rule to follow: that they are not allowed to engage in two of the same tasks consecutively. That is, they are not allowed to proceed to trial B from trial A (or vice versa) of the same task. The number of tasks attempted, number of rule breaks, along with an excess time spent on one task (>271 seconds) is recorded. These variables contribute to an overall profile score out of 4. The examinee was scored based on the number of subtasks attempted, number of rule violations, and duration spent on each (Wilson et al., 1996).

Wisconsin Card Sorting Test (WCST; Kongs et al., 2000)

The WCST is a commonly employed test of executive functions and mental flexibility (Heaton et al., 1993; Rabin et al., 2005). For the purposes of the current study, the 64-card version was employed (Kongs et al., 2000). Vayalakkara, Devaraju-Backhaus, Bradley, Simco, and Golden (2000) demonstrated the high validity of the 64-card version in predicting scores on the full 128-card version, while simultaneously reducing administration time. Participants were shown four stimulus cards, each with distinct patterns with respect to colour, shape, and number it items on the card. They were then asked to place a response card under the stimulus card where they thought it would be correct. They are given simple feedback, "correct" or "incorrect" after each placement of the response card. The instructions are designed to be vague in an effort for the examinee to figure out the protocol as they attempt some trials. After 10 correct responses, the sorting rule was changed (colour, to form, to number) without the participant's awareness. For example, correct placement of the first 10 trials would be response cards placed under stimulus cards with the same colour. The test terminates after 6 sets of 10 correctly placed cards each, or when 64 response cards had been used. Scores were based on the number of trials to reach their first category, errors made, and the number of categories they were able to complete (Heaton et al., 1993).

Trail Making Test (Lezak et al., 2012)

The TMT is a commonly administered measure of attention, psychomotor speed, cognitive flexibility, and divided attention and is also in the public domain (Strauss et al., 2006). This test initially appeared on the Army Individual Test Battery in 1944 and was later modified for the Halstead Battery (Reitan, 1955). The test consists of two parts, part A and part B. Part A, also referred to as TMT A, consists of a paper with encircled numbers, from 1 to 25, arranged on the page seemingly at random. The participant is instructed to draw a continuous line to connect all the circles in sequential order. Part B, also referred to as TMT B, consists of a paper with encircled numbers (1 to 13) and letters (A to L) arranged on the page seemingly at random. Here, the participant is instructed to once again draw a continuous line to connect all the circled numbers and letters in alternating order (i.e., 1, A, 2, B, 3, C…etc.). In both parts, a sample trial was presented prior to test administration to insure adequacy of the instructions. The participant is instructed to complete the task as quickly as they can without lifting their pencil from the paper. When an error was made, the participant was notified and instructed to correct their mistake(s) and to continue until the test was completed (Strauss et al., 2006). The examinee was scored based on the number of errors made and the amount of time elapsed to completion for each part (Reitan, 1979).

Tower of London (TOL). Drexel University: 2nd Edition (Culbertson & Zillmer, 2001)

The TOL is a commonly employed measure of executive functions (Rabin et al., 2005). The stimuli for this test include two sets of the following: wooden board; three vertical pegs (a tall, medium, and short peg) placed at equidistant length on the length of the board; and three coloured beads (blue, green, and red). One set is for the examinee and the other is for the examiner. For each trial, the examinee's board has the bead pre-arranged in the starting position. The examiner's board is the goal board with the beads placed in a specific manner depending on the trial. The goal is for the examinee to move their beads from the starting position to mirror the examiner's board in as few moves as possible albeit with some rules in mind. The first rule is that the examinee is not allowed to place more beads on a peg than it will hold. The second rule is that they may move one bead at a time. There are 10 trials with two practice trails. Each trial is allocated a maximum of two minutes. The scores were based on the initiation, execution, and total time, along with the total correct moves and total moves (Culbertson & Zillmer, 2001).

Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999)

The WASI is a test of cognitive functioning with four subtests that form the performance IQ and verbal IQ. The PIQ and VIQ together produce a full IQ score. The verbal IQ is calculated with the combined scores of Vocabulary and Similarities subtests, whereas the performance IQ is the combined scores of the Block Design and Matrix Reasoning (Wechsler, 1999). For the purposes of the present study, only the Similarities and Matrix Reasoning subtests were considered due to missing data. A brief description of these tests follows.

Similarities

This subset assessed the examinee’s ability to form abstract concepts as well as verbal reasoning. The examinee was presented with pairs of words of increasing conceptual difficulty and asked to relate the similarities of the items presented. For example, all adult examinees started from item #7 (“Grapes - Strawberries”) and ended at item # 26 (“Freedom – Law”; Wechsler, 1997).

Matrix Reasoning

This subtest assessed the examinee’s adaptive capacity, abstract reasoning skills and the ability to negotiate visual information. The examinee was presented with a cluster of pictures or patterns of increasing difficulty and a group of response stimuli. They were asked to choose one response stimuli that they believed to fit with the pattern presented (Wechsler, 1997).

4.5.3 Procedure

As with Study 2, participant demographics such as number of days since date of accident, GCS scores, and age were derived from the patient’s self-report within the context of a comprehensive clinical history and substantiated by way of file review, both undertaken by a board licensed clinical neuropsychologist. Participants were administered the above noted tests by a psychometrist under the supervision of a registered clinical neuropsychologist. This study was conducted using archival data and as such, there exist missing data for various reasons that include time restrictions, compliance, and necessity as per the direction from the supervising neuropsychologist. The examinees were grouped according to employment status (employed or unemployed) as determined by the examinees' self-report on vocational status and when available corroborated by way of a collateral interview with the patient's significant other (e.g., spouse, child, or parent). Employment status was also compared against whether or not participants were receiving income replacement benefits.

4.5.4 Statistical Analysis

Mean performance on the neuropsychological test measures between groups were analyzed on SPSS Version 22.0 for Windows (IBM Corp, 2013). A Bonferroni correction was also performed due to the number of comparisons made ($\alpha = 0.0025$). Cohen's d effect size values were computed for all comparisons.

4.6 Results

There was a significant difference of age between employment groups ($p = 0.004$), with the unemployed group comprising of older participants. No other significant differences were found among any other demographic variables between groups.

The performance of the unemployed group was generally worse than those of the employed group across all of the tests administered. Due to the relatively low group sample sizes (i.e., < 100), the distributions of the dependent variables were examined for normality using Shapiro-Wilk's test for normality. Approximately half of the distributions were found to be non-normal ($p < 0.05$). The non-normal distributions were log transformed but this did not result in reliable parametric distributions for all variables. Thus, non-parametric inferential statistics were employed. Mann-Whitney U comparisons were conducted between employment groups on all neuropsychological test measures and Cohen's d was used to calculate effect sizes. Test performance scores are summarized in Table 7. There were no statistically significant differences between employment groups on TMT errors (both parts), WCST, TOL, Similarities, and Action Program ($p < 0.05$). Statistically significant differences were observed between groups (with positive values indicating better performance by the employed group) on the TMT A ($U = 667.0$, $n_1 = 30$, $n_2 = 70$, $p = 0.004$), TMT B ($U = 696.5.0$, $n_1 = 30$, $n_2 = 69$, $p = 0.01$), Matrix Reasoning ($U = 665.0$, $n_1 = 29$, $n_2 = 69$, $p = 0.009$), Rule Shift Cards ($U = 195.0$, $n_1 = 18$, $n_2 = 37$, $p = 0.005$), Zoo Map ($U = 159.0$, $n_1 = 15$, $n_2 = 34$, $p = 0.028$), Key Search ($U = 203.5$, $n_1 = 17$, $n_2 = 37$, $p = 0.024$), and Modified Six Elements ($U = 216.5$, $n_1 = 19$, $n_2 = 37$, $p = 0.01$). However, after Bonferroni correction for the number of comparisons used was employed (α = 0.0025), no statistically significant differences were observed on any measure.

Due to the archival nature of the data, control over administration could not be implemented and thus, sample sizes fluctuated across measures. To this end, the smallest sample size comparisons were observed on the TOL and BADS subtests in the employed group. Since effect sizes are not dependent nor influenced by sample size (Zakzanis, 1998), a brief summary of the effect size findings will follow to supplement the non-parametric comparisons. Small effect size estimates were observed on all measures from the WCST (d = |0.02| to |0.38|), Similarities (d = 0.25), and Matrix Reasoning (d = 0.36). Small to medium effect sizes were observed on all measures from the TOL (d = |0.06| to |0.54|), with medium effect sizes for all timed-based scores (see Table 7). Small effect sizes were found on the errors from the TMT but medium effect sizes on the overall time (d ≈ |0.60|). Medium to large effect sizes were observed across all BADS subtests with the largest found in the Modified Six Elements (d = 0.77) and Rule Shift Cards (d = 0.86).

Table 7.

Performance on Neuropsychological Tests stratified by employment status.

	Employed			Unemployed					
	n	Mean	SD	*n*	Mean	SD	*p*	*p corrected*	*d*
Trail Making Test									
Trial A (s)	30	34.7	11.44	70	48.4	26.56	0.004	ns	-0.60
Trial A Errors	30	0.3	0.69	70	0.3	0.60	0.389	ns	-0.06
Trial B (s)	30	76.4	28.80	69	112.6	75.00	0.01	ns	-0.56
Trial B Errors	30	0.4	0.68	69	0.7	1.27	0.495	ns	-0.25
Wisconsin Card Sorting Test									
Errors	29	17.1	10.29	65	20.8	10.74	0.075	ns	-0.34
Categories completed	30	3.5	1.43	65	2.9	1.57	0.091	ns	0.38
Trials to 1st category	29	14.9	11.80	65	20.6	17.72	0.062	ns	-0.35
Failure to maintain set	30	0.3	0.48	65	0.3	0.54	0.882	ns	-0.02
Tower of London									
Correct score	18	3.7	2.49	45	3.9	2.67	0.927	ns	-0.06
Move score	18	34.7	16.99	45	36.5	24.39	0.819	ns	-0.08
Initiation time (s)	18	56.6	27.05	45	75.2	49.55	0.148	ns	-0.42
Execution time (s)	18	203.6	49.21	44	268.8	153.55	0.328	ns	-0.49
Total time (s)	18	260.2	44.94	44	343.6	180.62	0.226	ns	-0.54
WASI									

Similarities	26	37.2	7.94	59	35.2	7.67	0.205	ns	0.25
Matrix Reasoning	29	25.3	7.28	69	21.6	11.43	0.009	ns	0.36
BADS									
RSC	18	3.8	0.38	37	3.2	0.82	0.005	ns	0.86
Zoo Map	15	3.2	1.08	34	2.7	0.84	0.028	ns	0.53
Action Program	9	3.7	0.50	26	3.2	1.07	0.25	ns	0.46
Key Search	17	3.7	0.61	37	2.7	1.49	0.024	ns	0.75
MSE	19	3.7	0.48	37	3.2	0.70	0.01	ns	0.77

Abbreviations: BADS (Behavioural Assessment of Dysexecutive Syndrome); TEA (Test of Everyday Attention); MSE (Modified Six Elements); RSC (Rule Shift Cards); WASI (Wechsler Abbreviated Scale of Intelligence).

4.7 Discussion

The present study compared the predictive ability of an ecologically oriented battery (the BADS) with traditional pen-and-paper neuropsychological test measures with respect to RTW outcome in patients with mTBI. As a follow-up, the present study also examined which test could best differentiate between employment status. The results indicated that following correction of family-wise error via Bonferroni correction factor, neither the BADS nor traditional tests could differentiate employment status in patients with mTBI who were in the post-acute period of recovery.

Previous studies have reported the Rule Shift Cards, Modified Six Elements, and Action Program subtests to demonstrate good ecological validity as indexed by the significant correlations that these subtests had with informant ratings of the DEX (Bennett et al., 2005; Burgess et al., 1998; Evans et al., 1997; Norris & Tate, 2000). It is important to note that there is a difference in outcome measures used to index ecological validity; the DEX was in these studies whereas employment was status was used in the current study. Keeping this difference in mind, the significance reported in the present study ran counter to those of previous studies as none of these subtests were shown to differentiate between employment groups. Likewise, the Zoo Map subtest had been found to be significantly correlated not only to clinician ratings but also other informant ratings of patient functioning on the DEX (Norris & Tate, 2000). However, Wood and Liossi (2006) reported that they were also unable to find any significant correlations between performance on the Zoo Map subtest and informant ratings on the DEX. The same study also reported poor ecological validity for the Key Search subtest – a finding that was echoed in the present study.

As noted by Zakzanis (1998; 2001), it would be poor science to equate, and thus conclude that, a lack of significant findings automatically suggests that there were no effects. Zakzanis (2001) pointed to the importance of considering effect sizes not in their absolute value, but in the context in which these values were to be interpreted. Though differences in performance between the employed and unemployed groups did not reach significance for the BADS following the Bonferroni correction, it is noteworthy to highlight that the effect sizes for this battery as a whole tended to be larger in relation to those of other tests. Here, and as shown in Table 7 the Action Program (d = 0.46), and the Zoo Map subtests (d = 0.53) both demonstrated moderate effect sizes. In addition, the Key Search (d = 0.75), Modified Six Elements (d = 0.77), and Rule Shift Cards (d = 0.86) demonstrated large effect sizes. In light of this, the trend toward larger effect sizes for the BADS may be indicative of the fact that the BADS was more sensitive to employment status in comparison to the other neuropsychological test measures.

Medium effect sizes were observed on TMT A and B and a large effect size was found on the failure to maintain set metric from the WCST. Studies examining the ecological validity of the TMT and the WCST have produced mixed results. Consistent with some studies (e.g., Fortin, Godbout & Braun, 2003; Norris & Tate, 2000) reporting poor ecological validity of both the TMT and WCST, the present study also found that neither versions of the Trail Making Test were good predictors of disability. However, Bennett and colleagues (2005) reported that version B of the TMT as well as the percentage of correct response and perseverative errors on the WCST significantly correlated with informant ratings on the DEX.

Finally, findings for the TOL generally appear to be consistent with previous literature elucidating the limited utility of the test in being able to differentiate between persons with head injuries and healthy controls (Cockburn, 1995). To this end, only the timed scores (i.e., initiation, execution, and total time) demonstrated moderate effect size estimates ($d = |0.42|$ to $|0.54|$), whereas the correct score and move score yielded very small estimates ($d = |0.06|$ to $|0.08|$). Unexpectedly, the unemployed group produced slightly better scores on average than the employed group on the correct score and move score measures. This difference, however, was not close to being statistically significant. It is worth noting that to the best of our knowledge, no research has been conducted on the utility of the TOL in specifically predicting employment status in a population of post-acute patients with mTBI.

The finding that performance between the unemployed and employed groups failed to be significantly different on any measure could be explained by several factors. First, Chaytor and Schmitter-Edgecombe (2003) suggested that one reason why some neuropsychological test measures may fail to demonstrate ecological validity was that patients with TBIs may employ the use of compensatory strategies in testing environments whereas they would not do so in the real-world. In a follow-up study, Chaytor, Schmitter-Edgecombe and Burr (2006) reported that the degree to which a neuropsychological test demonstrated ecological validity was in part mediated by the amount of compensatory strategies utilized by the examinees. Therefore, accounting for these strategies in an assessment could result in a more ecologically valid assessment. In context of the present study, it is plausible that patients with TBIs were making a conscious effort to keep track of their failed attempts and the alterations made at each attempt to complete the task, whereas they would not have done so in the real-world.

Second, these authors along with others (Levine, Dawson, Boutet, Schwartz, & Struss, 2000; Norris & Tate, 2000) raised concerns regarding the disconnect that exists between the testing environment and the environment in which everyday living occurs. While testing environments are generally structured, occurring in quiet environments with minimal distractions, the workplace is rarely ever free of external factors that could easily impinge on an individual’s ability to sustain attention, demonstrate cognitive flexibility and so forth. In a litigating sample, the motivations for intentionally magnifying error may be greater given the incentive of secondary benefits.

The ability to RTW is undoubtedly an essential part of being able to resume pre-injury way of life (Machamer et al., 2005), as it enables patients with mTBI to access, by way of providing financial stability, many forms of services to aid in the process of reintegration. Being able to properly assess a patient’s ability to RTW also enables various agencies such as insurance companies to properly allocate financial resources and assistance to those who require it. Thus, it is crucial that neuropsychological test measures are able to properly predict a patient’s ability to RTW in the real-world.

There have been few suggestions as to why neuropsychological test measures lack ecological validity. It may be that these tests are framed by a firm set of structures and rules, with the examiner often prompting responses from the participants. However, in reality, this is not the case as real-life decisions are made in ambiguous situations that lack structure (Levine et al., 2000; Norris & Tate, 2000). Wilson (1993) argued that the problem may be that these tests are not detailed enough to provide adequate information regarding the functional abilities of patients and the problems they will have to face resulting from their cognitive deficits.

A second limitation of neuropsychological test measures as argued by Burgess and Alderman (1990, as cited by Wilson, Evans, Emslie, Alderman, & Burgess, 1998) and Chaytor and Schmitter-Edgecombe (2003) is that neuropsychological test measures only address a small sample of a wide range of possible behaviours that could be exhibited. Burgess and Alderman (as cited in Wilson et al., 1998) assert that patients with TBIs who suffer from dysexecutive syndromes often have the constituent parts of executive abilities intact. The problem, they suggest, may be in the actual initiation and synchronized use of these constituent abilities in order to carry out coherent actions such as multi-tasking and monitoring ongoing behaviours. For example, an individual may normally have the ability to sustain attention to simple stimuli. However, in situations that also require constant monitoring of one's own behavior, the individual could experience difficulties. Consequently, sampling a small subset of behaviours runs the risk of an over-exaggeration of executive functionality, and an erroneous prediction of real-life capabilities (Chaytor & Schmitter-Edgecombe, 2003).

Finally, and as mentioned in the introduction, many neuropsychological test measures that are commonly employed in a clinical context were originally developed for research purposes (Holt et al. 2011; Manchester et al., 2004). There is an inherent problem with this, as the requirements that must be met for measures developed for research differ from those that are necessary for the purposes of clinical assessments (Burgess et al., 2006). Hence, tests that were developed to elucidate the role of specific executive abilities in certain tasks may prove themselves sufficient for a research setting. However, performances on these measures may not necessarily translate to functionality in the real-world.

4.7.1 Conclusion

The lack of significant findings in the majority of the tests examined highlight the importance of other facets of a complete neuropsychological assessment such as the clinical interview and behavioural observations (Wilson, 1993). Again, as mentioned above, it is important that the relatively larger effect sizes observed for the BADS not be discounted, but rather be considered in clinical settings where time may be limited and only certain tests can be administered to inform clinical opinions of diagnoses and prognoses. To this end, considering the small effect sizes in most of the neuropsychological test measures (see Table 7), there did seem to be a lack of sensitivity in these tests in detecting patients who were experiencing disability in everyday living. In light of this, it may be beneficial for clinicians to work hand-in-hand with other professionals such as occupational therapists when opining on the prognoses of patients who have sustained a mTBI. Burgess and colleagues (1998) suggested a similar idea, stating that it may be fitting to assess patients in a variety of settings – namely those in which they experience difficulties. Furthermore, researchers may benefit from investigating other forms of assessment that could prove to be more ecologically valid (i.e. virtual reality testing; e.g., Campbell et al., 2009; Tippett et al., 2008).

4.7.2 Limitations and Future Directions

Limitations include the dichotomy of employment outcome. Stratification of RTW based on level (i.e., full-time, part-time, modified duties/hours) and type of employment (e.g., cognitively demanding, physical labour intense, etc.) would improve the specificity of the results. Second, this study only assessed the ecological validity of neuropsychological test measures as it pertains to patients' employment status. To this end, the findings cannot be generalized to suggest that these tests are completely limited in their ecological validity – additional studies examining other realms of everyday living such as cooking and home maintenance may provide different results (see Zakzanis & Grimes, 2016). Third, sample sizes across each comparison were inconsistent. Sizes were especially small in the employed group who completed the BADS. Due to the archival nature of the data, there was missing data due to time restrictions, compliance, and necessity as per the direction from the supervising neuropsychologist. Future studies aiming to replicate or validate the current findings should utilize a cross-sectional study design.

Given the lack of statistical relationship between the WCST and TOL performance with RTW in the current study, more research is required to fully validate the clinical utility of these tests in predicting employment status in the real-world. Future studies could examine the clinical utility of the neuropsychological test measures in predicting the disability in other facets of everyday living such as cooking and home maintenance. Other research endeavors could investigate whether the ecological validity of these neuropsychological test measures would be influenced by the different types of occupations examined. It may be advisable for researchers to develop specific neuropsychological test measures tailored to specific occupational groups to better enhance the ecological validity of neuropsychological assessments. Studies pursuing such endeavors are crucial in fully understanding the clinical utility of neuropsychological test measures as it pertains to real-world disability prediction in patients with mTBI. Lastly, virtual reality testing is a burgeoning field that may yield measures that outperform existing neuropsychological test measures in the prediction of disability (Campbell et al., 2009; Tippett et al., 2008).

Chapter 5

5 Study 4: The Ecological Validity of Two Novel Virtual Reality Tests of Attention and Executive Functions

Within the realm of mTBI research, it has been argued that current test measures of attention are not sufficiently sensitive to assess the various aspects of attention involved in everyday life. Moreover, these test measures were not initially designed to answer questions regarding real-world functioning but rather to help ascertain the relationship between brain and behavior and hereafter, measure strengths and weaknesses as they relate to cognition. The adaptation of traditional pen-and-paper neuropsychological test measures can be used to accurately predict real-world functioning would be akin to using a map to estimate a country's gross national product. This is inappropriate, and a new resource would be needed. Logically, it would seem more appropriate to develop new neuropsychological test measures that can measure cognitive strengths and weaknesses and predict functional outcome. It was proposed that some attentional problems might only arise in situations that resemble real-world environments that are less structured, more complex, and are examined over a longer period of time (Sloan & Ponsford, 1995). The limited ecological validity of traditional pen-and-paper neuropsychological test measures has been acknowledged in the research literature and a call to order has been put out to researchers to address this issue (Marcotte & Grant, 2010; Sbordone & Purisch, 1996).

With the limits of traditional pen-and-paper neuropsychological test measures in mind, some researchers have sought to develop batteries that are ecologically oriented. That is, they were originally developed with the goal of performance being predictive of functional outcome whilst framing the tests to resemble real-world activities. The BADS is an ecologically oriented battery that was developed to assess executive dysfunction. Similarly, the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994) is an ecologically oriented battery that was developed to assess attention. The TEA was developed in keeping with Posner and Peterson's (1990) model of attention where the process of attention consists of three systems; *selection*, *vigilance*, and *orientation*. Robertson et al. (1994) conducted a principal component analysis on the subtests of the TEA. The model revealed factors that were congruent with Posner and Peterson's model: *visual selective attention/speed*, *attentional switching*, *sustained attention*, along with *auditory-verbal working memory*. The entire TEA battery is based on a fictional trip to Philadelphia and contains ecologically relevant tests such as searching through maps, telephone directories, and listening to the lottery ticket broadcast. The battery consists of eight subtests: Map Search, Elevator Counting, Elevator Counting with Distraction, Visual Elevator, Elevator Counting with Reversal, Telephone Search, Telephone Search While Counting, and Lottery. In the Map Search subtest, the examinee must look through a map of Philadelphia and mark certain symbols (i.e., gas station symbol or restaurant symbol) within two-minutes. In the Elevator Counting subtest, the examinee is told that they are to pretend that they are on the ground floor of a hotel and are about to use the elevator. However, the elevator's visual floor indicator is not functioning. Accordingly, the examinee is instructed to listen to a pre-recorded tape recording of beeps that signify an elevator passing a floor. The examinee is asked to count the number of beeps and inform the examiner which floor they are on based on the number of beeps. There are variants of the elevator subtest. For example, in the Elevator Counting with Distraction, there are high and low tones, which the examinee must pay attention to one type of tone in order to determine which floor they land on. In the Elevator Counting with Reversal subtest, the examinee is presented with a recording that indicates when the elevator goes up a floor and when the elevator goes down a floor. The examinee is to keep track and once again indicate which floor they land on. In the Visual Elevator subtest, the examinee is presented with a Visual Stimulus that indicates if the elevator went up or down a floor. In the Telephone Search subtest, the examinee is instructed to look for key symbols on pages of a simulated telephone directory. The Telephone Search While Counting subtest is similar to the Telephone Search subtest but the examinee is also presented with an audio recording where they are to listen and count the number of tones they hear. In the Lottery subtest, the examinee is instructed to listen for their 'winning number' from a 10-minute pre-recorded audiotape.

The authors of the TEA argue that this test measure gives measurements of clinical and theoretical aspects of attention; has three versions and can be administered to a patient on three separate occasions with parallel materials; is able to identify different patterns of attention; and is able to be used on a wide range of patient populations. In addition, there is evidence that the TEA battery is able to discriminate between patients with minimal and mild Alzheimer's disease (Green, Hodges, & Baddeley, 1995). In another study, a regression analysis revealed that the TEA was better at predicting general functional outcome than traditional tests of memory and attention in a sample of patients with multiple sclerosis (Higginson, Arnett, & Voss, 2000). Furthermore, the authors present findings that suggest that the TEA is able to detect normal age effects in attentional skills in a sample of healthy controls (Robertson et al., 1994). In addition, the authors argue that the TEA's use of everyday materials resemble real-life scenarios. In other words, they argue that the TEA is high in verisimilitude.

Despite the reported advantages of the TEA, there have been relatively few studies looking at its clinical utility in a mTBI population. Bate and colleagues (2001) aimed to examine the TEA's clinical utility as a test of attention that could be used to differentiate between patients with TBI and controls. Thirty-five patients who had sustained a severe TBI were compared to 35 age and education matched controls. The TEA along with other test measures of attention (i.e., Modified Colour-Word subtest of the Stroop, SDMT, Digit Span, RSAT, and the Paced Auditory Serial Addition Test) were administered. A logistic regression analysis revealed a model of best fit that included the Map Search ($\beta = 0.63$, Wald $\chi^2 = 6.2$, $p = 0.013$) followed by the Modified Colour-Word subtest of the Stroop ($\beta = 0.56$, Wald $\chi^2 = 5.6$, $p = 0.018$). When the TBI group was further stratified into early (< 12 months post-injury) and late (> 24 months post-injury) recovery, it was found that those in the early recovery group showed deficits on tests of visual selective attention (Map Search and Telephone Search) and sustained attention (Lottery). Patients in the late recovery group only exhibited deficits on tests of visual selective attention. The authors suggested that this finding indicated potential recovery in attentional function after 12 months post-injury.

Similar results have been shown for patients who were in the acute recovery period following a mTBI. Mathias et al. (2004) administered the Visual Elevator, Telephone Search, and Telephone Search while Counting subtests of the TEA along with tests of fluency (COWAT and Ruff Figural Fluency Test) and memory (Rey Adult Verbal Learning Test) on a group of 40 patients who had sustained a mTBI and a matched control group ($n = 40$). Results indicated a performance deficit on the Visual Elevator (accuracy and time score), Telephone Search (time per target), Ruff Figural Fluency Test, and Rey Adult Verbal Learning Test (total trials 1-5, immediate recall, and delayed recall scores) significantly differentiated the mTBI group from the control group. The authors suggested that deficits in attention, non-verbal fluency, and verbal memory are prevalent one month after a mTBI. Thus, there appears to be some support for the clinical utility of the TEA to be used on a mTBI related population.

While there are some advantages of the TEA as noted above, there are some notable disadvantages. A review of the literature indicates that the TEA children version has gained some enthusiasm, especially as a diagnostic test for ADHD (see Barkley, 1998), whereas the use and utility in an adult TBI population has not. Thus, more basic studies in adult populations of TBI are needed to provide a more accurate estimation of the sensitivity and specificity of the TEA. Moreover, poor psychometric norms and ceiling effects in the profile scores of the TEA have been reported (Chan, Lai, & Robertson, 2006; Strauss et al., 2006). While the raw scores of some tests on the TEA can be transformed into a standard score, others provide a very narrow range of interpretation. For example, the profile score of the Elevator Counting test ranges from 1-7, but the manual interprets a score of 7 as "normal", a score of 6 as "doubtful", and a score of 5 or less as "definitely abnormal" (Robertson et al., 1994, p. 15). Moreover, the sample size of the closed head injury patients used to validate the TEA was only 15. In addition, although the test is purported to be high in ecological validity, many of the characteristics of traditional (non-ecologically valid) tests are present. For example, administration of the TEA requires a highly structured environment. The test design also eliminates all distractions that would normally be found in real life and there is a relatively short sampling window. Moreover, some subtests of the TEA draw on a range of cognitive processes in addition to attention. To this end, visual scanning, motor abilities, and planning are important cognitive domains that factor into test performance on the Map Search and Telephone Search tests.

Overall, the TEA attempts to combine traditional measures of cognition in tasks that resemble everyday living. Evidence from the literature supports their use and utility in detecting the impact that cognitive deficits have on functional ability as compared to traditional pen-and-paper tasks (Marcotte & Grant, 2010). However, these tests do not directly measure behaviour in real-life contexts (Chaytor & Schmitter-Edgecombe, 2003) and research is lacking on how performance on these tests *directly* relate to everyday functioning (Chan et al., 2008; Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999; Robertson & Schmitter-Edgecombe, 2017; Spooner & Pachana, 2006).

5.1 Naturalistic Assessments

Naturalistic assessments can be considered a step forward along the spectrum of ecologically valid test development. Here, participants complete commonplace activities in a realistic environment whilst being observed by a researcher or clinician. Naturalistic assessments have been defined as "observable, rule-based, open-ended tasks completed in an environment that either mimics the real-world or is the real-world" (Robertson & Schmitter-Edgecombe, 2017, p.

18). Since these assessments offer contextually relevant environments, it is purported that task performance is related to functional abilities. Indeed, naturalistic assessments are high in verisimilitude, thus arguing for their potential use as an analogue to real-world functional ability. While traditional pen-and-paper neuropsychological test measures are designed to be completed in a distraction-free, highly structured environment, naturalistic assessments allow for distractions such as environment-typical noise and movement from other individuals. An example of a type of naturalistic assessment with the most research interest is the Multiple Errands Test (Shallice & Burgess, 1991).

In the Multiple Errands Test, participants are instructed to complete three sets of tasks in an actual shopping centre. The instructions range from purchasing products, to meeting an appointment time, to looking up information (i.e., current exchange rates for currency). The participants are also told to complete the instructions while following a few rules, such as limiting the amount of money to be spent and not going into room or store unless they were going to complete a purchase. In the original study, the authors recruited three high IQ patients who had sustained traumatic injuries with subsequent damage to their prefrontal lobes. All patients were tested on traditional pen-and-paper neuropsychological test measures and their performance was within normal limits on tests of perception, language, and intelligence tests. Moreover, two of the three patients also performed within normal limits on tests of “frontal lobe function” (p. 731) such as the TOL and the TMT. However, when tested on the Multiple Errands Test, all patients performed poorly as compared to controls, took longer with each individual task, and incurred more total errors than the controls. The authors concluded that the Multiple Errands Test was more sensitive to functional impairment as compared to traditional pen-and-paper neuropsychological test measures. Their study demonstrated the potential clinical utility of the Multiple Errands Test, especially for high functioning individuals who perform within normal limits on traditional pen-and-paper neuropsychological testing.

In a follow up study, Alderman and colleagues (2003) developed a simplified version of the Multiple Errands Test which differed from the original in three ways; (1) concrete and clearer rules; (2) simplified task demands; and (3) the use of an instructions and record sheet by the participant. Their purpose was to determine the utility of the simplified version of the Multiple Errands Test on a broad range of clinical populations. Namely, they recruited a mixed sample of inpatients and outpatients (n = 50) who were in neurorehabilitation due to TBI, stroke, and cerebral tumors along with a group of neurologically healthy participants to serve as controls ($n = 46$). Their main results were that performance on the simplified version of the Multiple Errands Test was able to discriminate between the two groups with medium to large effect sizes ($d = 0.38$ to 0.84). In addition, the neurological group had more total number of errors, made more types of errors and, were not able to complete as many tasks as the control group. Similar to the findings reported by Shallice and Burgess (1991), they found that the majority of their neurological participants were able to perform within normal limits on traditional pen-and-paper neuropsychological test measures. As for the ecological validity of the task, performance on the simplified version of the Multiple Errands Test was associated with scores on the DEX (Burgess, Alderman, Wilson, Evans, & Emslie, 1996); a behaviour rating scale assessing behavioural manifestations of executive impairment in every day functioning. Finally, the authors reported that their results support the clinical utility of the simplified version of the Multiple Errands Test with good sensitivity in high functioning individuals with executive dysfunction.

Adaptations of the Multiple Errands Test have been developed to site-specific utility with iterations adapted for use in hospitals and malls (Cuberos-Urbano et al., 2013; Dawson et al., 2009; Maeir, Krauss, & Katz, 2010; Morrison et al., 2013; Tranel, Hathaway-Nepple, & Anderson, 2007). In addition, there is some evidence to suggest that these variations are sensitive measures of real-world functioning and cognitive dysfunction (see Dawson et al., 2009; Robertson & Schmitter-Edgecombe, 2017).

While naturalistic assessments certainly offer many advantages over traditional pen-and-paper neuropsychological test measures with respect to ecological validity, there are numerous limitations that are inherent. From a logistics perspective, naturalistic tasks are executed in the natural environment. Accordingly, the tasks, environment, and distractors are unique to each specific site. To be used in a new facility, the naturalistic task would potentially need to be modified and psychometrically re-validated for the new environment. Furthermore, due to the lack of control of the external environment, naturalistic assessments have questionable test-retest reliability. Each administration of the task would bring with it a unique set of distractors (e.g., number of people in the facility, noise level, signs displayed, peak versus off-peak hours, etc.). Naturalistic tasks are also quite time and resource intense. The majority of naturalistic tasks were reported to be 20 min – 120 minutes in length (Robertson & Schmitter-Edgecombe, 2017) and required two examiners to administer in addition to the use of recording devices. In addition, there are safety and mobility concerns limiting its use (Rand, Basha-Abu Rukan, Weiss, & Katz, 2009). From a methodological standpoint, there are some concerns with prior research in naturalistic assessments. Namely, with the exception of the Multiple Errands Test, many of the naturalistic assessments were introduced as preliminary and with small sample sizes using broadly defined experimental groups (Robertson & Schmitter-Edgecombe, 2017). The majority of the literature is also focused on the use of naturalistic tests as a rehabilitation tool rather than an assessment tool. This implies a functional difference in the use and clinical utility of the task (i.e., charting progress over time vs snapshot of functioning). Moreover, while the above review indicates that performance on naturalistic assessments may be sensitive to real-world functional ability, little research has been conducted to determine the specificity to certain neurological disorders/diseases. Certainly, little to no research has been undertaken to determine the relationship between performance on naturalistic assessments and RTW or other functional outcome measures specifically in patients with mTBI.

In an effort to improve RTW outcome, naturalistic assessments with vocational themes have been developed. Such examples include office and classroom-like environments where participants are instructed to complete clerical work tasks (Wolf et al., 2008), review stock inventory, make reservations to a restaurant (Lamberts et al., 2010), calculate fines (Novakovic-Agopian et al., 2014), and learn and retain information in a classroom-like setting (MacLennan & MacLennan, 2008). Robertson & Schmitter-Edgecombe (2017) reviewed these vocational tasks with respect to functional outcome. They found that there is some evidence to support the relationship between performance on vocational tasks and executive functioning. However, they noted that many of these studies were preliminary and that the findings need to be replicated and validated. In addition, they noted that a direct relationship between performance on vocational environments and everyday functioning was lacking. Moreover, most of these studies utilized an experimental group with mixed neurological etiologies and none consisted of a strictly mTBI sample.

5.2 Virtual Reality

Naturalistic assessments are highly reflective of real-world environments in that they take place in real environments, but the major limitation is that they lack psychometric strength and overall experimental control (i.e., stimulus delivery, response inhibition, control of distractors, etc.). Here, the use of VR in assessments may potentially overcome these limitations. VR is defined as “an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer in which one’s actions partially determine what happens in the environment” (Virtual reality, n.d.). In VR, individuals are partially immersed into a simulated three-dimensional VE where they can interact and manipulate aspects of that environment in real time (Rose, Attree, Brooks, & Johnson, 1998; Schultheis et al., 2002). There are a number of advantages in the implementation of VR and VEs for neuropsychological assessment. The most relevant of which is the potential to simulate real-world situations thereby allowing for assessments high in verisimilitude. This can be achieved as complete control is given to the test developer with respect to modeling and controlling the interactive environment, including but not limited to the setting, stimulus delivery, distractors, timing, and measurement of all inputs. Psychometrically, VR and associated peripherals allows for multiple different types of data to be captured easily, precisely, and simultaneously (i.e., movement data, eye-gaze fixation; van der Ham, Faber, Venselaar, Kreveld, Loffler, 2015). To this end, the types of data that can be captured is limited to the advancements in sensory technology, which is already quite broad and growing (Rizzo et al., 2004). VR allows for seemingly an unlimited number of customizations to meet the task complexity, action-reaction behaviours, and feedback requirements for nearly any type of situation or patient-type. Equally, VR allows for the assessment of quite possibly unlimited number of expansive virtual worlds, all confined with in the limited space of the laboratory setting. Furthermore, behaviours which may be physically impossible to carry out in the real-world are possible in VEs. For example, one is able to teleport from one area to another in an effort to reduce unnecessary travel time. Relative to naturalistic assessments, it is less resource intensive as all that is required is a computer system along with one trained staff member as opposed to an entire physical environment with multiple staff members. Once an environment has been developed, it allows for the infinite number of administrations with perfect inter-rater reliability in a cost-effective manner (Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001). Another advantage is that VR may provide access to assessment and rehabilitation for those with transportation or mobility issues (i.e., telerehabilitation; see Larson, Feignon, Gagliardo, & Dvorkin, 2014) and reduce the impact of safety concerns (Weiss, Bialik, & Kizony, 2003). VEs have been found to be more engaging and motivating as compared to traditional pen-and-paper neuropsychological test measures which is important as there is an established link between motivation, effort, and successful completion of rehabilitation (Campbell et al., 2009; Rizzo et al., 2004). A further advantage is that VEs also provide an increased approximation of the natural environment and the subject is able to use multiple cognitive domains while completing a task as they would in the real-world. In other words, since neuropsychological functioning in the real-world is multi-factorial, VEs can be tailored to parallel these environments and to ultimately increase verisimilitude and in turn ecological validity. In addition, VEs can be developed to particularly measure those cognitive constructs that have historically been difficult to measure using traditional pen-and-paper tasks (i.e., executive and attentional functions; Lezak et al., 2012).

5.3 Virtual Reality is the Future of Neuropsychology

It has been argued that the tools used in neuropsychology have lagged behind other clinical neurosciences (see Miller & Barr, 2017) and that the use of technology, specifically computers, is long overdue. Furthermore, as per Parsons & Kane (2017), the implementation of technology in clinical neuropsychology is inevitable to keep up with the advancements in other areas of the clinical neurosciences. Indeed, Bilder (2011) noted in his *state of neuropsychological assessment* that "neuropsychology needs to embrace computerized assessment" (pg. 12). Similarly, this sentiment was echoed in the presidential address at the 2011 annual conference of the National Academy of Neuropsychology wherein it was stated that VR is one of the emerging areas in the future of neuropsychology. Clinicians also agree that VR is the future of intervention. In a poll published in the journal *Professional Psychology: Research and Practice*, VR was ranked to be the 3rd and 5th most likely intervention that would increase over the next decade (Norcross, Hedges, & Prochaska, 2002). Lastly, the test publishing industry also has interests in this emerging field. *The Psychological Corporation*, the largest publisher of standardized psychological assessment test measures, began conducting research and development for VR in 2004 (DMW, 2004). Therefore, it appears that VR and computerized assessment have been universally accepted as the next evolutionary step in neuropsychology.

Accordingly, researchers have contributed to the literature with newly developed VEs for use in neuropsychology and related fields. For example, VEs have been developed with the intention to measure component cognitive processes such as memory (Canty et al., 2014; Sauzon et al., 2016), attention (Pollak et al., 2009; Rizzo et al., 2000), spatial abilities (Beck et al., 2010; Serino, Morganti, Di Stefano, & Riva, 2015), and executive functions (Cipresso et al., 2014; Raspelli et al., 2011; Robinson & Brewer, 2016). Moreover, virtual versions of existing traditional tests have been developed such as the virtual block design (Jeon, Clamann, Kaber, &

Currie, 2013), virtual Tower of Hanoi (Robinson & Brewer, 2016), and virtual WCST (Elkind et al., 2001). Furthermore, VEs have been developed to assess functional outcome for iADLs such as cooking (Manera et al., 2015), mobility (Lengenfelder et al., 2002), shopping (Carelli, Morganit, Weiss, Kizony, & Riva, 2008; Josman, Klinger, & Kizony, 2008; Klinger, Chemin, Lebreton, & Marie, 2006), attending school (Rizzo et al., 2002), working in an office (Jansari et al., 2014; McGeorge et al., 2001), and navigating specific environments like cities (Brown, Kerr, & Bayon, 1998; Campbell et al., 2009; Costas, Carvalho, & de Aragon, 2000) and crossing the street (Titov & Knight, 2005). With respect to RTW, VEs have been developed to train individuals in different situations such as driving (Mahoney, 1997), diving (Froehlich, 1997), sky diving (Hue, Delannay, & Berland, 1997), acting as a first responder (Bliss, Tidwell, & Guest, 1997), service duty (Knerr, Breaux, Goldberg, & Thurman, 2002; Stone, 2002), and mock surgery (Gallagher, McClure, McGuigan, Crothers, & Browning, 1999).

As noted by Rose, Brooks, and Rizzo (2005) *"*the use of VR has the potential to present some of these neuropsychological test measures in a more ecologically valid way*"* (p. 244). One of the earliest studies aimed at developing a more ecologically valid version of a traditional pen-and-paper neuropsychological test was developed by Pugnetti et al., (1995). These researchers designed a VR-analogue of the WCST. The VE consisted of 32 rooms and connecting doorways. The objective was to find the exit by going through different doorways. The doorways would either lead to another room or would lead to a dead end. The doorways were of variable shapes and colours serving as cues, analogous to colours, forms, and number of objects as the category cues for the WCST. Here, matching criteria was changed every seven consecutively correct doorway passages. Pugnetti and colleagues (1998) administered this task along with the pen-and-paper version of the WCST to a sample of 34 with mixed neurological pathologies along with 29 control subjects. Their results indicated that the number of categories achieved and total number of errors on the WCST and the VR task were able to differentiate between the two samples. The total number of correct responses on the WCST was able to differentiate between the two samples, whereas the total number of correct responses on the VR task was not. The authors concluded that the VR test had similar psychometric properties as the WCST and could distinguish patients from controls just as well as the traditional neuropsychological test measures could. One advantage of the VR task was that significant differences were observed between the two groups after the first category, whereas differences were only observed after the fourth category achievement on the WCST. The authors attributed this to the more cognitively demanding and multi-domain oriented VR task as compared to the pen-and-paper test. Their findings, however, were limited by their lack of homogeneity of the sample and low sample sizes. Despite these limitations, they provided interesting preliminary evidence that a VR task could be developed to resemble a traditional pen-and-paper neuropsychological test with promising psychometric results.

In a more recent experiment, Elkind et al., (2001) developed a different virtual version of the WCST called Look For A Match. Here, participants are in a virtual beach and their objective is to deliver one of four items to sunbathers (sodas, popsicles, beach balls, or frisbees) sitting under different umbrellas which are distinguished by their colours, numbers, or type of umbrella. After each delivery, a verbal feedback is given indicating a correct or incorrect response. The authors modeled the matching pattern after the WCST in that the first match was to colour, the second match was to type of umbrella, and the third match was to number. Similar to the WCST, category shifts occur every 10 trials. The task terminated after six successful category changes or 128-trials were administered, which ever was first. These authors administered the VR task along with the WCST to 62 healthy controls to determine the concurrent validity of the Look For A Match test with respect to executive functions. Results indicated that subscores of the WCST correlated significantly to their Look For A Match counterpart (i.e., total number of errors made, non-perseverative errors, conceptual level responses, trails to first category, and the number of failure to maintain sets that occurred). The authors also noted that performance scores indicated that the VR task was more difficult as compared to the WCST, which was consistent with 60% of the participants self-reporting that the VR task was more difficult than the WCST (though not a significant finding). Moreover, it was reported that the majority of the participants (77%) found the VR task more enjoyable. The authors concluded that the Look For A Match measured the same cognitive constructs as the WCST, was more difficult, but was more enjoyable. As a result, the authors suggested that the "Look For A Match may be a more useful measure of executive functions…and might prove to be more ecologically valid than the WCST" (Elkind et al., 2001, p. 495).

5.4 Present Study

Based on the above review, there appear to be many potential benefits of the utilization of VR and VEs in neuropsychological assessment. Most importantly, the literature indicates that VR test measures that resemble real-life situations have good ecological validity. To date, however, there have been no studies using VR to validate RTW in patients with mTBI. Given the interest in vocational outcome by way of referral sources (e.g., employers, insurers), the objective of the current study was to develop effective and valid psychometric tools that are able to predict vocational outcome. Accordingly, we set out to build upon previous work using VR by developing novel VR assessment tools and investigating their predictive ability as it pertains to RTW in a sample of patients in the post-acute recovery period of mTBI.

Two separate tasks were developed: the Office Task and the Conveyor Belt Task (CBT). Based on the work from Pugnetti and colleagues (1998) and Elkind et al. (2001), the Office Task was designed as a VR-analogue of the WCST measuring aspects of executive functioning. The premise behind the Office Task was to reflect sorting mail in an office environment. The CBT was designed to mimic assembly line work measuring aspects of attention. The premise behind the CBT was that the participant was a quality control manager at an assembly plant and had to review various products for deficiencies as they came down the belt. Components of attention were tested as sub-trials with various difficulty trials in each.

The aims of the current study were as follows:

1. Investigate the concurrent validity of the VR tasks as compared to currently utilized tests of attention and executive functioning.

2. Compare performance between traditional pen-and-paper neuropsychological test measures, ecologically oriented tests (i.e., BADS and TEA), and the VR tasks between patients with mTBI who were employed or unemployed.

3. To determine the ecological validity of traditional pen-and-paper neuropsychological test measures, sub-measures from the BADS and TEA, and the VR tests in an effort to identify predictors of employment status following mTBI.

An overarching objective of this study was to improve ecological validity of neuropsychological assessment tools in patients with mTBI who were in the post-acute period of recovery using RTW status as the functional outcome.

Based on previous work on VR and ecological validity, the following hypotheses are proposed:

1. Both VR test measures will show a statistical relationship with their traditional pen-and-paper neuropsychological test counterparts illustrating concurrent validity. More specifically, performance on the Office Task will be positively related to the WCST and performance on the CBT modules will be related to traditional tests of attention.

2. Based on the review of the literature, test performance between those who are employed and unemployed will be indistinguishable on traditional pen-and-paper neuropsychological test measures; performance between the two groups will be statistically different on the Modified Six Elements and Map Search; and performance between the two groups will be statistically different on the VR test measures.

3. Based on an ecological validity spectrum, performance on the VR test measures will be better predictors of employment status than ecologically oriented tests, which will be better than traditional pen-and-paper based neuropsychological test measures.

5.5 Method

5.5.1 Participants

The sample consisted of adults who had sustained a mTBI and were recruited from a private practice neuropsychology clinic. All participants were referred for a neuropsychological examination by way of their attorney, family physician, or treating neurologist or by way of their insurer secondary to having been involved in a motor vehicle accident for the purpose of insurer benefit entitlement. The criteria for a mTBI was operationally defined in keeping with the definition set by ACRM (Kay et al., 1993): (1) GCS scores at the time of injury between 13 and 15, or in the instance where a GCS was not made available in the medical documentation, a subjective report of feeling disoriented or confused immediately after the impact but without LOC; (2) duration of LOC of less than 30 minutes; (3) a period of posttraumatic amnesia that was less than 24 hours; and (4) the absence of positive neuroimaging findings. Severity of injury was assessed by way of available medical records. All participants met a minimum of grade 8 English reading level determined by their performance on the WRAT4-Reading (Wilkinson & Robertson, 2006).

At the time of the assessment, all participants received a comprehensive clinical evaluation from a licensed clinical neuropsychologist that included but was not limited to an intake interview, standardized neuropsychological battery, psychological questionnaires, and VR test measures. Since the participants were drawn from a litigating sample, evaluation of performance validity was necessary. Accordingly, the inclusion criteria for all participants with respect to credibility of responding was as follows: a score of 45 or higher on Trial 2 or the Retention trial of the Test of Memory Malingering (Tombaugh, 1996); 9/15 or higher on both the Rey Fifteen Item Test (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002) and Recognition follow-up; and less than the cut-off *T*-score on the four validity indices on the Personality Assessment Inventory (PAI; Morey, 1991). With respect to the latter, the specific *T*-score cut-off scores were derived from the manual and were <72 on the inconsistency scale (ICN); <74 on the infrequency scale (INF); <91 on the negative impression management (NIM) scale; and <68 on the positive impression management (PIM) scale. Moreover, the PAI was administered to exclude any potential psychopathological co-morbidities that has been shown to effect cognitive performance (see Zakzanis, 1999).

Initially, 57 participants were recruited for this study, 53 of which consented to participate. Three additional participants were removed from the analysis as they did not meet PAI validity inclusion criteria. The final analysis included 50 participants, of which 23 had returned to some form of gainful employment (i.e., full-time, part-time, modified duties/hours) and 27 who had not. Demographic information along with GCS scores, years of education, pre-morbid FSIQ estimates, days since injury, and handedness of the samples are presented in Table 8. Overall, there were no significant demographic differences ($p > 0.05$) between the two groups. The average age of the employed group was 42.1 years ($SD = 13.14$), which was similar to the unemployed group ($M = 46.1$ years; $SD = 15.35$). Distribution of gender was relatively equal (43.5% female in employed group; 55.6% in the unemployed group) as was handedness at roughly 96% (right) in both groups. There were no significant differences between the two groups with respect to days since injury, GCS scores, or FSIQ estimates[10]. All participants

[10] Based on the North American Adult Reading Test (Blair & Spreen, 1989).

provided informed written consent. The current study was conducted in accordance with human ethics standards and research ethics board approval from the University of Toronto.

Table 8.

Demographic characteristics of the sample (Study 4).

	Employed (n = 23)		Unemployed (n = 27)	
	Mean	SD	Mean	SD
Age (years)	42.1	13.14	46.1	15.35
Gender (% female)	43.50%		55.60%	
Education (years)	14.6	3.71	13.9	3.72
FSIQ Estimate*	108.1	7.01	104	7.16
GCS	14.7	1.09	14.9	0.37
Days since injury	674.8	454.63	734.5	413.99
Handedness (% Right)	95.70%		96.30%	

* Based on the North American Adult Reading Test
Abbreviations: FSIQ (Full Scale Intelligence Quotient); GCS (Glasgow Coma Scale).

5.5.2 Measures

Various test measures have been used to assess executive functions and the components of attention. The test measures selected in the current study were based on prior results in addition to the underlying abilities that are assessed by each of the ecologically valid and VR tests.

The traditional pen-and-paper neuropsychological test measures employed in the current study were chosen due similar face validity as with the ecologically valid and virtual reality tests. To be more clinically applicable, the tests were chosen from a list of commonly employed tests of attention and executive functions (see Rabin et al., 2005).

Brief descriptions of each of the test measures administered in this study are detailed below. For more detailed descriptions, the reader is advised to refer to Lezak and colleagues (2012), Strauss and colleagues (2006), along with the respective test manuals. Task instructions for the VR tasks (Office Task and CBT) can be found in the Appendix. Moreover, Table 9 presents all the test measures employed in the study stratified by way of their respective component of attention or executive functioning.

Table 9.

Components of Attention and Executive Functions Stratified by Measure.

Domain	Component	Traditional Measure	Ecological Measure	VR Measure
Executive Functions	Cognitive flexibility	TMT B	Rule Shift Cards	Office Task
	Planning, problem solving, organizing, and monitoring behaviour	WCST	Modified Six Elements	
	Planning	TOL		
Attention	Selective Attention	SDMT & RSAT	Map Search	CBT Selective Attention
	Sustained Attention	RSAT & TMT A	Elevator Counting	CBT Sustained Attention
	Divided Attention	TMT B		CBT Divided Attention
	Attention Control	RSAT	Map Search & Rule Shift Cards	CBT Attention Shift

Note: Some measures have overlapping domain and component characteristics. CBT = Conveyor Belt Task; RSAT = Ruff 2 & 7 Selective Attention Test; SDMT = Symbol Digit Modalities Test; TMT = Trail Making Test (A or B); TOL = Tower of London; and WCST = Wisconsin Card Sorting Test.

In order to assess the credibility of responding put forth on the neuropsychological test measures, participants completed the following.

Test of Memory Malingering (Tombaugh, 1996)

The Test of Memory Malingering is a brief visual memory test used to discriminate between those patients with real memory impairments and those malingering. The Test of Memory Malingering contains three trials (1, 2, and retention) and has a learning and recall component for the first two trials. In trial 1, 50 line drawings are presented one at a time, for three-seconds each with one-second intervals, to the examinee who is told to memorize the drawings to their best ability. The examinee is then presented with 50 pages with two line-drawn items on the page, one at the top and one at the bottom. One of the two items is identical to the ones presented in the learning trial. The examinee is asked to identify the one that they saw previously. The examiner provides feedback to the examinee after each selection by stating “correct” or “incorrect”. The administration and learning items for trial 2 are identical to trail 1 but with the items presented in a mixed order. The retention trial is administered following a 15 to 20-minute delay after completion of the first two trials. All participants from both samples produced scores that were credible and above the cut-off score of ≥45 on trial 2 or the retention trial as per the test manual.

Rey Fifteen-Item Test (Boone et al., 2002)

The Rey Fifteen Item Test was developed by Andre Rey and formally described by Lezak and colleagues (2012). This measure is a recall task in which 15 items arranged in 5 rows (3 items per row) are presented for 10 seconds. Immediately following the presentation of the stimuli, the examinee is asked to reproduce all 15 items from memory. Examinees who accurately reproduced fewer than 9 of the items were deemed to be suspected of non-credible test findings (see Lezak et al., 2012). The recognition trial consists of the original 15 items along with 15 distractor items, for a total of 30 items. The examinee is asked to circle all of the items that were on the initial administration (Boone et al., 2002). As per Boone and colleagues (2002), the combined recall and recognition score was calculated using the following formula: free recall score + (recognition hits – recognition false positives). A passing combined score was greater than or equal to 18 which was found to have increased sensitivity (71%) and higher specificity (92%) for non-credible test findings than the administration of just the Rey Fifteen Item Test recall alone (Boone et al., 2002). All participants in the current study scored higher than 18 on the Rey Fifteen Item Test with recognition.

Personality Assessment Inventory (PAI; Morey, 1991)

The PAI is a self-administered personality and psychopathology instrument that yields a broad range of clinically relevant information. The PAI contains 344-items which constitute 22 non-overlapping scales:11 clinical scales; 5 treatment scales; 2 interpersonal scales; and 4 validity scales. The validity scales are labeled ICN, INF, PIM, and NIM. The ICN scale reflects the reliability of consent similar responses, with high scores on this scale suggestive of inconsistent responding. The INF scale reflects the degree of random responding or responses in an atypical manner, with high scores suggestive of inappropriate attention given to the items on the test. The NIM scale includes test items taken collectively to indicate exaggerated unfavorable impression or extremely bizarre and unlikely symptoms. Elevated scores on the NIM scale is suggestive that the respondent attempted to portray themselves in an especially negative manner (Morey, 1991). In clinical practice and in the research literature, the NIM scale is often interpreted to indicate symptom exaggeration and possible malingering. In the current study, all examinees did not exceed the threshold score for the 4 validity scales.

Wide Range Achievement Test – Fourth Edition – Reading (WRAT4-Reading; Wilkinson & Robertson, 2006)

The WRAT4-Reading was administered to all participants to ensure that they met the minimum English reading ability to complete the PAI questionnaire. As per the manual, a stimulus sheet with 15 letters along with 55 irregularly pronounced words were presented to participants. Participants were asked to read aloud, annunciating and pronouncing the irregular words as well as they could. All participants met the minimum English reading level (Grade 4) to complete the PAI.

North American Adult Reading Test (NAART; Blair & Spreen, 1989)

In order to control for pre-morbid IQ differences between the two groups, the NAART was administered. This is an adapted version of the National Adult Reading Test (NART; Nelson, 1982), which was originally developed for the UK population, modified using Canadian and American pronunciation rules. This test requires the participant to pronounce and annunciate 61 phonetically irregular words. The error scores are then converted to estimate FSIQ scores using the conversion tables provided by Blair and Spreen (1989).

In order to assess attention, the following tests were administered.

Ruff 2 & 7 Selective Attention Task (RSAT; Ruff & Allen, 1996)

The RSAT served as a measure of visual attention, both sustained and selective (Lezak et al., 2012). The test consists of 20 trial administered in consecutive 15-second intervals. Each trial consists of three rows of mixed numbers and capital letters spanning the width of the page. During the trials, the participant is instructed to cross out all 2’s and 7’s that appear in the trail from left to right, starting from the top row. At the end of each time interval, they are instructed by the examiner to move onto the next trial. This is repeated until completion of all 20 trials. The duration of the RSAT is 5 minutes long. Raw scores for the RSAT were obtained by counting all the correct items in each trial and calculating the total speed and total accuracy of the scores as per the RSAT manual (Ruff & Allen, 1996). The RSAT was chosen due to its sensitivity to TBI populations (Bate et al., 2001).

Symbol Digit Modalities Test (SDMT; Smith, 1973)

The SDMT is symbol-substitution paper and pencil test that measures selective and sustained visual attention in addition to visual scanning and psychomotor speed (Lezak et al., 2012; Shum et al., 1990). The SDMT requires the examinee to substitute rows of geometric symbols for numbers using a key which is located at the top of the page. The test is 90 seconds long and performance is measured by the total number of correctly substituted items. Errors are also measured but due to the low frequency of the number of errors in the sample, errors were excluded from the analysis. This test was chosen due to research support in discriminating between TBI and healthy control participants (Bate el al., 2001; Ponsford & Kinsella, 1992). square completed correctly on each trial.

Trail Making Test (TMT; Reitan, 1955)

The TMT is a commonly administered measure of attention, psychomotor speed, cognitive flexibility, and divided attention and is also in the public domain (Strauss et al., 2006). This test initially appeared on the Army Individual Test Battery in 1944 and was later modified for the Halstead Battery (Reitan, 1955). The test consists of two parts, part A and part B. Part A, also referred to as TMT A, consists of a paper with encircled numbers, from 1 to 25, arranged on the page seemingly at random. The participant is instructed to draw a continuous line to connect all the circles in sequential order. Part B, also referred to as TMT B, consists of a paper with encircled numbers (1 to 13) and letters (A to L) arranged on the page seemingly at random. Here, the participant is instructed to once again draw a continuous line to connect all the circled numbers and letters in alternating order (i.e., 1, A, 2, B, 3, C…etc.). In both parts, a sample trial was presented prior to test administration to insure adequacy of the instructions. The participant is instructed to complete the task as quickly as they can without lifting their pencil from the paper. When an error was made, the participant was notified and instructed to correct their mistake(s) and to continue until the test was completed (Strauss et al., 2006). Completion time and the number of mistakes were recorded. However, due to the infrequent occurrence of errors made on the TMT by the participants, only the completion time was utilized in the analyses. In the current study, the TMT was included as it is a commonly utilized standardized measure of sustained attention (TMT A), divided attention (TMT B), along with executive functioning by way of mental flexibility (TMT B). Moreover, this test has been shown to be a sensitive measure in mTBI samples (Lange et al., 2005).

Map Search. Test of Everyday Attention (TEA; Robertson et al., 1994; 1996)

The Map Search is a subtest from the TEA battery that is a measure of selective attention (Robertson et al., 1996). Here, participants are presented with a colour map of the Philadelphia area with various symbols indicative of landmarks (e.g., rest stops, restaurants, gas stations). Participants are asked to imagine that they are on a road trip in the area and to circle all specified landmarks in case they wish to make a stopover. The specified landmark varies based on which of the three versions of the test are administered. For the purposes of this study, all participants were administered Version A which featured the knife and fork symbol representing restaurants. Participants were initially given a blue dry-erase marker to circle all landmarks. This was switched with a red dry-erase marker at 60seconds. The trial lasted 2 minutes in duration. The purpose of the colours were to track the number of items circled in the first and second minute of the trial. Total correct circled items in the first minute and the entire trial are converted into a scaled score as per the manual. For the purposes of the current study, the total correct score over the entire trial were used in the analysis. This test was chosen as it is a purported ecologically valid test that conceptually resembles the selective attention trail of the CBT. Moreover, the Map Search has been shown in the literature to be the best TEA subtest to distinguish between a TBI population and healthy controls (Bate et a., 2001).

Elevator Counting. Test of Everyday Attention (TEA; Robertson, Ward, Ridgway, & Nimmo-Smith, 1994, 1996)

The Elevator Counting is a subtest from the TEA battery that is a measure of sustained attention (Robertson et al., 1996). Originally, it was developed by Wilkins, Shallice, and McCarthy (1987) as a validated measure of sustained attention and adapted into the TEA by Broks and colleagues (1988). Participants are instructed to pretend that they are in an elevator where the visual floor indicator is malfunctioning. They are presented with a series of tones from a tape recorder and told that each tone represents the elevator arriving at a different floor. Accordingly, participants are to listen to all tones and establish which floor they have arrived at. This subtest was included as it was conceptually similar to the sustained attention trial of the CBT.

In order to assess executive functions, the following tests were administered.

Tower of London (TOL). Drexel University: 2nd Edition (Culbertson & Zillmer, 2001)

The TOL is a commonly employed measure of executive functions (Rabin et al., 2005). The stimuli for this test include two sets of the following: wooden board; three vertical pegs (a tall, medium, and short peg) placed at equidistant length on the length of the board; and three coloured beads (blue, green, and red). One set is for the examinee and the other is for the examiner. For each trial, the examinee's board has the bead pre-arranged in the starting position. The examiner's board is the goal board with the beads placed in a specific manner depending on the trial. The goal is for the examinee to move their beads from the starting position to mirror the examiner's board in as few moves as possible albeit with some rules in mind. The first rule is that the examinee is not allowed to place more beads on a peg than it will hold. The second rule is that they may move one bead at a time. There are 10 trials with two practice trails. Each trial is allocated a maximum of 2 minutes. Scores that were derived from this measure include total correct score (where the number of moves equates to the minimum moves for that trial), total move score, total initiation time, total execution time, total trial time (a combination of initiation and trail time), total time violations (greater than 1 minute), and number of rule violations. Due to the infrequency of time and rule violations, these were not included in the analysis. The TOL was included in the current study due to conceptual problem-solving similarities with the Office Task.

Wisconsin Card Sorting Test (WCST; Kongs et al., 2000)

The WCST is a commonly employed test of executive functions and mental flexibility (Heaton et al., 1993; Rabin et al., 2005). For the purposes of the current study, the 64-card version was employed (Kongs et al., 2000). Vayalakkara et al. (2000) demonstrated the high validity of the 64-card version in predicting scores on the full 128-card version, while simultaneously reducing administration time. Participants were shown four stimulus cards, each with distinct patterns with respect to colour, shape, and number it items on the card. They were then asked to place a response card under the stimulus card where they thought it would be correct. They are given simple feedback, "correct" or "incorrect" after each placement of the response card. The instructions are designed to be vague in an effort for the examinee to figure out the protocol as they attempt some trials. After 10 correct responses, the sorting rule was changed (colour, to form, to number) without the participant's awareness. For example, correct placement of the first 10 trials would be response cards placed under stimulus cards with the same colour. The test terminates after 6 sets of 10 correctly placed cards each, or when 64 response cards had been used. Scores were based on the number of correct response card placements, number of errors, number of perseverations, number of perseverative errors, conceptual level responses, number of categories completed, trials to first category, and number of failure to maintain sets (see Heaton et a., 1993). This test was chosen as it was deemed to be conceptually, and at face-value, similar to the Office Task.

Rule Shift Cards. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

The Rule Shift Cards is a subtest from the BADS battery that assesses cognitive flexibility and set shifting. There are two trials in this subtest, both utilizing 20 playing cards presented on a flip book. In the first, which is considered the practice trial, the participant is instructed to respond "yes" to seeing red playing cards and "no" to black cards. On the second, which is considered the scored trial, the rules are switched to "yes" if the card was shown was the same colour as the preceding card, and "no" if it was different. For both trials, the rules were typed out and placed in view of the examinee throughout the test. The scores were based on the time taken to respond to trial 2, the number of errors made, time violations (if overall time was >67 seconds), and the profile score as per the manual (Wilson et al., 1996). Since time violations were infrequent in the entire sample, which led to a floor effect, this was removed from the analysis. This subtest was included as it examined similar components of executive functions as the Office Task and the other traditional and executive test measures employed in the study, thereby allowing for conceptual comparisons to be made.

Modified Six Elements Test. Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)

The Modified Six Elements is a subtest from the BADS battery that assesses planning and multitasking abilities. Here, participants are to complete three different tasks within 10-minutes. The first task is to dictate an event, the second task is arithmetic problems, and the third task involves naming line drawn pictures. Each task has two trials labeled 'A' and 'B'. As per the instructions, the examinees are told that they are to attempt to do at least some of all six subtasks within the 10-minute timeframe. However, they are also given one rule to follow: that they are not allowed to engage in two of the same tasks consecutively. That is, they are not allowed to proceed to trial B from trial A (or vice versa) of the same task. The number of tasks attempted, number of rule breaks, along with an excess time spent on one task (>271 seconds) is recorded. These variables contribute to an overall profile score out of 4. The Modified Six Elements was chosen due to its sensitivity to executive dysfunction (Gouveia, Brucki, Malheiros, & Beuno, 2007; Rochat et al. 2009) and similarities to the Office Task with respect to planning and multitasking.

The following VR test measures were employed.

The VR test measures were developed with the broad, long-term intention of potential wide scale usage in healthcare facilities. As such, they were developed to be administered on standard computer systems that would be commonly found in healthcare facilities. Moreover, and for these reasons, the decision to develop a non-immersive (i.e., without the use of a head-mount display) VE was made. The VR platform was developed using the Unreal Tournament 2003 3D engine and platform (Epic Games, 2002). It has been tailored to assess selective, sustained, divided attention, along with a divided attention with a distractor through the CBT and planning and set-shifting via the Office Task. The tasks were executed on a Windows-based PC with the following minimum requirements: Intel Pentium® Dual-Core CPU with at least 2.20GHz processing power, 4 Gigabytes of RAM, 20 GB free disk place, High Definition Intel Graphics 4400. The tasks were displayed on a 15-inch colour LCD monitor with a first-person perspective. Navigation was done through the use of a keyboard for the CBT and a gamepad for the Office Task (Logitech GamePad, Model F310, Logitech Inc.). A brief description of each task will follow.

Office Task

The Office Task was loosely inspired by the WCST and designed to elicit aspects of executive functioning such as reasoning and problem solving in a real-world scenario. To this end, participants are immersed in a VE of an office building and were instructed to deliver various packages to four presented doors. The four doors that packages can be delivered to were all labeled with different signs identifying the business inside. From left to right, the rooms were labeled with the following information: “401-410 The Doctor’s Office”, “411-420 Riverdale

Florists”, “421-430 Shutterbugs Photography”, and “431-440 A Touch of Class Catering”. The packages, however, were designed to have limited information on them. Packages were labeled with either the appropriate room number, printed paraphernalia associated with one of the four rooms, or the exact sign/logo from the door was presented on them. Participants were briefed with instructions that they were a courier that needed to deliver packages that were not always clearly marked (full instructions can be found in the Appendix). Participants were told to use the limited information on the packages to deliver them to the correct door by using their reasoning and problem-solving skills.

At the start of the test, participants were spawned in the elevator of the office floor. Following the verbal recitation of the instructions and confirmation that they understood, they were given the gamepad and instructed on basic navigation through the environment. Following 1-2 minutes of orientation, they were instructed to open the elevator doors via the [start] button on the gamepad where they were able to walk into the office floor. Past the elevator doors, the office floor resembled that of a large elevator lobby. The floor contained two posters placed on the wall opposite of the elevators, three windows with sunlight illuminating inwards, four doors with the labels as described above, and a mail trolley full of packages in the middle of the lobby (see Figure 2 for the layout). The participant was instructed to walk over to the trolley, where they would automatically pick up a package (see Figure 3). They were then instructed to deliver each package to one of the four doors that were directly in front of the trolley. The four doors were placed equidistant apart on the length of the wall and had a green and red light above each door. Participants were instructed to navigate to the door where they intended to deliver the package and to press a button on the gamepad to submit their package. After each delivery, a “CORRECT DOOR” in green text or “INCORRECT DOOR” in red text would appear on the screen. Each package delivery constituted a trial and following delivery, they were instructed to obtain another package and continue delivering. A total of 48 trials were administered for the entire task. The task was programmed to record the total number of correct deliveries, the total number of incorrect deliveries, number of perseveration errors (making more than two consecutive incorrect deliveries), and failures to maintain set (making an incorrect delivery following three consecutive correct deliveries). Due to the lack of perseverative errors and failures to maintain set from the entire sample leading to floor effects, these two variables were excluded from the analysis.

Figure 2. Overall Bird's Eye View of the Office Task.

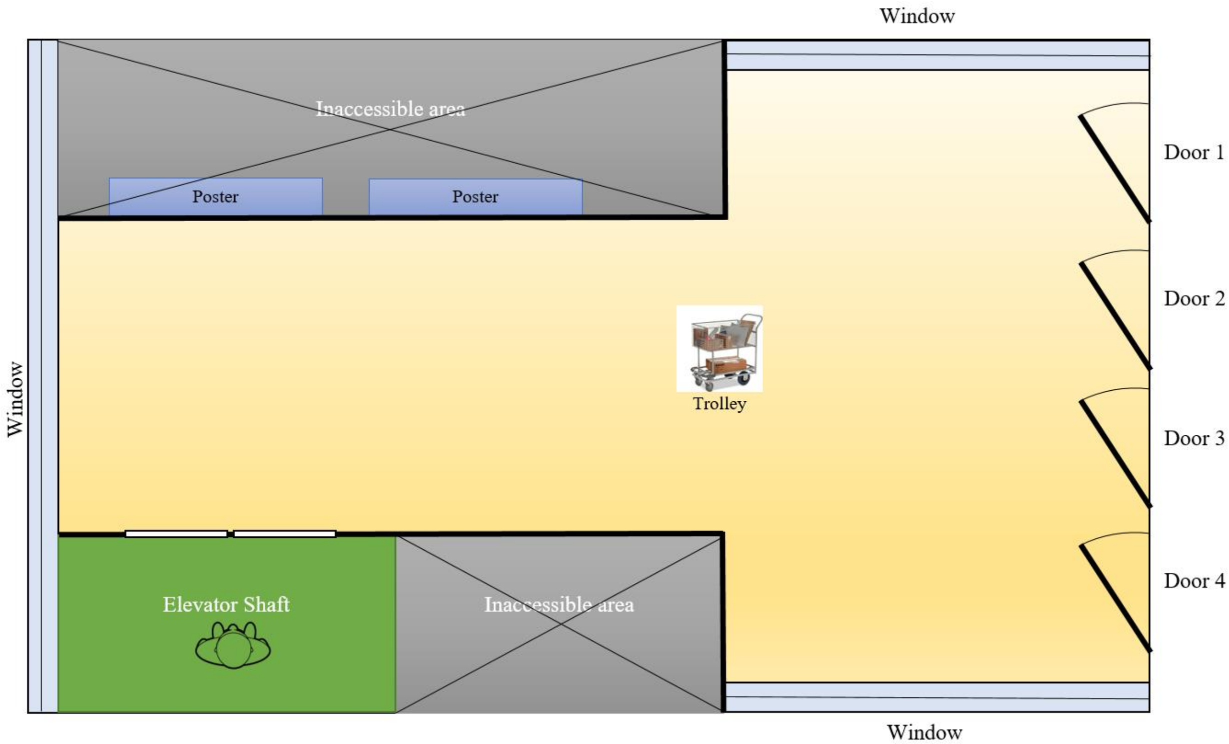


Figure 3. Screenshot of the Office Task with the Trolley and Two of the Four Doors in View.

Conveyor Belt Task (CBT)

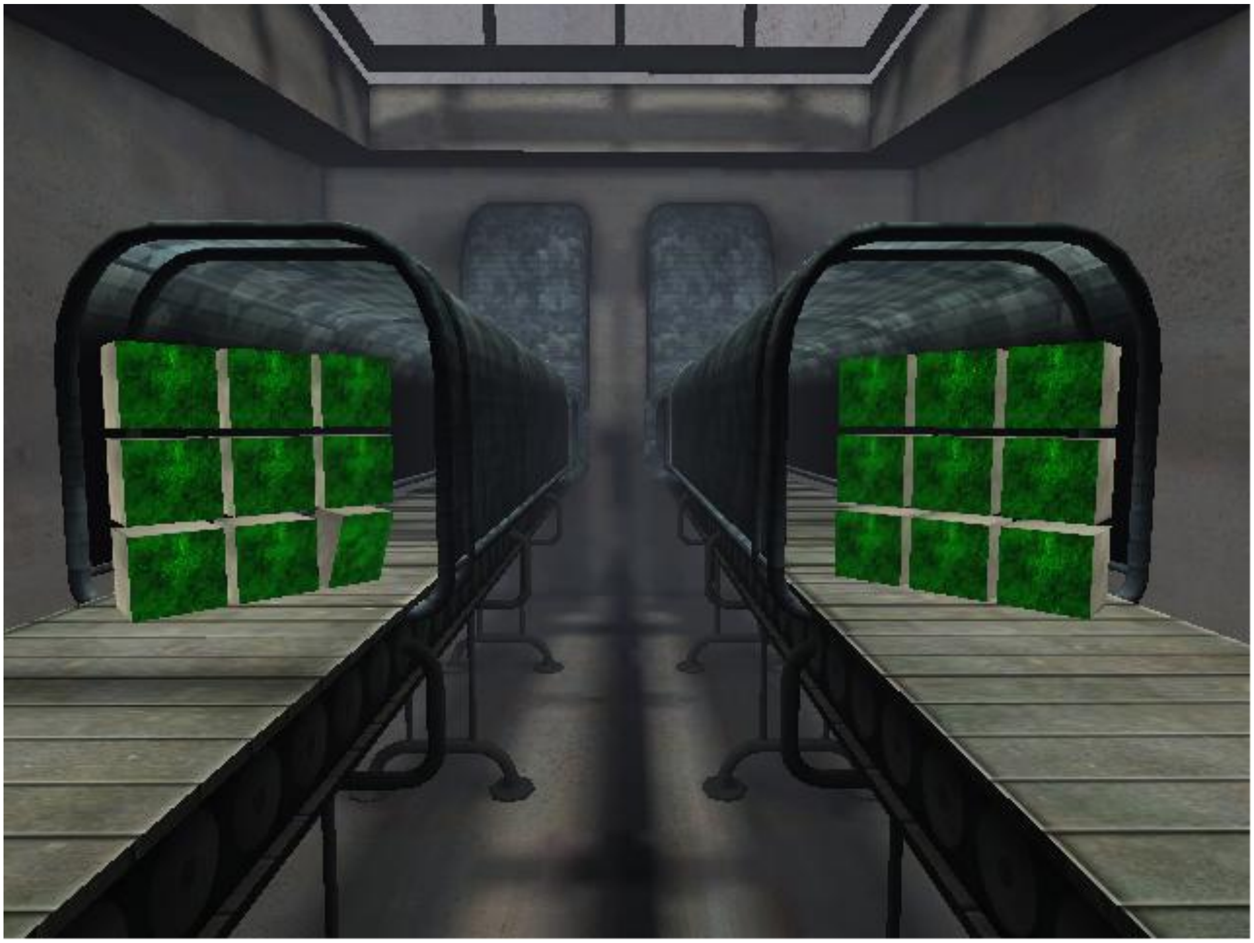
This test was designed to reflect a factory assembly position whereby the participant is asked to sustain, divide (with and without distraction), or selectively use their attention to complete the task. The task was modeled after the RSAT by way of the selective attention and sustained attention trials. Four subtasks testing different components of attention were developed; selective attention, sustained attention, divided attention, and attention shift. With respect to the latter, attention shift was nearly identical to the divided attention trail but with the inclusion of a distractor stimulus (explained below). Each subtask contained several trials of increased difficulty. While the specific instructions and aspects of the VE for each subtask varied, basic commonalities exist. To this end, the trials all began with the participant being spawned into an industrial environment where they were faced with one or two conveyor belts. Using the keyboard and mouse, they were free to peer around but movement around the VE was restricted due to the nature of the stimulus delivery. Moreover, a running score was displayed on the bottom right of the screen as a type of feedback to the examinee. In addition, parameters such as the number of trails, speed of stimulus delivery, number of points at the start, points for correct rewards, points deducted for errors, and frequency were common and customizable across all subtasks. The premise of the CBT is that the participant is factory worker in a toy manufacturing company. Exact instructions recited to each participant for each subtask can be found in the Appendix. These parameters were based on pilot data that indicated that increasing difficulty could be achieved through adjustments of the parameters. Specifically, increasing the speed of stimulus delivery and frequency of the stimulus items reflected an increase in trial difficulty. A brief description of each subtask follows.

Selective Attention. Within the Selective Attention module, the participant is told that they are the quality control specialist at a Tic-Tac-Toe board facility and they need to ensure that all of the boards rolled down the conveyor belt are made out of any shape other than octagonal shapes. A board that contains an octagonal shape (target) is considered a malfunctioned product. They are instructed to press a keyboard key whenever they see a malfunctioned product. For every successful detection, participants receive 100 points; for every false detection they would be deducted 50 points from their score. The Selective Attention module contained four difficulty levels, each progressively increasing the speed of stimulus delivery. The duration for each difficulty level was approximately two-minutes long. A total of 30 trials per difficulty level were presented, 15 of which were target trials whereas the remaining 15 were distractor trials. The total correct trials, number of omissions (false negative), commissions (false positive), and final score in points were extracted for analysis.

Sustained Attention. In the Sustained Attention module, the participant is told that they are the quality control specialist for a toy company and will be instructed to press a keyboard key whenever they witness a malfunctioning item (target) rolling down the belt. As the difficulty levels increased, the malfunctioning item went from large and noticeable to nuanced and specific. In essence, the spotlight of attention was decreased aiding to the increase in difficulty of sustaining attention (see van Zomeren & Brouwer, 1994). In the first level (least difficult), the participants are told that the product-sorting machine is malfunctioning and all items are being rolled down the globe conveyor belt. They are instructed to press a keyboard key when they detect all items that are not globes. In the second level, the participant in instructed that some globes do not have weights installed on the stand causing them to fall over. They are directed to press a keyboard key when they witness a globe that falls over so that it can be reconditioned. In the third level, the participant is instructed that some globes are missing a coat of rust-proof paint. Here, they are directed to press a keyboard key when they witness a globe rusting (turning to red). In the last difficulty level (most difficult), the participant is told that some globes are missing a small weight causing them to rotate slightly. They are directed to press a keyboard key when they witness the globes rotating slightly on the conveyor belt so that it can be tagged for reconditioning. The duration of each difficulty level is approximately two-minutes in length. The correct number of trials, number of omissions, commissions, and total point score were extracted for analysis.

Divided attention. In the Divided Attention module, participants are told that they are placed back as the quality control specialist for the Tic-Tac-Toe conveyor belts. While in this capacity, they will be responsible for two conveyor belts and will be instructed to detect (by pressing one of two keyboard keys) any Tic-Tac-Toe boards that do not have *nine full-squares*. Nine full-square are emphasized as a square may be missing (level one), or replaced by a circle (level two), or replaced by a parallelogram (level three; Figure 4), or an octagon (level four). The duration of each trial was approximately two minutes. The total number of correct trials, number of omissions, commissions, and final points score was extracted for analysis.

Figure 4. Screenshot of the Third Trial of the Divided Attention Module of the Conveyor Belt Task.



Attention shift. In the Attention Shift module, participants are told that they are once again a quality control specialist in charge of the Tic-Tac-Toe boards and need to ensure that all products have nine full-squares on them. In addition, they are told that an error detection arrow that is placed in the middle of the two boards is malfunctioning in that it may or may not detect the conveyor belt that has the malfunctioned item. To this end, they are instructed to ignore the arrow and detect the malfunctioned item on either the left or right conveyor belt by pressing one of two keyboard keys. In essence, the administration of the Attention Shift module is similar to that of the Divided Attention module with the added distractor arrow. The Attention Shift module contains two difficulty levels that vary by speed of presentation. The duration of each difficulty level is approximately two-minutes. Total correct, omissions, commissions, and total points score were extracted for data analysis.

5.5.3 Procedure

Following informed consent, a questionnaire was administered to gather basic demographic information such as age, gender, and handedness. Education was determined via a structured clinical interview with the number of years of formal schooling completed. Employment status was also obtained via self-report during the structured clinical interview and corroborated with their present income replacement benefits status, where available. A neuropsychological battery containing test measures of effort, pre-morbid IQ, attention, executive functions, and the VR tests were administered in a predetermined and counterbalanced order with respect to the VR tests. To this end, half of the study participants were initially administered the VR tests followed by the traditional and ecologically valid tests whereas the VR tests were administered at the end of the battery for the remaining half. This was done in an effort to control for the possible confound of fatigue and order effects. The order of administration of the non-VR tests was consistent for all participants in the following order: Rey Fifteen Item Test → Test of Memory Malingering → NAART → TMT → Modified Six Elements → Rule Shift Cards → Map Search → Elevator Counting → SDMT → TOL → RSAT → WCST → PAI. The entire test battery took approximately 2.5 hours per participant. Each test was administered in a standardized manner by a trained psychometrist under the supervision of a licensed clinical neuropsychologist.

5.5.4 Statistical Analysis

Descriptive and inferential statistics were calculated using SPSS Version 22.0 for Windows (IBM Corp., 2013). Pearson product correlations were calculated between the Office Task and all executive tests (TMT B, TOL, Rule Shift Cards, Modified Six Elements, WCST) and between the CBT Task and all tests of attention (TMT A, RSAT, SDMT, Elevator Counting, Map Search). A repeated measures one-way ANOVA was utilized to determine differences between the various difficulty levels of the selective, sustained, and divided subtests of the CBT across groups. Effect sizes were also calculated in order to articulate the magnitude of differences in means between groups using percent overlap to further describe differences (Wilkinson & The APA Task Force on Statistical Inference, 1999; Zakzanis, 2001). Ecological validity was assessed via a stepwise discriminant functions analysis modeling performance scores that best predict RTW status.

5.6 Results

At the time of injury, all participants (n = 50) were employed. At the time of the assessment and data collection, 23 (46%) were employed in some capacity (full-time, part-time, modified duties/hours) and 27 (54%) were not employed in any capacity. Performance on neuropsychological, ecologically oriented, and VR tests are presented in Table 10 (a-c).

Table 10a.

Performance on Neuropsychological Test Measures Stratified by Employment Status.

	Employed (n = 23)		Unemployed (n = 27)			
	Mean	SD	Mean	SD	p	$\eta 2$
Neuropsychological test measures						
Trail Making Test						
Trial A (s)	31.5	11.65	46.6	21.99	0.003*	0.164
Trial A Errors	0.3	0.45	0.3	0.62	0.644	0.004
Trial B (s)	75.8	43.89	116.9	61.01	0.010*	0.131
Trial B Errors	0.5	0.73	0.5	0.99	0.690	0.003
Ruff 2&7 Selective Attention Test						
Total speed	95.1	22.42	75.7	19.73	0.002*	0.181
Total accuracy	102.2	16.54	90.5	22.36	0.044*	0.082
Symbol Digit Modalities Test	50.0	11.48	39.9	10.89	0.003*	0.174
Wisconsin Card Sorting Test						
Correct trials	47.3	10.40	45.2	12.21	0.532	0.008
Errors	16.7	10.40	18.8	12.21	0.532	0.008
Perseverative responses	10.7	5.52	13.0	5.98	0.165	0.040
Perseverative errors	7.6	3.80	9.4	5.48	0.190	0.360
Non-perseverative errors	7.6	7.61	6.7	5.48	0.631	0.005
Conceptual level responses	40.3	15.97	39.1	16.19	0.794	0.001
Categories completed	3.2	1.75	3.2	1.62	0.981	0.000
Trials to first category	11.0	4.20	16.2	14.39	0.102	0.055
Failure to maintain set	0.4	0.58	0.3	0.73	0.761	0.002
Tower of London						
Correct score	3.9	2.45	3.9	1.75	0.919	0.000
Move score	33.4	18.31	27.3	13.50	0.183	0.037
Initiation time (s)	69.4	28.74	85.1	39.58	0.120	0.050
Execution time (s)	215.0	100.51	237.6	94.38	0.416	0.014

Total time (s)	288.9	108.30	322.8	101.12	0.258	0.027
Time violations	0.7	1.10	1.0	1.16	0.489	0.010
Rule violations	0.1	0.34	0.5	1.16	0.167	0.039

* = $p < 0.05$ (two tailed).

Table 10b.

Performance on Ecologically Oriented Tests Stratified by Employment Status.

	Employed (n = 23)		Unemployed (n = 27)			
	Mean	SD	Mean	SD	p	$\eta2$
Ecologically Valid Tests						
BADS Rule Shift Cards						
Time	37.9	11.98	43.3	10.18	0.088	0.060
Errors	0.7	1.03	1.8	3.37	0.130	0.047
Time violations	0.0	0.00	0.0	0.19	0.361	0.017
Profile score	3.6	0.59	3.3	1.20	0.334	0.019
BADS Modified Six Elements						
Tasks attempted	5.2	1.24	4.8	1.45	0.260	0.026
Rule violations	0.4	0.95	0.4	0.97	0.972	0.000
Raw score	4.5	1.59	3.9	1.49	0.212	0.032
Time violations	0.4	0.89	0.3	0.47	0.632	0.005
Profile score	3.0	1.50	2.8	1.33	0.706	0.003
TEA Map Search Raw	67.5	10.89	54.8	16.89	0.003*	0.167
TEA Elevator Counting Raw	6.7	0.76	6.0	1.47	0.046*	0.080

Abbreviations: BADS (Behavioural Assessment of Dysexecutive Syndrome); TEA (Test of Everyday Attention). * = $p < 0.05$ (two-tailed).

Table 10c.

Performance on Virtual Reality Tests Stratified by Employment Status.

		Employed (*n* = 23)		Unemployed (*n* = 27)			
		Mean	SD	Mean	SD	*p*	*η2*
VR Tasks							
Office Task							
Correct deliveries		41.4	13.14	46.2	4.48	0.101	0.055
Incorrect deliveries		1.3	1.46	0.9	1.60	0.373	0.017
Failure to maintain set		0.9	0.92	0.7	1.21	0.539	0.008
Perseverations		0.1	0.39	0.1	0.46	0.638	0.005
Conveyor Belt Task Selective Attention							
Trial 1	Correct	13.1	2.47	13.3	2.25	0.755	0.002
	Omissions	1.9	2.47	1.7	2.25	0.755	0.002
	Commissions	0.2	0.65	0.3	0.66	0.647	0.004
	Final score	1204.4	374.44	1231.5	343.39	0.791	0.001
Trial 2	Correct	13.4	1.85	11.7	3.50	0.048*	0.079
	Omissions	1.6	1.85	3.3	3.50	0.048*	0.079
	Commissions	1.2	1.47	3.0	3.74	0.033*	0.091
	Final score	1200.0	295.04	852.2	535.20	0.008*	0.138
Trial 3	Correct	12.2	2.52	10.0	4.45	0.043*	0.083
	Omissions	2.8	2.52	5.0	4.45	0.043*	0.083
	Commissions	0.3	0.56	0.4	0.93	0.645	0.004
	Final score	1067.4	383.35	735.2	657.18	0.038*	0.087
Trial 4	Correct	9.5	2.87	8.7	4.64	0.447	0.012
	Omissions	5.4	2.96	6.3	4.64	0.385	0.016
	Commissions	6.7	4.63	8.4	4.93	0.214	0.032
	Final score	347.8	299.04	161.1	500.06	0.124	0.049
Conveyor Belt Task Sustained Attention							
Trial 1	Correct	8.4	2.54	7.9	2.01	0.438	0.013
	Omissions	1.6	2.54	2.1	2.01	0.438	0.013

	Commissions	1.9	2.38	1.9	2.22	0.931	0.000
	Final score	665.2	496.68	590.7	390.52	0.556	0.007
Trial 2	Correct	8.7	1.39	8.2	1.99	0.238	0.029
	Omissions	0.0	0.00	0.0	0.00	N/A	N/A
	Commissions	0.8	0.90	1.9	2.28	0.034*	0.091
	Final score	834.8	176.09	720.4	280.54	0.097	0.056
Trial 3	Correct	8.9	1.52	8.0	1.87	0.094	0.057
	Omissions	0.0	0.00	0.0	0.00	N/A	N/A
	Commissions	0.5	0.67	1.9	2.04	0.003*	0.173
	Final score	863.0	171.37	709.3	268.91	0.022*	0.104
Trial 4	Correct	6.7	1.70	5.7	2.45	0.111	0.052
	Omissions	0.0	0.00	0.0	0.00	N/A	N/A
	Commissions	0.9	1.33	1.5	2.74	0.305	0.022
	Final score	623.9	199.36	490.7	259.45	0.050	0.077
Conveyor Belt Task Divided Attention							
Trial 1	Correct	27.0	4.36	21.4	8.49	0.007*	0.143
	Omissions	2.8	4.19	7.3	7.52	0.015*	0.118
	Commissions	0.2	0.39	1.3	5.36	0.307	0.022
	Final score	2543.5	654.58	1709.3	1274.00	0.007*	0.143
Trial 2	Correct	22.4	6.87	14.9	8.38	0.001*	0.196
	Omissions	7.2	6.91	13.6	8.06	0.004*	0.159
	Commissions	0.4	0.66	1.5	4.22	0.229	0.030
	Final score	1852.2	1030.94	727.8	1257.54	0.001*	0.196
Conveyor Belt Task Attention Shift							
Trial 1	Correct	25.1	6.33	22.2	8.26	0.175	0.038
	Omissions	4.8	6.34	5.9	6.83	0.558	0.007
	Commissions	0.0	0.00	1.9	5.95	0.142	0.044
	Final score	2269.6	949.01	1835.2	1238.32	0.176	0.038
Trial 2	Correct	26.1	5.14	16.6	8.58	<0.001*	0.311
	Omissions	3.4	4.89	11.4	8.14	<0.001*	0.258
	Commissions	0.4	0.49	2.0	4.95	0.128	0.048

Final score	2413.0	771.39	983.3	1286.84	<0.001*	0.311

Abbreviations: VR (Virtual Reality). * = $p < 0.05$ (two-tailed).

5.6.1 Correlations

As per Table 9, Pearson product-moment bivariate correlation analysis was conducted on the entire sample in an effort to determine concurrent validity between the CBT modules and traditional and ecological neuropsychological test measures. The results are organized in Table 11 and Table 12 (a-d) and significant findings discussed by module below.

Table 11.

Correlations Between the Office Task and Measures of Executive Functions.

	Office Task			
Executive Functions Tests	Correct Deliveries	Incorrect Deliveries	Failure to Maintain Set	Perseverations
WSCT				
Correct	-0.054	0.038	-0.023	0.053
Errors	0.054	-0.038	0.023	-0.053
Perseverative Responses	0.102	-0.038	0.027	-0.122
Perseverative Errors	0.128	-0.031	0.021	-0.078
Nonperseverative Errors	-0.104	-0.118	-0.111	-0.041
Categories Completed	-0.048	-0.022	-0.076	0.032
Conceptual Level Responses	-0.052	0.073	0.001	0.069
Trials to First Category	0.031	0.020	0.065	-0.035
Failure to Maintain Set	0.020	0.052	0.084	0.015
TOL				
Correct	-0.163	-0.229	-0.204	-0.175
Move Score	0.250	0.141	0.124	0.121
Initiation Time	-0.057	-0.171	-0.099	-0.219
Execution Time	-0.04	0.121	0.127	0.034
Total Time	-0.057	0.108	0.146	0.002
TMT				
A	-0.049	0.078	0.071	0.046
A Errors	0.181	-0.005	-0.021	0.045
B	-0.174	0.160	0.148	0.161
B Errors	0.054	0.267	0.356*	0.113
Rule Shift Cards				
Time	-0.147	0.246	0.268	0.191
Errors	-0.055	0.375**	0.378**	0.126

Time Violations	0.047	-0.008	0.022	-0.035
Profile Score	0.028	-0.437**	-0.450**	-0.207
Modified Six Elements				
No. Tasks Attempted	-0.031	-0.019	-0.016	0.040
Rule Violations	0.057	0.271	0.153	0.402**
Score	-0.088	-0.162	-0.069	-0.155
Time Violations	0.105	-0.124	-0.148	-0.121
Profile Score	-0.101	-0.084	-0.002	-0.085

Abbreviations: TMT = Trail Making Test, TOL = Tower of London, & WCST = Wisconsin Card Sorting Test

** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).

Selective Attention

A Pearson product-moment bivariate correlation analysis was conducted between the correct score on the four CBT Selective Attention difficulty levels and the SDMT, RSAT, and Map Search. There was a significant positive correlation between RSAT total speed (r = 0.335, n = 50, p = 0.017, two-tailed) and total accuracy (r = 0.528, n = 50, p < 0.0001, two-tailed) with difficulty level 3 of the CBT Selective Attention module. Moreover, significant positive relationships were found between SDMT and difficulty level 1 (r = 0.489, n = 50, p < 0.0005, two-tailed), difficulty level 2 (r = 0.329, n = 50, p = 0.02, two-tailed), and difficulty level 3 (r = 0.525, n = 50, p < 0.0001, two-tailed). As for the Map Search, positive correlations were found between difficulty level 1 (r = 0.323, n = 50, p = 0.02, two-tailed) and difficulty level 3 (r = 0.665, n = 50, p < 0.0001, two-tailed). To this end, difficulty level 3 of the Selective Attention module and Map Search had the strongest relationship relative to all other correlations in this module.

Table 12a.

*Correlations Between the Selective Attention Module of the CBT and Measures of Attention***.**

	Difficulty Levels			
Attention Measures	1	2	3	4
RSAT				
Total Speed	0.183	0.24	0.335*	0.078
Total Accuracy	0.231	0.269	0.528**	0.162
SDMT	0.489**	0.329*	0.525**	0.074

Map Search	0.323*	0.266	0.665**	0.061

Abbreviations: RSAT = Ruff 2 & 7 Selective Attention Task; SDMT = Symbol Digit Modalities Test
** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).

Sustained Attention

A Pearson product-moment bivariate correlation analysis was conducted between the correct score on the four CBT Sustained Attention difficulty levels and TMT A, RSAT, and Elevator Counting. There was a significant positive correlation between difficulty level 1 and RSAT total speed (r = 0.315, n = 50, p < 0.05, two-tailed). Correlations between difficulty levels 2, 3, and 4 with RSAT total speed yielded non-significant results (p > 0.05). Significant positive correlations were found between RSAT total accuracy difficulty levels 1 (r = 0.367, n = 50, p < 0.01, two-tailed) and 4 (r = 0.348, n = 50, p < 0.05, two-tailed) of the Sustained Attention module. Exploratory analysis revealed significant negative relationships between the number of omissions and commissions on difficulty level 1 with RSAT speed (omissions r = -0.315, n = 50, p < 0.05, two-tailed; commissions r = -0.284, n = 50, p < 0.05, two-tailed) and accuracy (omissions r = -0.367, n = 50, p < 0.05, two-tailed; commissions r = -0.281, n = 50, p < 0.05, two-tailed). No significant relationships were found between any of the difficulty levels of the Sustained Attention module and TMT A (completion time or errors) or Elevator Counting.

Table 12b.

Correlations Between the Sustained Attention Module of the CBT and Measures of Attention.

	Difficulty Levels			
Attention Measures	1	2	3	4
RSAT				
Total Speed	-0.246	-0.179	0.061	-0.172
Total Accuracy	0.315*	0.182	0.127	0.006
TMT A	0.367**	0.258	0.172	0.348*
Elevator Counting	0.081	0.106	0.220	0.023

Abbreviations: RSAT = Ruff 2 & 7 Selective Attention Task, TMT = Trail Making Test.
** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).

Divided Attention

A Pearson product-moment bivariate correlation analysis was conducted between the correct score on the four CBT Divided Attention difficulty levels and TMT B. A significant negative correlation was observed (r = -0.500, n = 50, p < 0.01, two-tailed) between TMT B completion time and difficulty level 2 of the Divided Attention module. No significant relationships were found between difficulty level 1 of the Divided Attention module and TMT B completion time or errors.

Table 12c.

*Correlations Between the Divided Attention Module of the CBT and Measures of Attention***.**

	Difficulty Levels	
Attention Measures	1	2
TMT B	-0.242	-0.500**

Abbreviations: TMT = Trail Making Test.

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

Attention Control

A Pearson product-moment bivariate correlation analysis was conducted between the correct score on the four CBT Attention Shift difficulty levels and RSAT, Modified Six Elements, and Rule Shift Cards. There was a significant positive correlation between Map Search and difficulty level 1 ($r = 0.357$, $n = 50$, $p < 0.05$, two-tailed) and difficulty level 2 ($r = 0.549$, $n = 50$, $p < 0.01$, two-tailed). Indeed, a stronger relationship with difficulty level 2 was found with the Map Search raw score. Relationships between total accuracy scores on the RSAT and difficulty level 1 and 2 approached significance with $p = 0.076$ & $p = 0.056$, respectively but did not cross the threshold alpha of 0.05. No significant correlations were found between RSAT total speed or any of the BADS sub-scores with either difficulty level 1 or 2 of the Attention Shift module.

Table 12d.

Correlations Between the Attention Shift Module of the CBT and Measures of Attention.

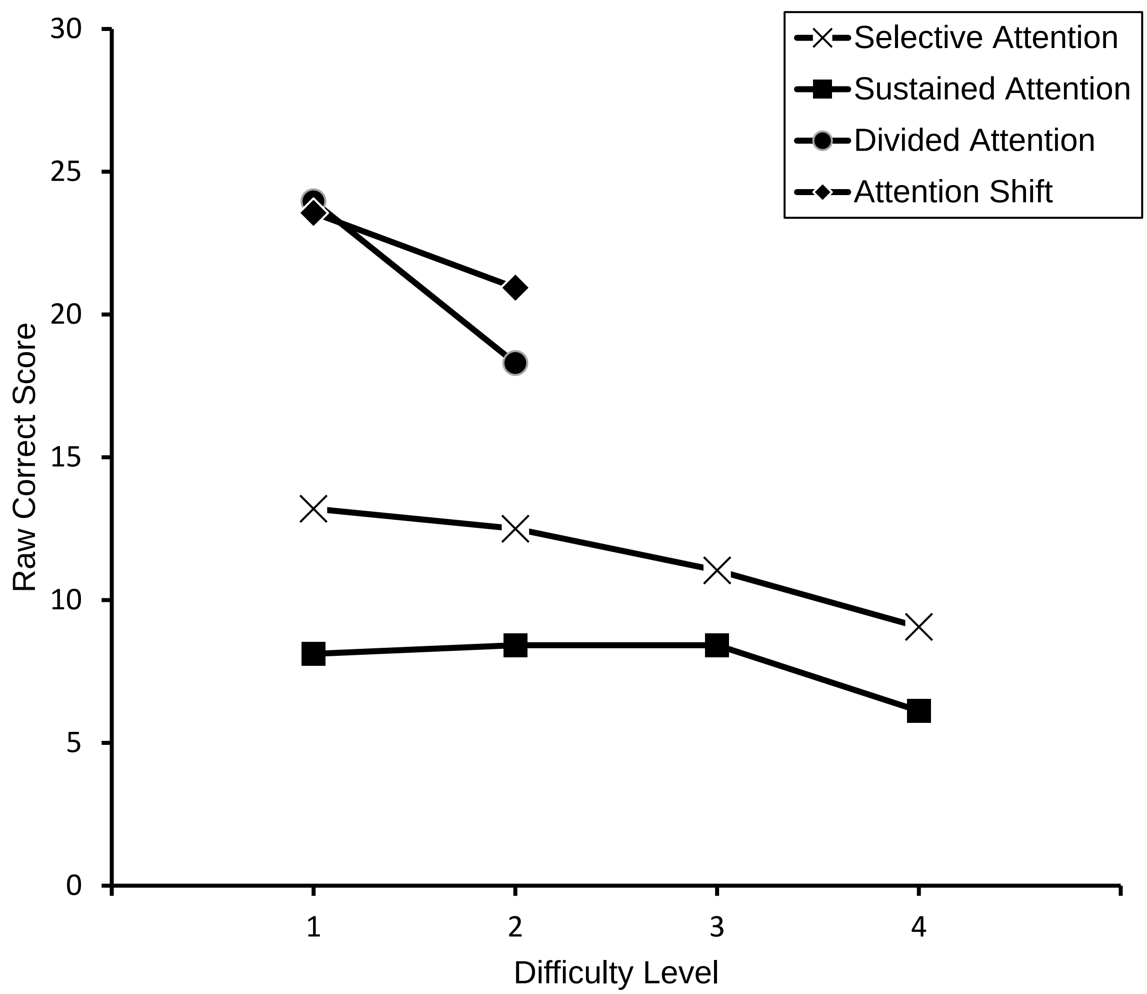
	Difficulty Levels	
Attention Measures	1	2
RSAT		
Total Speed	0.216	0.209
Total Accuracy	0.253	0.272
Map Search	0.357*	0.549**
Rule Shift Cards		
Total Time	-0.084	-0.150
Errors	-0.104	-0.063
Time Violations	0.124	0.001
Profile Score	0.068	0.015

Abbreviations: RSAT = Ruff 2 & 7 Selective Attention Task.
** Correlation is significant at the 0.01 level (two-tailed).
* Correlation is significant at the 0.05 level (two-tailed).

5.6.2 Difficulty Levels Analysis

Although the specific parameters for each CBT module difficulty levels were determined through a pilot study with healthy undergraduate students, it was uncertain if these results would translate into our mTBI sample. In an effort to determine if increasing difficulty within each module reduced performance across the sample, a repeated measures ANOVA was employed. Since the distributions of the total correct performance across all difficulty levels on all modules were non-normally distributed, nonparametric analyses were performed. The Friedman test is a nonparametric equivalent of the one-way within-subject ANOVA and was employed. Overall, as difficulty level increased, mean score decreased significantly (Selective Attention $\chi^2(3, 50) = 40.4$, $p < 0.001$; Sustained Attention $\chi^2(3, 50) = 44.0$, $p < 0.001$; Divided Attention $\chi^2(1, 50) = 20.5$, $p < 0.001$; Attention Shift $\chi^2(1, 50) = 7.0$, $p < 0.01$)). Figure 5 displays the mean raw correct scores of all CBT modules across difficulty levels.

Figure 5. Line Graph of the Mean Raw Correct Score Across Difficulty Levels in the Selective Attention, Sustained Attention, Divided Attention, and Attention Shift Modules of the Conveyor Belt Task.



Across the entire sample, the difficulty level analysis confirmed an inverse relationship between difficulty level and performance on the CBT modules. This result suggested the higher difficulty levels were more difficult with decreasing performance scores. One exception to this was between levels 1, 2 and 3 of the Sustained Attention module. Performance on these first three trials were not statistically significant from each other suggesting that there was no change in experienced difficulty. Rather than change specific parameters such as speed of stimulus delivery or frequency, each of these levels was conceptually different. On all levels, the main component of attention was sustained attention, however, the stimulus delivery varied between difficulty levels as compared to the other modules. For example, in difficulty level 1, participants had to identify non-similar objects (non-globes), whereas on difficulty levels 2 and 3 participants were instructed to identify subtle changes to a similar stimulus (falling over or colour change). By design, this was done in an effort to manipulate the attention spotlight by increasing the subtly of the stimulus change. Lastly, the inter-stimulus interval between target stimuli was less in difficulty level 4 as compared to the other levels. In other words, the stimuli were presented at a faster pace. Similar speed, frequency, and inter-stimulus interval may have contributed to the lack of differences in performance across difficulty levels 1-3. In addition, the differences in the design of the target stimuli may have also factored into the lack of performance difference.

5.6.3 Discriminant Analysis

Across all test measures, higher standardized scores were indicative of better performance. Group differences were examined using ANOVA with RTW (yes or no) serving as the independent variable and neuropsychological scores as the dependent variables. Descriptive statistics for all scores were reviewed and the following variables showed ceiling or floor effects and were therefore, dropped from the discriminant functions analysis.

- TMT number of errors
- WCST failure to maintain set
- TOL time violations
- TOL rule violations
- Rule Shift Cards total errors
- Rule Shift Cards time violations
- Rule Shift Cards profile score
- Modified Six Elements rule violations
- Modified Six Elements time violations
- Elevator Counting correct score
- Office Task incorrect deliveries
- Office Task failure to maintain set
- Office Task perseverations
- CBT Selective Attention T1 – commissions
- CBT Selective Attention T3 – commissions
- CBT Sustained Attention T2 – omissions
- CBT Sustained Attention T3 – omissions
- CBT Sustained Attention T4 – omissions
- CBT Divided Attention T1 – commissions
- CBT Divided Attention T2 – commissions
- CBT Attention Shift T1 – commissions
- CBT Attention Shift T2 – commissions

Within the data, 11 cases were missing values due to either noncompliance or computer error on the VR tasks. This accounted for a range of 0 to 3 missing values (out of 50) across all predictor variables with a total of 107 individual data points. Since discriminant analysis does not tolerate missing values per case, missing values were replaced. To determine the type of imputation to be used, a test of normality revealed that all test measures were skewed (Shapiro-Wilk $p < 0.05$), with the exception of WCST number of non-perseverative responses in t-scores (Shapiro-Wilk $p = 0.57$). Accordingly, a generalized median imputation of the series was used to replace the missing values (Burke, 2001).

To determine the least number of significant predictors that would predict return to work status, a step-wise discriminant analysis was performed. Wilk's lambda method was employed with criteria for entry set at $p = 0.05$ and removal at $p = 0.15$ (Tabachnick & Fidell, 1996). The predictors were entered en masse. The disproportionate impact of outliers was minimized using the winsorizing method as per Tabachnick and Fidell (1996). The dependent variable (criterion) was set as RTW and age, education, NAART FSIQ, along with all sub-scores on the TMT, RSAT, SDMT, WCST, TOL, BADS (Modified Six Elements & Rule Shift Cards), TEA (Elevator Counting & Map Search), Office Task, and CBT as predictor variables. A total of 50 cases were analyzed; 23 that returned to work and 27 that had not. Univariate ANOVA revealed significant between group differences on parts A and B of the TMT, RSAT (total speed and accuracy), SDMT, Map Search and Elevator Counting, along with total correct score, number of commissions, and final score of trial 2 of the Selective Attention module of the CBT; total correct score, number of omissions, and final score of trail 3 of the Selective Attention module of the CBT; number of commissions of trail 2 of the Sustained Attention module of the CBT; number of commissions and final score of trial 3 of the Sustained Attention module of the CBT; final score of trail 4 of the Sustained Attention Trial; correct selection, number of omissions, and final score of trials 1 and 2 of the Divided Attention module of the CBT; and correct selection, omissions, and final score of trial 2 of the Attention Shift module of the CBT. The discriminant analysis, however, revealed two functions with better discrimination at the final step ($\lambda = 0.586$, $p < 0.001$) which included the predictors: the number of correct selections on trial 2 of Attention Shift module of the CBT and the total speed (t-score) of the RSAT ($D_i = -4.812 + 0.055$[Ruff total speed t-score] + 0.114[Attention Shift Trial 2 correct]). The value of this function was significantly different for those who returned to work and those who did not ($\chi^2 (2) = 25.1$, $p < 0.001$). All other variables were removed from the function as they did not meet the entry criteria. The null hypothesis was rejected as per

Wilks' Lambda test for the null hypothesis ($p < 0.05$). The correlations between two predictor variables and the discriminant function suggested that the number of correct selections on trial 2 of Attention Shift module of the CBT and the total speed (t-score) of the RSAT were the best predictors of employment. Both were positively correlated suggesting higher scores (better performance) was more likely to predict RTW. Overall, the discriminant function successfully predicted outcome for 77.8% of the participants who did not RTW and 87.0% of participants who did. The function successfully predicted 82.0% of criterion membership with 80.0% split-half cross-validation support. Group size classification for this function predicted a specificity of 81.5% (chance of false negative results; Type II errors) and sensitivity of 82.6% (chance of false positive results; Type I errors).

5.6.4 Effect Sizes

Effect size estimates were calculated for each predictor and displayed in Table 10 (a-c). An examination of the effect sizes revealed that performance across test measures were small to large as per Cohen's (1988) heuristic. Mean performance between the groups on the TMT (A & B), RSAT (total speed and total accuracy), SDMT, Map Search, Elevator Counting, and all modules of the CBT were statistically significant. The effect sizes of these predictors revealed the following observations. First, measure from all points on the ecological validity spectrum (traditional, ecologically oriented, and VR) statistically discriminated between the two groups. Second, nearly all of these test measures would fall under the domain of attention and nearly all of them have some sort of timed aspect implicating processing speed. Third, the largest effect sizes, in descending order, were observed on the CBT (Attention Shift difficulty level 2, d = 1.32, Divided Attention difficulty level 2, d = 0.97, Sustained Attention difficulty level 3, d = 0.67; Selective Attention difficulty level 1 d = 0.79), RSAT (total speed, d = 0.93), SDMT (d = 0.90), Map Search (d = 0.88), TMT A (d = -0.84), TMT B (d = -0.76), Elevator Counting (d = 0.58).

5.7 Discussion

The present study sought to develop novel VR tasks of attention and executive function with ecological validity in mind. Specifically, performance on traditional pen-and-paper neuropsychological test measures, ecologically oriented tests (BADS and TEA), and the novel VR tasks were compared in a sample of patients in the post-acute recovery period of mTBI who were employed or unemployed at the time of testing. The general objective of the study was to improve ecological validity with the development of VR tasks that would have better predictive ability of employment than commonly employed pen-and-paper neuropsychological test measures. A discussion of the results as they pertain to the specific aims of the current study will follow.

Concurrent validity of the VR tasks was examined via correlational analysis with traditional tests of executive functions and attention (as per Table 9). With respect to the Office Task, a significant association with the sub-scores and their WCST equivalents were not found after correcting for multiple comparisons via Bonferroni correction (Table 11). Moreover, no associations were found on the TOL sub-scores. Number of errors on TMT B were associated with failure to maintain sets on the Office Task. In addition, the number of errors on the Rule Shift Cards was positively associated with incorrect deliveries and failure to maintain set, however, correct deliveries were not associated with completion time or the overall profile score on the Rule Shift Cards. Similarly, positive correlations were found between the number of errors on the Office Task and the Modified Six Elements. Specifically, the number of perseverations on the Office Task were significantly related to the number of rule violations on the Modified Six Elements. Correct scores were not associated between these two tasks. These findings suggest that the Office Task demonstrates good concurrent validity for measuring error but not for target responses.

Though the Office Task was designed as a partial VR analogue to the WCST, there are some key differentiating characteristics which may have contributed to the lack of correlation between equivalent sub-scores. First, both tests have an element of cognitive flexibility. The WCST changes the target category every 10-correct trials without announcement, whereas the Office Task elicits cognitive flexibility on *each* trail. As noted above, the participant must use the limited information presented on the package to reason which door it should be delivered to. Second, the WCST requires intact working memory in order to remember which sorting principle is the current one. Since the package contents are visible throughout the trial, the utilization of an intact working memory is not necessary for successful completion of the Office Task. Working memory has indeed been argued as a critical subcomponent of executive functions in nearly all currently accepted theoretical models of executive functioning (see Baddley, 1986; D'Esposito et al., 1995). Thus, the lack of working memory involvement on the Office Task may reduce its sensitivity to executive functioning. Since executive dysfunctions become more evident in the most complex of aspects of human consciousness activity (Zillmer, Spiers, & Culbertson, 2008), the lack of recruitment of this critical subcomponent of executive functions from the Office Task may have made the task too easy, causing a ceiling effect. This leads into the last point, which was the perceived level of difficulty of the Office Task. Informally, many participants noted that the task was easy, unstimulating, and even boring. Since data was not collected regarding this, analysis into those who felt this way versus those who did not was not available. However, since there were no differences with respect to task performance by way of employment status on the Office Task, it may be argued that such sentiments might have been experienced by both. Indeed, a review of the mean correct deliveries reveals that the unemployed group actually delivered a higher number of correct packages ($M = 46.2$, $SD = 4.48$) than the employed group ($M = 41.4$, $SD = 13.14$). This finding was not statistically significant ($p = 0.101$) but a medium effect size was observed ($d = 0.51$).

The concurrent validity of the CBT modules was found to be more robust than the Office Task. The following discussion is based on the correlational findings (see Table 12 a-d) and is organized by module.

The Selective Attention module was found to have strong concurrent validity suggesting that this module was tapping into the same construct as the traditional pen-and-paper neuropsychological test measures of selective attention. Specifically, difficulty level three was significantly and positively correlated with all other test measures of selective attention with the strongest association with Map Search ($r = 0.665$, $p < 0.0001$). Interestingly, difficulty level 4 was not significantly associated with any of the test measures. An exploratory correlational analysis was conducted separately between difficulty level 4 and all other traditional test measures, yielding no significant associations. As per the configurations and the difficulty level analysis (see Figure 5), the parameters for difficulty level 4 may have been too hard as compared to the other difficulty levels. The data suggests that difficulty level 4 may have been measuring a unique cognitive construct relative to the traditional test measures.

Good concurrent validity was found with respect to the Sustained Attention module as per the moderate and statistically significant correlations found on difficulty levels 1 and 4 on the RSAT. Difficulty level 1 of the Sustained Attention module required that the participant press a keyboard key on the appearance of any non-globe related items on the conveyor belt. On the RSAT, participants are required to mark target numbers (2 & 7) and ignore all other items. The administration and task requirements of both tests are very similar and thus, it is not surprising that there is a significant relationship between test performances on these test measures. Moreover, both are cancellation tasks to an extent and as such, each presented stimulus must be processed to some extent. This implies the involvement of speed of processing (reaction time) and inhibition/accuracy components of attention. Indeed, there was a significant negative relationship between the number of omissions and commissions on difficulty level 1 with RSAT speed and accuracy. Thus, total speed and accuracy were indeed affected on the Sustained Attention module as they were on the established test measures of attention. Taken together, these findings support the hypothesis that the Sustained Attention module of the CBT has good concurrent validity for measuring speed of processing and inhibition/attention control.

Difficulty level 2 and 3 of the Sustained Attention module did not correlate with any of the established test measures of sustained attention. As a review, the target stimulus was globes falling over in difficulty level 2 and globes rusting (colour change) on difficulty level 3. Difficulty levels 2 and 3 could have been too easy as reflected by the higher mean scores and lack of errors on difficulty levels 2 and 3 compared to 1 and 4. Further support that difficulty levels 2 and 3 were perhaps perceived as easier than trials 1 and 4 can be observed via the difficulty level analysis (Figure 5). As can be seen, performance on difficulty levels 2 and 3 were actually better than levels 1 and 4. Despite the parameters being devised from a pilot study with healthy undergraduate controls, the results of the current study suggest that future explorations with an mTBI sample modify the parameters in levels 2 and 3 in an effort to produce a linear increase in performance difficulty and in turn to improve concurrent validity.

The lack of significant correlations between any difficulty level of the Sustained Attention module and Elevator Counting and TMT A may be due to ceiling effects and short sampling window. The performance score generated by the Elevator Counting has been reported to have ceiling effects as the score ranges from 1-7 (Chan et al., 2006; Strauss et al., 2006). The means on the Elevator Counting test in our current sample were 6.7 (*SD* = 0.76) and 6.0 (*SD* = 1.47) for the employed and unemployed groups, respectively. These means were statistically significant but the effect size here was medium (Cohen, 1988) with an percent overlap of approximately 66% (Zakzanis, 2001). Moreover, both Elevator Counting and TMT exhibit a short sampling window. Though various models of attention have been proposed in cognitive psychology, the general consensus is that sustained attention involves maintaining attention control over a period of time (Mirsky et al., 1991; Peterson & Posner, 2012). Moreover, sustained attention tasks that are longer with generally unbroken periods of time have been reported to be more sensitive to subtle attention deficits (van Zomeren & Brouwer, 1994). The TMT and Elevator Counting are short duration tasks with a combined duration range of 16-116s in our sample, whereas the RSAT and the Sustained Attention module of the CBT are considered longer duration tasks, as they require 120-300s to complete. Thus, our findings suggest that the Sustained Attention module of the CBT is conceptually related to sustained attention tasks with longer sampling windows such as the RSAT.

With respect to the Divided Attention module, good concurrent validity was observed as per the moderate negative relationship with TMT B. The negative relationship indicates that higher (better) performance on the Divided Attention module was associated with lower (better) times on the TMT B test. This finding supports the obvious face validity of being a divided attention task as the participant is responsible for two conveyor belts at the same time. The significant relationship was observed on difficulty level 2 only. Level 1 may have been too slow to elicit any sort of meaningful cognitive load. Two relevant limitations are worth noting. The first is that only one traditional measure of divided attention (TMT B) was used to establish concurrent validity. The second is that the target on the Divided Attention module of the CBT was always either in the right or left. Thus, participants could use a strategy whereby they focus on one conveyor belt to determine where the target stimulus was. This is a design flaw and in retrospect, should have been corrected. Follow-up studies with additional traditional divided attention test measures in addition to a mix of trials with no target stimulus should be administered in an effort to determine the robustness of the current findings.

The Attention Shift module of the CBT correlated with Map Search from the TEA (r = 0.549, p < 0.01). No other correlations were found between any of the other traditional tests or ecologically oriented tests and the Attention Shift module. These results provide some support for concurrent validity with respect to attention shift. The Map Search subtest of the TEA has attentional control/attentional inhibition properties in that the participant must search and compare each stimulus to determine if it is a target or a distractor. For successful performance on this task, the decision to act or ignore the stimuli requires inhibitory control to reduce the number of errors in the form of omissions and commissions. Based on factor analysis studies by others (Bate et al., 2001; Chan et al., 2000; Robertson et al., 1996) the Map Search test loads onto visual selective attention factors. In this sense, the Attention Shift module may be more concurrently valid with respect to a more challenging selective attention task rather than one that has more loadings onto an attentional shift or inhibitory control construct. In the current study, the correlation between the RSAT, a test of selective and sustained attention (Bate et al., 2001), approached statistical significance with the Attention Shift module (p = 0.06). Further evidence for this has been found and studies with the children's edition of the test of everyday attention. Not only did the Sky Search (Map Search equivalent of the children's version of the TEA) load it onto the selective attention factor, but it was also a significant predictor and discriminating between children with ADHD and healthy controls (Manly et al., 2001).

The ecological validity of the CBT was demonstrated by the results of the discriminant functions analysis, whereby performance on the second difficulty level of the Attention Shift module was able to predict group membership. In addition, the RSAT total speed performance was also able to distinguish those who were employed versus unemployed, though to a lesser degree than the CBT Attention Shift difficulty level 2 performance. Combined, these yielded a model that was able to successfully predict group membership with 82% overall accuracy. Specifically, the model was able to predict those who were unemployed at a better accuracy (87%) versus those who were employed (77.8%). Potentially, this finding may be due to uneven sample sizes, with a larger unemployed sample. Moreover, the sensitivity and specificity of this model were found to be 82.6% and 81.5%, respectively, which may also account for the higher accuracy of the unemployed group. Of the 54 total predictors entered into the discriminant functions analysis, only these two predictors were found to have some statistically significant ability to predict group membership.

The prediction accuracy and sensitivity and specificity of the current study are generally higher than what has been reported in prior studies. Paniak, Toller-Lobe, Melnyk, and Nagy (2000) investigated variables such as PTA, premorbid alcohol use, injury age, premorbid psychological problems, gender, previous TBIs, seeking or receiving financial compensation, bodily injuries, SES, education, and pre-injury adverse life events in their ability to predict employment outcome in a sample of 118 patients with mTBI who were 3-4 months post-injury. Their model found that only age and seeking or receiving financial compensation were predictive of employment status three-four months after injury and was accurate with a 79% correct classification rate. Neuropsychological variables were not investigated, and thus cannot be directly compared with the current study. Interestingly, our classification accuracy rate is similar to that of Paniak and colleagues despite having different predictor variables contribute to each model. In Paniak’s study, neuropsychological variables were not used as predictor variables whereas in the current study, they were in addition to demographic variables. Klonoff et al., (2007) found that tasks of processing speed (visual scanning, visuospatial skills and memory) were able to differentiate employment status in a sample of mixed neurological patients. It should be noted, however, that their sample only consisted of 4/38 patients who sustained a mTBI. Using a sample of 483 head-injured and workers compensation-seeking patients, Bowman (1996) conducted a discriminant functions analysis to determine the best predictors of employment outcome. Their results indicated a model where neuropsychological test performance (specifically on the block design and the Wechsler Memory Scale) in addition to injury age, gender, time since injury, occupational functioning at the time of the injury, MMPI score, and whether or not the individual was involved in rehabilitation programs predicted employment outcome to 72% accuracy. This model, however, was only able to account for 34% of the overall variability. A large portion of the variability (66%) remained unexplained by unmeasured variables. This finding is important in that their sample was Canadian and thus comparable to our sample in the current study. Moreover, the sample employed in the Bowman (1996) study were all patients who had made it through the Worker's Compensation Board system to receive a neuropsychological assessment and thus had greater difficulties with cognitive functioning. In other words, it can be argued that their sample contained the severest of the head injured group. Unfortunately, these authors did not report the proportion of the mTBI proportion in their sample and as such, heterogeneity of the samples may limit further comparison. Gollaher and colleagues (1998) found that preinjury employment and education (in years) were the only two predictors that were able to correctly classify employment status to 75% accuracy. Though their sample consisted of TBI patients, severity was not reported and it is uncertain how much influence more severe injuries have contributed to this finding.

The findings from the current study suggest that the number of errors made may not be sensitive to distinguish employment status in a sample of post-acute period patients with mTBI. This is in contrast to some previous and related studies. For example, Dawson and colleagues (2009) employed a variant of the Multiple Errands Test (called the Baycrest Multiple Errands Test) on an ABI sample and healthy controls for the purposes of determining the psychometric properties of the Baycrest Multiple Errands Test. Their results suggested that the total number of errors, not types of errors, contributed to the discrimination between the two groups. Their findings echoed that of Alderman and colleagues (2003) who reported a dichotomous pattern of behaviour on the simplified version of the Multiple Errands Test; *rule breakers* and *task failures*. Both studies concluded with discussions highlighting the theoretical importance of error monitoring on executive processes in neurological patients. Namely, that the number of errors may be a clinical effect that distinguishes patients from healthy controls. Moreover, these finding were replicated in another ABI sample but with the use of a virtual version of the Multiple Errands Test called the Multitasking in the City Task (Jovanovski et al., 2012). Similarly, Rand and colleagues (2009) employed yet another VR version of the Multiple Errands Test and found similar results; the total number of errors produced on the task significantly separated the ABI group versus the control group. These studies used mixed ABI samples composed of patients who sustained strokes along with patients of mixed severities of TBI at various stages in their recovery.

The aim of the current study was not to differentiate between healthy participants and patients with mTBI and thus a control group was not employed. To this end, a direct conceptual comparison of the studies reviewed above versus the findings of the current study cannot be made, though, support for the error hypothesis has been found elsewhere (see Dean & Sterr, 2013). Aside from the obvious differences in patient characteristics of sample from the current study versus the ones reviewed above, the lack of errors made on the Office Task in the entire sample may explain this finding. On the Office Task, the average number of incorrect deliveries made was approximately 1 for each group, which was not statistically significant. Moreover, this finding was extended to that of traditional pen-and-paper neuropsychological test measures and ecologically oriented tests. Indeed, none of the types of errors made on the TMT, WCST, TOL, Rule Shift Cards, or Modified Six Elements were able to distinguish between employment statuses in our post-acute period mTBI sample. The results of the current study add to the body of research by suggesting a limitation to the discriminate power of the number of errors on tests of executive functions with respect to RTW status. To examine this phenomenon further, a comparison between mTBI samples and healthy controls would be required to determine the reliability of this finding. It was not the aim of the current study to determine the cause of such errors, but it is postulated to be related to a dysfunctional attention response as will be discussed next in conjunction with the findings of the CBT.

In partial contrast to the second and third hypotheses, performance on some traditional pen-and-paper neuropsychological test measures differentiated employment status in our mTBI sample. Moreover, mean performance on ecologically oriented tests were also significantly different between employment groups. However, and in support of the hypothesis, mean performance on the CBT also differentiated the groups. As per Table 10 (a-c), performance on the following test measures were statistically different between employment groups: TMT A (time), TMT B (time), RSAT (speed and accuracy scores), SDMT, Map Search, Elevator Counting, and all modules of the CBT. None of the executive test measures, including the Office Task, differentiated the two groups. These results suggest that tests of attention with a timed factor are able to differentiate between employment status in patients with mTBI in the post-acute period of recovery.

Support for the third hypothesis was found by way of an effect size analysis. The largest effect sizes were observed on the CBT. Specifically, the Attention Shift (d = 1.32) and Divided Attention modules (0.97) produced the largest effect sizes, however, all the modules produced effect sizes that were in the medium to large range. Following this, large effect size estimates were observed on ecologically oriented tests such as the Map Search (d = 0.88) and traditional tests such as the RSAT (d = 0.93) and SDMT (d = 0.90). In addition, percent non-overlap (Cohen, 1988) values can be used to examine the differences in magnitudes and can be used as a proxy to interpret how ecologically valid each of these predictors are. To this end, the highest percent non-overlap was observed in the Attention Shift module (65%), followed by the Divided Attention module (47-55%), then the Map Search, SDMT, and RSAT total speed (all 52%), followed by the TMT (43-47%), and finally the Elevator Counting (38%). Generally, the trend suggests that the CBT was the most ecologically valid, followed by a relative tie between the ecologically oriented tests and traditional tests. On average, however, the ecologically oriented tests had slightly higher values of percent non-overlap than the traditional tests, but this may have been an artifact of unequal sample sizes between the groups as there were more traditional test predictors (k = 5) than ecologically oriented ones (k =2). Alternatively, these findings suggest that tests that use the verisimilitude approach are more ecologically valid than those that use the veridicality approach. This is corroborated by previous studies that compared a VR measure with traditional pen-and-paper neuropsychological test measures and found relatively large effect sizes in favour of the VR task (Cipresso et al., 2014; Dawson et al., 2009; Man, Poon, & Lam, 2013; Pollak et al., 2009). For example, Dawson and colleagues (2009) compared the performance of 27 patients with acquired brain injury (TBI and stroke) and 25 matched controls on traditional pen-and-paper neuropsychological test measures and the Baycrest Multiple Errands Test. These researchers found large effect sizes between the TBI group and their matched controls on the Baycrest Multiple Errands Test (d = 0.75 to 1.02). These authors concluded that the Baycrest Multiple Errands Test was more ecologically valid than traditional tests.

It is noteworthy to mention that the more comparisons that are made, the increase in the likelihood of a Type I error (false-positive; Sato, 1996). The discriminant functions analysis is less prone to such errors and therefore more robust. To this end, the discriminant functions analysis yielded a final model that consisted of only two test measures, The RSAT, and the CBT (Attention Shift difficulty level 2), which predicted employment status with high accuracy (82.0%). As per the manual and validated by other analyses (see Messinis et al., 2007) the total speed (number of targets identified) and total accuracy (correct targets divided by possible targets) score of the RSAT is considered a measure of sustained attention. As per the concurrent validity results, the Attention Shift module can be interpreted as a challenging sustained attention task with a distractor. By adding a distractor stimulus (malfunctioning arrow) in addition to two target stimuli (conveyor belts), the number of items to attend to in parallel increases. Logically, such information would take longer to process. In light of the mental slowness commonly reported by patients who continue to be symptomatic following mTBI, poor performance on this task may have been due to increased cognitive load. Tipper and Baylis (1987) found that active inhibition of distractors might be a mechanism associated with efficient selection. Kewman, Yanus, & Kirsch (1988) extended this finding to a head injured group. They found that processing of greater information loads through a binaural voice experiment differentiated between head injured and control groups. Similar findings have also been reported elsewhere (Conway, Tuholski, Shisler, & Engle, 1999) and the *response conflict* was coined by DeLand (1992) considering patients required longer time to process information in light of distractions. Cognitive control tasks such as the Stroop and Flanker Test have also been shown to differentiate between controls and patients in the post-acute phases of mTBI. Losoi et al. (2016) found no differences between controls and patients with mTBI who were 1- or 6-months post-injury on the Stroop test. However, they reported a significant difference between 1- and 6-months post-injury (p = 0.01) on the Stroop test. The authors attributed this finding to practice effects and improvements in cognition from 1 to 6 months post-injury. It should be noted that they conducted 42 separate t-tests and did not correct for multiple comparisons. In contrast, other studies did not find any differences between groups on the Stroop test (Kraus et al., 2007; Mangels et al., 2002). Pontifex and colleagues (2012) used a modified Flanker task in a sample of athletes with a self-reported history of concussions and healthy controls. The concussions were on average 3.6 years since the date of the assessment. They found that the concussion group had decreased response accuracy on the flanker task while no other differences were found on the ImPACT. The authors speculated that these deficits may be the result of inefficient neural resource allocation (Broglio, Pontifex, O'Connor, & Hillman, 2009). Taken together, there is evidence to support that cognitive control tasks such as the stroop and flanker task may capture long-term cognitive deficits following a mTBI. Studies investigating the ecological validity of these tests with respect to RTW should be further explored.

From a cognitive neuroscience perspective, response conflict on attentional tasks fit in with the vigilance attention system as per Posner's Anterior and Posterior Attention Model (Posner & Peterson, 1990). The vigilance attention system has been postulated to be responsible for achieving and maintaining an alert state whilst actively avoiding irrelevant distractors. Tests of selective and sustained attention have been shown to be associated with the vigilance attention system (Peterson & Posner, 2012). Since the RSAT and Attention Shift module of the CBT are sustained attention tests that require focusing on targets and avoiding irrelevant distractors, it would be plausible that the vigilance attention system would be active during their administration. Response conflict may have contributed to increased latency times and errors in the unemployed group as compared to the employed group. To this end, our findings may have shown indirect support for deficits in the vigilance attention system in employment status in patients with mTBI who are in post-acute period of recovery.

Studies have shown that response to rapidly presented and externally paced stimuli is positively associated with neurologic severity of injury and overall cognitive impairment following TBI as opposed to internally paced stimuli (Baribeau, Ethier, & Braun, 1989). Such findings can be related to impaired information processing speed, drain on attentional resources, and disrupted attention effort following a TBI (Ponsford & Kinsella, 1992; Schmitter-Edgecombe, 1996). It has been postulated that these impairments are related to the anterior attention network with involvement of the anterior cingulate gyrus (Deary, Ebmeier, MacLoud, & Dougall, 1994; Posner & Rothbart, 1992). Although speculative, since the RSAT and the CBT are both externally paced test measures of attention, it may be that the anterior attention network may be indirectly important for RTW outcome. Further research is necessary to validate this speculation.

Our finding that performance on attention, but not executive test measures, was predictive of employment status in an mTBI supports previous studies. Research examining the neuropsychological variables that predict RTW status in patients with mTBI is limited. Accordingly, broad comparisons with our findings with that of mixed severity TBI follows. Lengenfelder et al. (2002) examined driving fitness as outcome using a VR task in patients who sustained a severe TBI. They concluded that the number of errors on test measures of attention distinguished the TBI and healthy control group. Ruff et al., (1993) used demographic, medical, neuropsychological, and psychosocial variables to determine RTW in a sample of patients with 'severe head trauma'. They found that speed of information processing, age, and 'verbal intellectual power' was most predictive of employment return. Haboubi, Long, Koshy, & Ward (2001) conducted a retrospective study examining 1255 patients with 'minor head injuries' and found that 38% were unable to RTW. The most common symptoms reported by those who were unable to RTW were fatigue, headache, dizziness, irritability, sleep disturbances, poor concentration, and poor memory. To this end, similar RTW rates and symptoms reported were also found in a previous study conducted by Levit et al., (1994).

There is research support to the contrary of our findings. Drake et al. (2000) conducted a discriminant functions analysis on 121 military patients with mTBI and found that verbal memory, verbal fluency, and a speeded test of planning and strategy predicted work status at 3-15 months post-injury to an accuracy of 68.8%. Key differences between the current study and that of Drake and colleagues (2000) are military versus litigating mTBI samples and the limited number of attention and memory tasks employed. As to the latter, only one measure of attention was administered in Drake's study, which was the Paced Auditory Serial Addition Test limiting the sensitivity of this study to attention deficits. A meta-analysis conducted by Ownsworth & McKenna (2004) found that executive functioning, which was defined with concepts such as concept formation, complex attentional skills, mental flexibility, mental programming, and planning ability, were the most reliable predictors of RTW in a mixed sample of TBI. It can be argued, however, that both attention and executive functioning were predictive as *complex attentional skills* included tasks of divided and selective attention. Crepaeu and Scherzer (1993) conducted a meta-analysis to determine the predictors of work status after severe TBI and found that executive dysfunction, emotional disturbances, deficits in activities of daily living, and less vocational rehab services best predicted employment status.

Taken together, additional research with a mTBI population is required to confirm our finding that sustained attention deficits are correlated with unsuccessful RTW.

5.7.1 Limitations

As with all studies, limitations are present and need to be discussed. The study had a relatively small sample size limiting inferential statistics and the power of the findings. However, "effect size *d* are not dependent on nor influenced by sample size" (Zakzanis, 1999, p. 12). Further, the sample comprised of a convenience sample of litigating patients with mTBI limiting the generalizability of the findings. The generalizability and reliability of the current findings needs to be replicated in a non-litigating sample. Furthermore, all patients were in the post-acute period of recovery but time since injury was quite broad (94 days to 1650 days). To this end, sequential recruitment of patients in a hospital setting with multiple follow-up assessments during the acute, sub-acute, and post-acute periods would control for potential recovery time effects. Though not essential to the scope of the current study, including a matched control group would have improved the specificity of our findings and allow further discussion on the breadth (and limitations) of the VR tasks. A pilot study (not reported) was initially completed in an effort to obtain usable parameters on the CBT. However, the sample size was quite small (n = 9), they were not matched to age and education of the mTBI group and had not completed any of the other test measures employed in this study. Indeed, the outcome variable of interest, RTW, was self-reported by each patient and was dichotomous as opposed to stratified such as RTW with modified hours/duties as in other studies (Wäljas et al., 2014). Employment status confirmation by either the patient's employer or occupational therapist would have improved veracity of the patient's self-report. However, a follow-up with their employer may be considered intrusive and in breach of privacy. In addition, not all patients were assigned occupational therapists, thereby limiting the number of eligible participants to the study. As noted above, additional test measures especially of divided attention would have improved the number of comparisons made. Particularly and in retrospect, the Elevator Counting with Distractors subtest of the TEA would have been conceptually comparable to the Attention Shift module of the CBT and has been shown to discriminate between TBI and healthy controls (Bate et al., 2001).

The current study also was limited by the large number of comparisons that were conducted, especially in the correlation matrix. Conducting multiple correlations inflates Type I errors but Bonferroni corrections also increases the chances of increasing Type II errors. The difficulty is choosing an appropriate threshold for significance, balancing the risk of making Type I and Type II errors. As per Table 11, there were 23 correlations conducted per variable and it needs to be acknowledged that there is a very good chance that at least one of the significant findings ($p < 0.05$) would be due to random chance. In addition, it should be expected to observe one or more trends in the findings ($0.05 < p < 0.10$). However, if there are say 8 significant correlations, this would hint at an overall pattern that would be above chance. As per Table 11 on average only 1 correlation was found to be significant for each of the 4 Office Task variables. The significant probability values ranged from 0.001 to 0.007, indicating that even with a Bonferroni correction (0.05/23 = 0.002) only two of the comparisons would be found to be statistically significant. As a result, it is very much likely that the findings do not indicate a clear pattern and one cannot, in good conscious, conclude to have found a robust finding. With respect to the CBT comparisons, it is a different situation. There were only a few comparisons (range of 1 to 7 depending on the sub-scale of attention) per variable and there were clear trends in the findings with probability values that were below even Bonferroni thresholds.

Despite these limitations, the current study provides an important contribution to the ecological validity of assessment instruments in patients with mTBI who are in the post-acute period of recovery.

5.7.2 Conclusions

Taken together, the findings of this study suggest that the second difficulty level of the Attention Shift module of the CBT may be more of a litmus test for employment status for those who are in the post-acute stages following an mTBI. These results are encouraging given the lack of ecological sensitivity and specificity in predicting employment status in patients with mTBI in the literature. Certainly, the results of the current study do not warrant any clinical recommendations for employment prediction in an mTBI sample. This was not the intention. Potentially, these findings may provide a starting point for future studies to determine, with increased breadth and specificity, which particular parameters and combination of tests (traditional or virtual) would comprise an ‘ecological battery’ in an effort to place clinical neuropsychological diagnostic decision making as it pertains to employability on firmer scientific grounds.

Chapter 6

6 General Discussion

Despite neuropsychological testing being a standard component of the assessment of mTBI and decades of research, reasons for persistent symptoms past the typical recovery period remain complex and contentious. Broadly speaking, the consensus from the literature is that the majority of those who incur an mTBI will recover to premorbid cognitive abilities within three months (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005; Rohling et al., 2011). Some patients continue to report long-term cognitive difficulties, however, that can result in reduced employment capacity and disability (Cancelliere et al., 2014; Ponsford et al., 2008; Sherer et al., 2002).

Previous research has proposed numerous factors that contribute to resulting functional impairment following a mTBI (e.g., physical limitations, premorbid personality characteristics, psychosocial situation, caregiving resources, assistive technologies, and contextual/environmental factors), including limited sensitivity of our current objective tests (Silverberg & Millis, 2009). Some studies have provided support for the veridicality or predictive ability of neuropsychological test measures and RTW (Drake et al., 2000; Hayden, 1997) and others show no relationship (Mooney et al., 2005; Nolin & Heroux, 2006; Studerus-Germann et al., 2017; Wäljas et al., 2014).

Indeed, neuropsychology has been undergoing a paradigm shift with an emphasis on real-world functional ability rather than cognitive strengths and weaknesses. Batteries such as the BADS and the TEA have been developed that are ecologically oriented by way of verisimilitude (resemblance to real-world events), rather than veridicality (statistically associated to outcome but not reflective of it based on face validity). There is evidence that these tests are more ecologically valid than traditional pen-and-paper tests, which shows promise for their continued use. Much of this research has been conducted in dementia and aging (Armentano et al., 2009; Canali et al., 2007; Green et al., 1995; Perfetti et al., 2010; Robertson et al., 1994; van der Leeuw et al., 2017) and pediatric populations (Barkley, 1998), however. Few studies have been conducted on head-injured populations and even fewer on samples with mTBI (Chan et al., 2000; Erez et al., 2009). Moreover, some have questioned the quality of verisimilitude of the subtests as compared to traditional pen-and-paper neuropsychological test measures (Acker, 1990; Alderman et al., 2003).

In an effort to maximize verisimilitude, naturalistic assessments such as the Multiple Errands Test have been developed. Here, some inspiration may have been taken from the functional assessments that are conducted by occupational therapists. In these assessments, occupational therapists' comment on a patient's capacity (the limits of their abilities) and habits (the activities that they typically engage in) through demonstration of their targeted activities (for a review, see Rogers & Holm, 2016). In naturalistic assessments such as the Multiple Errands Test, the assessment takes place in a real-life environment such as a mall or hospital setting. The participant is given a list of tasks to complete (e.g., write and mail a letter, buy a shirt) and is monitored by the staff. Such tasks are thought to involve the dynamic interplay between cognitive domains in the real-world with sensitivity to aspects of attention and executive functions (Burgess et al., 2006). While some studies have reported that naturalistic assessments are uniquely sensitive to functional impairment as compared to traditional pen-and-paper based tests (Alderman et al., 2003; Clark et al., 2017; Dawson et al., 2009; Shallice & Burgess, 1991), a recent review concluded that sensitivity to functional impairment was limited and that more research was needed to validate this type of assessment in order to make recommendations for clinical use (Robertson & Schmitter-Edgecombe, 2017). Moreover, naturalistic assessments are also limited by way of administration costs, length of assessment, burden to physical resources, and lack of standardization within (i.e., distractors and external stimuli) and between assessment sites (Robertson & Schmitter-Edgecombe, 2017).

The equivocal nature of these findings underscores that better and more accurate tools are clearly needed to determine functioning secondary to neurocognitive changes following an mTBI. This is especially true in forensic evaluations of patients with mTBI where financial compensation is often dependent on the clinician's estimate of functional capacity.

Computerized VR-based assessments were developed in an effort to maximize on the above noted advantages of ecologically oriented tests and naturalistic assessments whilst minimizing their limitations. With the advancement of the technological revolution, VR-based assessments are becoming increasingly sophisticated, decreasing in cost, and strike a balance between simulating a real-life situation whilst controlling every aspect of stimulus delivery in a semi-naturalistic manner. It is somewhat poetic to think that the very aspect that caused the previous paradigm shift in neuropsychology, advancement in (neuroimaging) technology, can also be the solution.

Marcotte & Grant (2010) recommended that future studies foster the development and implementation of new measures with greater ecological validity. VR has been announced by the National Academy of Neuropsychology, researchers in the field, and industry as the future of neuropsychological assessment (Bilder, 2011; DMW, 2004; Norcross et al., 2002; Parsons & Kane, 2017). Accordingly, the present set of studies were undertaken in an effort to review, compare, and determine the ecological validity of traditional pen-and-paper, ecologically oriented, and two novel VR neuropsychological test measures in an effort to articulate predictive validity of RTW outcome in post-acute mTBI. The overall aim of these studies was to improve the ecological validity of neuropsychological assessment in patients with mTBI who report persistent symptoms past the expected recovery period. Before expanding on the implications of the findings, a brief summary of the main results for each study will follow.

6.1 Summary of Studies

The first study (Chapter 2) provided an up to date meta-analysis of the sensitivity of traditional pen-and-paper neuropsychological test measures in a sample of patients in the post-acute period of recovery following a mTBI using a random-effects model. The novel aspect of this study was the inclusion of only studies with patients in the post-acute (over 90 days) recovery period following the date of injury. We hypothesized that effect sizes on neuropsychological test performance would be nonsignificant from controls in the post-acute period of mTBI. This was derived from previous meta-analytic reviews that indicate minimal differences between controls with 95% confidence intervals that overlap with zero. A thorough literature search was undertaken and 31 studies comprising a total of 1469 patients with mTBI and 4281 controls were included in the analysis. Overall, 316 effect sizes were extracted from all primary studies, and they were grouped into the following cognitive domains: Global Cognitive Ability, Attention & Psychomotor Speed, Executive Functions, Fluency, Acquisition Memory, Delayed Memory, Language, and Visuospatial Ability.

The overall estimated mean effect sizes were small to moderate (g = -0.35, SD = 0.156) but statistically significant from zero ($p < 0.001$). By cognitive domain, mean effect size estimates were as follows: global cognitive ability (g = -0.42, SD = 0.351); attention & psychomotor speed (g = -0.30, SD = 0.207); executive functions (g = -0.23, SD = 0.230); fluency (g = -0.61, SD = 0.336); acquisition memory (g = -0.42, SD = 0.285); delayed memory (g = -0.32, SD = 0.239); language (g = -0.73, SD = 0.396); and visuospatial ability (g = -0.22, SD = 0.293). The largest effects were found in the domains of language and fluency, whereas executive functions and visuospatial ability were calculated to have the smallest effects. Multiple comparisons were conducted, however most of the estimates of the mean effect size across all eight domains were statistically different from zero at $p < 0.001$ with the exceptions of language ($p < 0.01$) along with global cognitive ability and visuospatial ability ($p < 0.05$). Moreover, heterogeneity statistics confirmed the use of a random effects model (Q = 1,368.1, $p < 0.001$; $\tau^2 = 0.125$).

The results of this study are in contrast to previous quantitative reviews that reported there were no differences in neurocognitive performance between mTBI and controls in the post-acute period of recovery (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005; Rohling et al., 2011). In the current study, all the effect size estimates were small to medium and statistically significant. However, there is still considerable overlap between the neuropsychological profiles of mTBI and controls. For example, in the domain with the largest effect size observed (language), there is a 57% overlap of scores obtained by the mTBI and healthy control groups. To be clinically meaningful, there should be minimal overlap (~5% or less) of scores between healthy controls and patients (Zakzanis et al., 1999). As such, it was concluded that although the effect sizes estimates were statistically significant, they are not practical or meaningful with respect to clinical decision making.

The second study (Chapter 3) examined the ecological validity of neuropsychological test measures in the post-acute period following an mTBI. Here, we explored the veridicality approach in determining ecological validity. The specific questions that were addressed were (1) whether traditional pen-and-paper neuropsychological test measures could differentiate between RTW status in sample of patients in the post-acute phase of recovery following mTBI; and (2) if so, which specific tests were most sensitive? Based on the results from the first study, it was hypothesized that performance on tests of language and verbal fluency should differentiate the two groups. Ecological validity was determined by examining the relationship between test performance and RTW status. The sample consisted of 46 patients who were in the post-acute period of recovery following mTBI. Of these patients, 13 were employed and 33 were unemployed. The following tests were administered in this study: RSAT, TMT, Digit Span, CVLT-II, RCFT, BVMT, COWAT, Vocabulary, and BNT. A logit regression analysis was conducted to determine if neuropsychological test performance could predict employment status. The results indicated that performance on none of the neuropsychological tests significantly predicted employment status. Overall, the findings from the second study suggested partial evidence, at best, for the veridicality approach in predicting RTW in a sample who were in the post-acute phase of recovery following a mTBI.

Study 3 (Chapter 4) examined the ecological validity of the veridicality and verisimilitude approach in predicting RTW status. Traditional pen-and-paper neuropsychological test measures were employed for their inherent veridicality approach, and the BADS was used for its verisimilitude approach. Similar to the second study, RTW status was used as an outcome measure. The specific aims were to determine if the BADS was more sensitive to RTW status than were traditional pen-and-paper neuropsychological test measures, and if so, which subtests from that battery were most predictive. Prior investigations with the BADS in mTBI samples were lacking, but it was hypothesized that the BADS would be more predictive of RTW status than traditional pen-and-paper neuropsychological test measures. The sample consisted of 102 participants who were in the post-acute period of mTBI; 30 of which who were employed and 72 were unemployed. The specific measures that made up the traditional tests were WCST, TMT, TOL, Similarities, and Matrix Reasoning (from the WASI). Five subtests from the BADS were administered; Rule Shift Cards, Action Program, Key Search, Zoo Map, and Modified Six Elements. Non-parametric comparisons were conducted, and family-wise error was controlled via a Bonferonni corrected alpha. The probability estimate of all comparisons were above the new threshold, and thus no measure was statistically significant. Effect size estimates indicated that the BADS produced larger effect sizes as compared to all other measures, however. Specifically, the Modified Six Elements, Zoo Map, and Key Search demonstrated the greatest effect sizes ranging from d = 0.53 to 0.86. Medium effect size estimates were found for the timed aspects of the TOL but small effect sizes were garnered for the move scores. This pattern suggested that aspects of attention such as processing speed rather than planning ability may have been implicated. This was further supported by the medium effect sizes found on the TMT A and B time scores. Overall, the findings of this study showed that the BADS was more sensitive to RTW status than the majority of traditional pen-and-paper neuropsychological test measures. Some traditional pen-and-paper neuropsychological tests of attention and executive functions, such as the timed aspects of the TOL and the TMT A and B, were also found to be predictive by way of effect size analysis, however. Overall, the hypothesis was supported, and it was concluded that the verisimilitude approach was more predictive of RTW status than the veridicality approach.

The fourth study (Chapter 5) was undertaken to investigate the ecological validity of novel VR tasks in comparison to the *ecologically oriented* tests (i.e., BADS & TEA) and traditional pen-and-paper neuropsychological test measures. The collective findings from the previous studies undertaken illustrate that the verisimilitude approach was more sensitive to RTW status than the veridicality approach. Moreover, partial support for the sensitivity of executive and attention tests was also found. Thus, the aim of this study was to develop and validate two novel VR-based assessments of attention and executive functions that were high in verisimilitude and to compare the RTW sensitivity and specificity of these tasks with traditional and ecologically oriented tests of attention and executive functions. It was hypothesized that the VR tasks would show a statistical relationship with their traditional test counterparts demonstrating good concurrent validity. In addition, it was hypothesized that the VR test measures would be more sensitive of RTW status than ecologically-oriented tests, which would be more sensitive than traditional pen-and-paper neuropsychological tests. Participants included 50 patients in the post-acute period of mTBI who continued to report persistent symptoms at the time of assessment. Of the sample, 23 were employed to some capacity and 27 were unemployed. The following tests were administered: RSAT, SDMT, TMT, Map Search, Elevator Counting, TOL, WCST, Rule Shift Cards, and the Modified Six Elements. Moreover, a VR-based test of executive functions called the Office Task was administered along with a VR-based test of attention called the CBT. The results indicated good concurrent validity for the VR Office Task for measuring error but not for target responses, whereas good concurrent validity was found throughout all modules of the CBT. Ecological validity was evaluated via a step-wise discriminant functions analysis and means comparison with effect size analysis. Of the 54 predictors entered into the discriminant functions analysis, only two predictors (CBT Attention Shift difficulty level 2 and to a lesser extent RSAT total speed) were statistically significant and able to predict group membership. This model was able to predict group membership at a 82% accuracy, which was supported by a split-half cross-validation follow-up analysis. Sensitivity of this model was found to be 82.6% and specificity was 81.5%. Effect size analysis also corroborated these results. It was found that the CBT was the most ecologically valid test measure, followed by ecologically oriented tests of attention, followed by traditional pen-and-paper neuropsychological tests of attention. The VR Office Task did not predict group membership nor was it sensitive to RTW status. In addition, traditional and ecological oriented measures of executive functions also produced null results. The conclusion drawn from these findings was that across all types of tests included in this study, performance on tests of attention were sensitive to RTW status in this sample but executive functioning was not. In addition, the VR attention test (CBT) was more ecologically valid than the ecologically oriented test of attention (TEA) and traditional pen-and-paper neuropsychological test measures of attention. These findings provide further support to the superiority of the verisimilitude approach over veridicality and the use of advanced technology in determining outcome defined as RTW.

6.2 On the Role of Attentional Functioning After mTBI

Impairments in speed of information processing has been argued as the core cognitive deficit following TBI (Stuss et al., 1989). Attention and specifically the aspect of response conflict or increased time needed to process information and ignore distractions (i.e., cognitive load) was consistently found to discriminate between mTBI and controls (as per Study 1) and employment status in patients with post-acute period mTBI (Studies 2, 3, 4). Problems with attention have long been known to be a subjective and objective complaint following TBI (Cicerone, 1996; van Zomeren & Brouwer, 1994) and have been associated with poor RTW (Ruff et al., 1993; Wehman et al., 2017). Few studies have examined this phenomenon in the post-acute period of recovery following mTBI, however. Ruffolo and colleagues (1999) administered reaction time measures and a brief battery of neuropsychological test measures to patients with mTBI within one-month post-injury and then again between 6-9 months post injury. They separated the groups into those who had returned to work and those who had not. These researchers found no differences across any measure, however, a test of attention (Paced Auditory Serial Addition Test) approached significance ($p = 0.06$). The authors suggested that measures of attention and speed of information processing could predict RTW in the post-acute period of recovery. Moreover, they go on to state that “such measures could be used to identify those with continued mTBI symptoms” (p. 396). The findings of the current set of experiments support this conclusion and extends it to a litigating population.

It is common for patients of mTBI to self-report feeling slowed, mentally fatigued, and not as sharp (Alves et al., 1993). Non-specific as these statements may be, clinically, they may be interpreted as disruptions in attention and information processing speed (Vanderploeg, Curtiss, Luis, & Salazar, 2007). However, such deficits are usually undetected on standard neuropsychological test measures, especially long after injury (Belanger et al., 2005; Binder et al., 1997; Frencham et al., 2005; Rohling et al., 2011). If a genuine deficit exists, then traditional measures may not be sensitive enough to measure them. To this end, Ozen and Fernandes (2012) examined the long-term effects of mTBI on accuracy and information processing speed on an experimental high and low-load attentional task. The Repetition Detection task (Bopp & Verhaeghen, 2007) was used. This task consisted of a high and low load condition. In the low load condition, participants had to visually identify repeated digits in a random string. For example, in the string ’96 73 93 70 93 78 47 98 27’ one would identify that ‘93’ was repeated. The high load condition was similar but with the additional rule that the repeated numbers must be within the same-coloured enclosed square. Thus, the high load condition required additional attentional demand. This test, in addition to standard neuropsychological test measures such as the digit span, TMT, CVLT-II, and computerized Stroop task were administered to 26 undergraduates who had self-reportedly sustained a single mTBI in their remote past and 31 healthy controls. These researchers found that the mTBI group took significantly longer in both the high and low load conditions. Further, they reported that accuracy between the groups was similar in the low load but in the high load condition, the mTBI group outperformed the healthy controls. The authors suggested that decreased processing speed was a lasting consequence of mTBI and that slowing down gave the unexpected benefit of allowing this group to be less susceptible to distracting information on the task. No differences were found between the groups on any of the neuropsychological test measures administered, suggesting that the findings on the high and low-load attentional task may be related to a higher-order type of construct rather than simple processing speed. The researchers conclude that higher order processing is a long-term consequence after even a single mTBI but that accuracy is spared.

The results from Study four are congruent with the findings of Ozen and Fernandes’ (2012) and others (Malojcic, Mubrin, Coric, Susnic, & Spilich, 2008; Ponsford & Kinsella, 1992). Total speed on the RSAT was found to be a predictor between RTW status in mTBI but not RSAT accuracy rates, as per the discriminant functions analysis in Study 4. The CBT Attention Shift performance score, which is a divided attention task with a distractor, was also found to be a significant predictor. Errors on this task, however, were not statistically significant and were not a part of the final model. Further evidence from Study 4 can be found from the effect size analysis. Performance, rather than errors, generally produced larger effect sizes separating the two employment groups.

Further support of slowed information processing but accurate performance has been published in detailed case studies. For example, it has been reported that multitasking within a set time-constraints increases cognitive load, which is processed at slower rates after TBI, impacting work productivity. However, accuracy was maintained despite the slowed production (Bootes & Chapparo, 2010).

There is also evidence from neuroimaging studies to corroborate slowed information processing speed following TBI. Sharp and colleagues (2011) employed functional magnetic resonance imaging (fMRI) to examine abnormalities of functional connectivity following mild to moderate TBI using attention and speed of processing tests. Their results illustrate that patients were slower and produced variable responses as compared to controls but were accurate. Their findings are congruent with the results of the current set of investigations. Whilst timed tests of attention and processing speed were significantly poorer in the unemployed group as compared to the employed group, there was no difference with respect to accuracy or the number or errors. Moreover, Sharp et al., (2011) also noted that lower functional connectivity in the default mode network was observed in patients as compared to controls and related this finding to an increase in cognitive load to maintain accurate task performance. The authors reported that this finding was associated with evidence of diffuse axonal injury and connectivity disruption within the adjacent corpus callosum. Similar conclusions have been found on other studies (Mayer et al., 2012). It should be noted, however, that their findings were not specific to mTBI.

From an ecological validity perspective, measures of attention employed in Studies three and four were the most accurate at predicting employment outcome. These studies provide evidence for the discriminatory power of some traditional tests. Moreover, Study 4 provided support for the use of the RSAT and SDMT as significant discriminators of employment status in mTBI. This is congruent with past studies that have also found these measures to be predictive of employment outcome (Anderson, Christensen, Kirkevold, & Johnson, 2012; Ruff et al., 1993; Yamout et al., 2013). Ecologically oriented tests of attention (subtests from the TEA) were only examined in Study 4, but these findings also provide support for their ability to discriminate between employment status. To this end, our findings provide preliminary support for the ecological validity (as measured by RTW status) for the use of the Elevator Counting and Map Search test in a post-acute sample of patients who sustained a mTBI. It is not clear if these results will translate in other areas of outcome. As such, additional research will be required to explore the extent to which performance on the TEA can predict other outcomes (e.g., iADLs such as cooking, cleaning, driving, managing personal finances, etc.).

Collectively, the results from the current set of studies support several premises. First, persistent symptoms of slowness in the post-acute period of mTBI (i.e., > 90 days) have been reported to affect functional outcome. Second, while these symptoms may not be corroborated on most traditional pen-and-paper neuropsychological test measures, measures such as the RSAT timed scores, but not errors, are sensitive enough to distinguish RTW status (Study four). Third, the effect size estimates from Study One and the results from Study four demonstrate that some traditional pen-and-paper tests of attention, ecologically oriented tests of attention, and attention modules of the CBT predicted RTW in mTBI. Both approaches of ecological validity, veridicality and verisimilitude, were found to be sensitive to RTW, however, based on the effect sizes and percent non-overlap values, tests of verisimilitude were found to be more sensitive and accurate at predicting group membership. Based on these data, it is concluded that the verisimilitude approach is more ecologically valid to RTW outcome following mTBI than the veridicality approach. The clinical implication of these findings is that ecologically oriented tests and especially VR tests such as the CBT may hold potential to answer the common referral question of functional outcome following mTBI.

6.3 On the Role of Executive Functioning After mTBI

Poor performance on tests of executive function was also found on some of the current studies, albeit less consistently as compared to performance on tests of attention. Despite a small effect size estimate of g = -0.23 in Study One between healthy controls and post-acute mTBI, this absolute value was consistently reported across previous meta-analyses. The results of Study 2 did not find any traditional pen-and-paper executive measure to be a significant factor in RTW outcome; however, this may have been due to the relatively small sample size and limited number of executive tests that were administered. Moreover, additional support from Study 3 showed that ecologically-oriented tests of executive dysfunction (i.e., BADS) demonstrated medium to large effect sizes. Study 4, however, found mixed results. Although the subtests from the BADS did not significantly differentiate the employed and unemployed groups, the effect size estimates were in the small to medium range, suggesting some discriminating power. Similar results were observed on the VR Office Task. Medium effect sizes were found for the number of correct deliveries between the employed and unemployed groups; however, this failed to meet the threshold for statistical significance. Moreover, the concurrent validity results suggested that the VR Office Task had good concurrent validity at detecting errors (as compared to other gold standard tests of executive dysfunction) but not for hits. Correct deliveries did not demonstrate good concurrent validity and thus may have been measuring a different construct than expected. It is posited that based on the face validity of the task, correct deliveries may have been measuring some aspect of sustained attention. This aspect may not be sensitive enough to differentiate the groups in an expected manner as the means of the number of correct deliveries from the employed group was lower than the unemployed group. Such a comparison of means was not statistically significant, but the observed trend was unexpected nonetheless. Since some sub-measures were aligned across traditional and VR tests of executive functions but others were not, this provides some evidence of *fractionation* (separation of different executive sub-processes) of the executive system (Shallice & Burgess, 1996).

Overall, findings from the present set of experiments provided mixed results with respect to the sensitivity of executive tests with RTW outcome in a sample of patients with mTBI who were in the post-acute period of recovery. Based on the findings from Study 3 and previous studies investigating the ecological validity of the BADS, the clinical reliability of the BADS needs to be further investigated. The relationship between clinical neuropsychology and executive functions has an ambiguous and sometimes conflicted history (Stuss & Levine, 2002). A potential explanation for why executive tests were not as reliable as measures of attention in the present set of experiments is that the concept of executive functions is complex, the definition of what subprocesses constitute it are debatable, and arguably, there is evidence that it controls nearly all aspects across cognitive domains. Logically, this would mean that activity from other domains influence its functioning.

As such, executive functions are more of an umbrella term that comprises a wide range of cognitive processes (see Damasio, 1995; Stuss & Benson, 1986). This has been an area of difficulty for neuropsychologists to develop and measure due to the varying operational definition of what constitutes an executive function and the multimodal involvement that may influence performance. For example, some researchers consider the term working memory to be a factor under attention, whereas others have argued that it is fractionated under executive functions. Accordingly, tests such as the SDMT and TMT B have been classified in both domains (Karr, Arenenkoff, & Garcia-Barrera, 2014; Pertab et al., 2009). Moreover, the majority of neuropsychological test measures of executive functions are aimed at measuring the behavioural manifestations of the dorsolateral region of the prefrontal cortex – an area with strong empirical evidence for its connection with executive functions. However, the ventral region of the prefrontal cortex is also involved in aspects of what would be considered executive function, such as emotional processing, inhibition, and self-regulation (Rolls, 2000). Arguably, these aspects are quite important to human behaviour yet are not routinely measured by neuropsychological assessment (Stuss & Levine, 2002). Hence, it is possible that a significant component of executive functions is missed by neuropsychological evaluations. To this end, neuropsychologists may be speculating on one's strengths and weaknesses with incomplete data. This topic is beyond the scope of the current work, but it may be a limiting factor of the design and implementation of the VR Office Task. Future iterations of this VR task would need to consider this limitation and implement modifications, such as socialization and interaction with other people and avatars.

6.4 Limitations

The findings from this set of studies are novel and provide an important contribution to the extant literature. However, they are not without limitations.

While the meta-analysis in Study One (Chapter 2) was thorough, inherent limitations to the use of a meta-analysis were present. For example, combining samples from multiple studies produces a heterogeneous sample with characteristics, such as patient recruitment, and variable test sensitivity and specificity values. Further, an analysis of moderating variables was not undertaken due to considerable variability of reporting across studies (as also experienced by Pertab et al., 2009) and lack of reporting standardization of such variables (Maas et al., 2017). In addition, although the inclusion criteria for this study was comprised of patients in the post-acute period of mTBI, the range of duration since injury was 90 days to 16 years. To this end, studies that included patients with mTBIs in the remote past may disproportionately include those with persistent symptoms that are not related to cognition (i.e., psychosocial, mood disturbance, pain, physical injuries, other behavioural manifestations, etc.).

Regarding sample size, the collective studies undertaken herein were comprised of relatively small and unequal sample sizes. This limited the power of the findings and the types of inferential statistics that were employed. Ideally, a larger sample size would be needed to employ multivariate statistics such as structured equation modelling to estimate multiple and interrelated dependent variables to RTW in a single analysis.

An additional limitation is the use of an archival database of patients with mTBI that were involved in litigation (Studies 2 & 3). The analyses employed in these studies allowed for the classification of *current* RTW status based on *current* functioning. A much more useful procedure would be to determine *future* RTW status based on *current* functioning. To achieve this, follow-up studies could recruit non-litigating participants consecutively through a tertiary care centre followed longitudinally (Cancelliere et al., 2017).

In addition, although the test measures were administered by a psychometrist under the supervision of licensed neuropsychologist, strict experimental control was not possible. To this end, the clinic from where the database was collected utilized a combination of the flexible battery and fixed battery approach in an effort to minimize fatigue and assessment duration (Rabin et al., 2016; Sweet et al., 2015). As a result, some test measures were not administered to some patients. Ideally, all test measures would be administered to all patients allowing for full and direct comparisons. In a clinical research setting, however, it is reasonable that some patients would not complete all test measures due to situational factors such as fatigue, time restrictions, or patients request to skip a test measure. Future replications of these studies should aim to include only patients that have completed all test measures of interest.

The use of a convenience sample of litigating patients with mTBI is another limitation (Study 4), as it limits the generalizability of the findings. Most of the sample were involved in litigation, and prior studies have shown that there is a higher rate of unemployment in this sample as compared to non-compensation-seeking patients with mTBI (Friedland & Dawson, 2001). Further, recovery has been shown to be affected by compensation (Cassidy et al., 2004). The generalizability and reliability of the current findings needs to be validated by replication in a non-litigating sample.

RTW outcome was dichotomous and was not stratified (e.g., by level of engagement of duties such as part-time or full-time and modification of hours and/or duties), which may have provided additional data. Further, there is evidence to support that job independence and decision-making latitude in the workplace is a predictor for good RTW outcome (Friedland & Dawson, 2001). Although RTW status was garnered by way of self-report and coupled by way of the patient's medical file and income compensation benefit status, evaluation of RTW status could have been improved by an occupational therapist's evaluation, review of employment records, and/or review of the patient's work environment. Such categorization, such as the use of the National Occupational Classification (Human Resources and Development Canada, 2006), could have been undertaken in Study 4. This would have reduced group sample sizes limiting analysis to descriptive statistics, however.

Another limitation of the current set of studies was that mood was not formally assessed and accounted for statistically as a moderator of cognitive performance or outcome. Elevated symptoms of depression have been shown to be associated with slowed information processing (Ozen & Fernandes, 2012) and some have indicated that it is one of many factors that may affect RTW after a mTBI (Dikmen et al., 1994; Ruffolo et al., 1999; Schultz, Law, Cruikshank, 2016). In systematic review of the factors affecting RTW after mTBI, Cancelliere and colleagues (2014) noted that those who were unable to return to gainful employment experienced greater physical ailments, depression, and anxiety. Those who were able to RTW were in better physical health and reported experiencing fewer symptoms of depression and anxiety. However, a more recent review by Schultz and colleagues (2016) cautioned against the robustness of these findings by stating that "evidence of predictors of disability among workers with depression, anxiety, and PTSD is emerging but is no by means strong" (p. 192). Replication using studies with improved methodologies was recommended before clinically meaningful recommendations could be derived.

A universal limitation of this type of research was that the measures administered across Studies 2 and 4 do not measure what the participant actually does in the real-world. They measure what they could do as per the instructions and bounds of the test in an effort to predict real-world behaviour. The VR tasks were not truly reflective of the real-world because the real-world is complex and people have free will and volition that contribute to the limited direction from external sources to influence their behaviours. True verisimilitude, however, is not possible (Goldstein, 1996) and developing a situation where participants have control to demonstrate their own structure is difficult and somewhat paradoxical (Lezak et al., 2012). Future advancements in VR technology, however, may allow for this to occur.

A further limitation to the CBT was the observation of ceiling effects on some of the difficulty levels in the Sustained Attention module. The parameters for these difficulty levels were devised from a pilot study that utilized healthy undergraduate controls. On this module, the parameters did not generalize and future iterations would require modifications of the stimulus delivery speed and frequency. It is difficult to strike the balance of making a task simple enough to be understood quickly whilst challenging enough to avoid ceiling effects (Marcotte & Grant, 2010). Aside from the Sustained Attention module, the results from the difficulty level analysis in Study 4 demonstrated a reasonable balance. A ceiling effect was observed on the VR Office Task as well, with participants indicating that it was mundane and too easy. Improvements to category switching and increased difficulty levels may improve the performance distribution of this task. For example, the doors could be unmarked and finding the correct door could be reached by trial and error. Furthermore, the correct door could alternate every ten consecutively correct deliveries. In essence, these two recommendations were inspired by the WCST category switches and the Iowa Gambling Task (Bechara, 2007) correct card deck. In turn, it may improve engagement as well. Lastly, since timed measures of processing speed were robustly linked to RTW status in the attention measures in Study 4, it could perhaps be implemented in the VR Office Task by implementing a time limit to deliver the packages.

The VR measures were limited not by way of imagination but rather budget. Future modifications could improve the realism of the task with upgraded graphics and transcoding it into a better platform (i.e., Unity 3D, https://unity3d.com/). Immersion into the task could be improved by using a head mounted display, such as the Oculus Rift (https://oculus.com/) and headphone sound. On face value, it stands to reason that it would be beneficial for immersion and thereby improve verisimilitude of the task, but it has been argued that graphics do not play a role in immersion as much as the actual activities that are carried out in the VE (Rizzo et al., 2005). Further, using additional peripherals would likely detract from its scalability and potential use by clinicians, as it would add cost and maintenance factors. As noted above, adding social interactions or distraction components to the VR Office Task could tap into the VPFC aspects of executive functioning, thereby improving its concurrent validity and potential ecological validity (Alderman et al., 2003; Stuss & Levine, 2002). In addition, multiple parallel versions of the CBT and VR Office Task could be developed in an effort to be utilized in rehabilitative contexts where multiple administrations would be necessary.

The results of Study 4 and other studies assessing the construct validity and sensitivity of VR applications in detecting neurocognitive deficits have shown promising results (Cipresso et al., 2014; Elkind et a., 2001; Jansari et al., 2014; Jovanovski et al., 2012; Koenig, Crucian, Dalrymple-Alford, & Dünser, 2010;

Lengenfelder et al., 2002; Man et al., 2016; Nir-Hadad, Weiss, Waizman, Schwartz, & Kizony, 2015; Parsons & Rizzo, 2008; Rand et al., 2009; Rizzo et al., 2006). The field in general will need to agree on standardized research methodologies and standardized outcome measures of clinical VR tools, as well as titrating VR-specific parameters like frequency and modality of stimuli and cues, complexity of VR tasks, immersion by way of user interfaces, graphical realism, and the gamification of the task (Parsons, McMahan, & Kane, 2018; Rizzo & Koenig, 2017). Further research and increased standardization will be required before clinical VR applications mature enough to be used by the practicing clinician, however.

Fortunately, the use of VR in clinical applications is in its infancy, suggesting much progress is possible ahead. The general availability of VR technology has only really existed for about 25 years. Ever changing and maturing hardware and software systems produced one-off versions of VR applications with variable results for the first 10-15 years (Rizzo & Koenig, 2017). Surveying the literature on clinical applications of VR technology, a staple comment usually involved something to the effect of *these VR results are promising but further research with larger sample sizes and different populations is needed*. The initial development of VR technology during this period was quite costly and time consuming. Recently, however, general VR development has gained in popularity driven by access to affordable design tools and VR hardware. Head mounted VR hardware is as low as $7 USD (Google Cardboard, Google Inc., 2018) and VR development software can be found free of charge (e.g., Unity3D, Unreal Engine, Amazon Lumberyard). In the last five years, we have seen that the demand for VR development has increased such that high schools and computer science curriculums have adopted this as a course (Rizzo & Koenig, 2017). In addition, Goldman Sachs (2016) conducted an analysis for VR and the different markets it could serve. Unsurprisingly gaming was the largest but healthcare was a small but growing market for the use of this technology. Steam (Valve Corporation, 2018) is an online platform for the distribution of games with over 1900 VR-enabled games. Such a distribution solution for clinically-related VR software could be feasible for clinicians and researchers. In sum, the decreasing cost of VR hardware, virtually no cost for VR software, increasing VR training programs, favourable and growing VR healthcare outlook from market surveys, popularity of VR games to expand its interest, and promising results from the use of VR in neuropsychological studies has created the momentum and support to the argument that VR is the future of neuropsychology (Bilder et al., 2011; Norcross et al., 2002; Parsons & Kane, 2017; Rizzo & Koenig, 2017).

6.5 Conclusions

While there were many aspects of mTBI functional outcome that the current research could have addressed (e.g., iADLs, vocational, recreational, social disability), using a multitude of modalities (e.g., mood disturbances, cognitive symptoms, pain, other behavioural manifestations, etc.), RTW was chosen to be studied as RTW is a general indicator of real-world functioning (Ownsworth & McKenna, 2004). To a degree, it transcends culture and other barriers; to live in society, one must work. “Work organizes life. It gives structure and discipline to life” (Clinton, 1993, n.d.). Indeed, the assessment of mTBI contributes to a major proportion of private practice referrals, with the most common question relating to capacity and prognosis of RTW (Rabin et al., 2005; 2012). This is also an important indirect indicator of a patient’s overall quality of life, as those who return to some sort of gainful employment report improved health, well-being, integrate better in social and community settings, are less dependent on external health and social services, and have an overall greater quality of life than do those who are unemployed (Corrigan et al., 1997; Steadman-Pare et al., 2001; Wehman et al., 2005). Reliably identifying predictors that are associated with poor RTW outcome allows those at risk to benefit from rehabilitation programs that may help them RTW as quickly as possible. This focuses our resources to those that actually need it, saving healthcare costs (Chen et al., 2012). Since the current methods of assessment show equivocal relationships with RTW following mTBI (see Studies 2-4), then the development, validation, and implementation of new measures with improved ecological validity is required (Alderman et al., 2003; Burgess et al., 2006; Dawson et al., 2009; Heaton & Pendleton, 1981; Jovanovski et al., 2012; Marcotte & Grant, 2010; Rizzo et al., 2000; 2004; Schultz et al., 2016; Schultheis et al., 2002; Tippett et al., 2009). Hence, the development of the VR measures used in Study 4. The current set of findings provide evidence that ecologically oriented tests such as the BADS and TEA are slightly better than traditional pen-and-paper neuropsychological test measures at determining RTW rates, but the CBT was the most predictive. The findings from this work benefits clinicians by providing research-based evidence of the ecological validity of traditional and ecologically oriented tests. Moreover, the findings from this work introduced novel neuropsychological tools with better sensitivity and specificity with respect to RTW. The natural next step would be to validate our findings with larger and broader samples. The corroborating evidence from future studies would provide the *evidence* that is demanded in *evidence*-based practices (Ruff, 2003; Schultz et al., 2016).

Review into the time lag between research findings and clinical implementation illustrates that it takes approximately 17 years for recommendations/findings to translate into clinical practice (Morris, Woodling, & Grant, 2011). It has been argued that VR is the future of clinical neuropsychology (Bilder, 2011; DMW, 2004; Norcross et al., 2002; Parsons & Kane, 2017). The results from the current series of studies demonstrates support for the use of VR as a testing method that is ecologically valid, adding support to the extant literature. Given the collective support for the implementation of VR, it is ready to place diagnostic clinical decision making on firmer scientific grounds.

References

Acker, M. B., & Davis, J. R. (1989). Psychology test scores associated with late outcome in head injury. *Neuropsychology, 3*, 123-133.

Acker, M. R. (1990). A review of the ecological validity of neuro-psychological tests. In D.E. Tupper & K.D. Cicerone (Eds.), *The neuropsychology of everyday life: Assessment and basic competencies* (pp. 19-55). Boston, MA: Kluwer Academic Publishers.

Airey, C. M., Chell, S. M., Rigby, A. S., Tennant, A., & Connelly, J. B. (2001). The epidemiology of disability and occupation handicap resulting from major traumatic injury. *Disability Rehabilitation, 23*, 509-515.

Alexander, M. P. (1995). Mild traumatic brain injury: Pathophysiology, natural history, and clinical management. *Neurology, 45*, 1253–1260.

Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society, 9*, 31– 44.

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *British Medical Journal*, 332, 1080.

Alves, W., Macciocchi, S. N., & Barth, J. T. (1993). Postconcussive symptoms after uncomplicated mild head injury. *Journal of Head Trauma and Rehabilitation, 8*(3), 48–59.

American Academy of Neurology. (1997). Practice parameter: The management of concussion in sports (summary statement). *Neurology, 48*, 581-585.

Andelic, N., Stevens, L. F., Sigurdardottir, S., Arango-Lasprilla, J. C., & Roe, C. (2013). Associations between disability and employment 1 year after traumatic brain injury in a working age population. Brain Injury, 26(3), 261–269.

Anderson, G., Christensen, D., Kirkevold, M., & Johnson, S. P. (2012). Post-stroke fatigue and return to work: A 2-year follow-up. *Acta Neurologica Scandinavica, 125*, 248-253.

Ao, B., Tobias, M., Ameratunga, S., McPherson, K., Theadom, A., Dowell, A., et al. (2015). Burden of traumatic brain injury in New Zealand: Incidence, prevalence and disability-adjusted life years. *Neuroepidemiology, 44*(4), 255–261.

Arango-Lasprilla, J. C., Ketchum, J. M., Gary, K. W., Kreutzer, J. S., O’Neil-Pirozzi, T. M., Wehman, P., et al. (2009). The Influence of Minority Status on Job Stability After Traumatic Brain Injury. *PM and R*, *1*(1), 41–49.

Armentano, C. G., Porto, C. S., Brucki, S. M., & Nitrini, R. (2009). Study on the behavioural assessment of dysexecutive syndrome (BADS) performance in healthy individuals, mild cognitive impairment and Alzheimer's disease. *Dementia & Neuropsychologia, 3*(2), 101-107.

Ashley, M. J., Schultz, J. D., Bryan, V. L., Krych, D. K., & Hays, D. R. (1997). Justification of postacute traumatic brain injury rehabilitation using net present value techniques: A case study. *Journal of Rehabilitation Outcomes Measurement, 1*(5), 33-41.

Asikainen, I., Kaste, M., & Sarnas, S. (1998). Predicting late outcome for patients with traumatic brain injury referred to a rehabilitation programme: A study of 508 Finnish patients 5 years or more after injury. *Brain Injury, 12*(2), 95–107.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Ball, C. G. (2015). Penetrating nontorso trauma: the head and the neck. *Canadian Journal of Surgery, 58*(4), 284–285.

Bannon, S., Gonsalvez, C. J., Croft, R. J., & Boyce, P. M. (2006). Executive functions in obsessive-compulsive disorder: State or trait deficits? *Australian and New Zealand Journal of Psychiatry*, *40*, 1031-1038.

Banville, F., Nolin, P., Lalonde, S., Henry, M., Dery, M. P., & Villemure, R. (2010). Multitasking and prospective memory: Can virtual reality be useful for diagnosis? *Behavioural Neurology*, *23*(4), 209–211.

Baribeau, J., Ethier, M.A., & Braun, C. (1989). A neurophysiological assessment of selective attention before and after cognitive remediation in patients with severe closed head injury. *Journal of Neurologic Rehabilitation*, 3(7), 1-92.

Barkley, R. A. (1998). *Attention Deficit Hyperactivity Disorder*. New York: Guilford Press.

Barkley, R A., Murphy, K. R., O'Connell, T., Anderson, D., & Connor, D. F. (2006). Effects of two doses of alcohol on simulator driving performance in adults with attention-deficit/hyperactivity disorder. *Neuropsychology*, *20*(1), 77-87.

Barr, W.B. (2001). Methodological issues in neuropsychological testing. *Journal of Athletic Training, 36*(3), 297-302.

Barwood, C. H. & Murdoch, B. E. (2013). Unravelling the influence of mild traumatic brain injury (MTBI) on cognitive-linguistic processing: a comparative group analysis. *Brain Injury, 27*, 671-6

Bate, A. J., Mathias, J. L., & Crawford, J. R. (2001). Performance on the Test of Everyday Attention and standard tests of attention following severe traumatic brain injury. *Clinical Neuropsychologist*, *15*(3), 405–422.

Bayless, J. D., Varney, N. R., & Roberts, R. J. (1989). Tinker toy test performance and vocational outcome in patients with closed-head injuries. *Journal of Clinical and Experimental Neuropsychology, 11*, 913–917.

Bechara A. (2007). *Iowa Gambling Task (IGT) Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Beck, L., Wolter, M., Mungard, N.F., Vohn, R., Staedtgen, M, Kuhlen, T., et al. (2010). Evaluation of spatial processing in virtual reality using functional magnetic resonance imaging (fMRI). *Cyberpsychology, Behavior, and Social Networking, 13,* 211-215.

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: a meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215-227.

Belanger, H.G., & Vanderploeg, R.D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(4), 345-357.

Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test-Revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Benedict, R. H. B., Cookfair, D., Gavett, M., Gunther, M., Munschauer, F., Garg, N., et al. (2006). Validity of the minimal assessment of cognitive function of multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society, 12*(4), 549-558.

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 International Neuropsychological Society, 11,* 606 – 613.

Benton, A. L., & Hamsher, K. deS. (1989). Multilingual Aphasia Examination. Iowa City: AJA Associates.

Bernstein, D.M. (1999). Recovery from mild head injury. *Brain Injury, 13*, 151–172.

Bernstein, D. M. (2002). Information processing difficulty long after self-reported concussion. *Journal of the International Neuropsychological Society, 8,* 673–682.

Bigler, E. D. (2001). The lesion(s) in traumatic brain injury: implications for clinical neuropsychology. *Archives of Clinical Neuropsychology*, *16*(2), 95-131.

Bigler, E. D. (2003). Neurobiology and neuropathology underlie the neuropsychological deficits associated with traumatic brain injury. *Archives of Clinical Neuropsychology*, *18*(6), 595-621.

Bigler, E. D. (2004). Neuropsychological results and neuropathological findings at autopsy in a case of mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *10*(5), 794-806.

Bigler, E. D. (2008). Neuropsychology and clinical neuroscience of persistent post-concussive syndrome. *Journal of the International Neuropsychological Society, 14*, 1-22.

Bilder, R.M. (2011). Neuropsychology 3.0: Evidence-based science and practice. *Journal of the International Neuropsychological Society, 17,* 7-13.

Binder, L.M. (1997). A review of mild head trauma. Part II: clinical implications. *Journal of Clinical and Experimental Neuropsychology, 19,* 432 – 457.

Binder, L. M., Rohling, M. L., & Larrabee, G. J. (1997). A review of mild head trauma. Part 1: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*, *19*, 421-431.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*(2), 129-136.

Blennow, K., Hardy, J., & Zetterberg, H. (2012). The neuropathology and neurobiology of traumatic brain injury. *Neuron*, *76*(5), 886–899.

Bliss, J.P., Tidwell, P.D., & Guest, M.A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence: Teleoperators and *Virtual Environments, 6*, 73–86.

Boelen, D.H., Spikman, J.M., Rietveld, A.C., & Fasotti, L. (2009). Executive dysfunction in chronic brain-injured patients: Assessment in outpatient rehabilitation. *Neuropsychological Rehabilitation*, *19*(5), 625-644.

Bohnen, N., Jolles, J., Twijnstra, A., Mellink, R., & Sulon, J. (1992). Coping styles, cortisol reactivity, and performance in a vigilance task of patients with persistent postconcussive symptoms after a mild head injury. *International Journal of Neuroscience, 64*(1-4), 97-105.

Bopp, K. L., & Verhaeghen, P. (2007). Age-related differences in control processes in verbal and visuospatial working memory: Storage, transformation, supervision, and coordination. *The Gerontological Society of America, 5*, 239 –246.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-item recognition trial: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology*, *24*(5), 561–573.

Bootes, K., & Chapparo, C. (2010). Difficulties with multitasking on return to work after TBI: A critical case study. *Work*, *26*, 207-216.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2010). *Comprehensive MetaAnalysis*. Biostat, Englewood, NJ.

Borg, J., Holm, L., Cassidy, J.D., Peloso, P.M., Carroll, L.J., von Holst., H. et al. (2004). Diagnostic procedures in mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine, 43S*, 61–75.

Bowman, M. L. (1996). Ecological validity of neuropsychological and other predictors following head injury. *The Clinical Neuropsychologist*, *10*(4), 382-396.

Broglio, S. P., Pontifex, M. B., O'Connor, P., & Hillman, C. H. (2009). The persistent effects of concussion on neuroelectric indices of attention. *Journal of Neurotrauma, 26*, 1463–1470.

Broks, P., Preston, G.C., Traub, M., Poppleton, P., Ward, C., & Stahl, S.M. (1988). Modelling dementia: Effects of scopolamine on memory and attention. *Neuropsychologia, 26*, 685-700.

Brooks, N., McKinlay, W., Symington, C., Beattie, A., & Campsie, L. (1987). Return to work within the first seven years of severe head injury. *Brain Injury, 1*(1), 5-19.

Brown, D. J., Kerr, S. J., & Bayon, V. (1998). The development of the virtual city: A user centred approach. In P. Sharkey, D. Rose, & J. Lindstrom (Eds.), *Proceedings of the Second European Conference on Disability, Virtual Reality and Associated Technologies (ECDVRAT)* (pp. 11–16). Reading, UK: University of Reading.

Brunswik, E. (1955). Symposium of the probability approach in psychology: Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217.

Bryant, R. A. (2008). Disentangling mild traumatic brain injury and stress reactions. *New England Journal of Medicine*, *358*, 525-527.

Bryant, R. A., O'Donnell, M. L., Creamer, M., McFarlane, A. C., Clark, C. R., & Silove, D. (2010). The psychiatric sequelae of traumatic injury. *American Journal of Psychiatry, 167*, 312-320.

Burgess, P. W., & Alderman, N. (1990). Rehabilitation of dyscontrol syndromes following frontal lobe damage: A cognitive neuropsychological approach. In R. L. Wood & I. Fussey (Eds.), *Cognitive rehabilitation in perspective* (pp. 183-203). Basingstoke: Taylor & Francis.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H. & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society,* 4, 547-58.

Burgess, P. W., & Alderman, N., Forbes, C., Costello, A., Coates, L. M., Dawson, D. R. et al. (2006). The case for the development and the use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*, 194-209.

Burgess, P. W., Alderman, N., Wilson, B. A., Evans, J. J., & Emslie, H. (1996). The Dysexecutive Questionnaire. In B. A. Wilson, N. Alderman, P. W. Burgess, H. Emslie, & J. J. Evans, *Behavioural assessment of the dysexecutive syndrome.* Bury St. Edmunds: Thames Valley Test Company.

Burke, S. (2001). Missing values, outliers, robust statistics & non-parametric methods. *LC-GC Europe Online Supplement*, 19-24.

Burke, W.H., Wesolowski, M.D., & Guth, M.L. (1988). Comprehensive head injury rehabilitation: An outcome evaluation. *Brain Injury, 2*(4), 313-322.

Capruso, D.X., & Levin, H.S. (1992). Cognitive impairment following closed head injury. *Neurologic Clinics, 10*, 879-893.

Campbell, Z., Zakzanis, K. K., Jovanovski, D., Joordens, S., Mraz, R., & Graham, S. J. (2009). Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: An fMRI case study elucidating the neural basis of planning by comparing the Tower of London with a three-dimensional navigation task. *Applied Neuropsychology, 4,* 295-306.

Canali, F., Brucki, S.M., & Bueno, O.F. (2007). Behavioural assessment of the dysexecutive syndrome (BADS) in healthy elders and Alzheimer's disease patients. *Dementia & Neuropsychologia, 2,* 154-160.

Cancelliere, C., Kristman, V. L., Cassidy, J. D., Hincapié, C. A., Côté, P., Boyle, E., et al. (2014). Systematic review of the prognosis after mild traumatic brain injury in adults: Cognitive, psychiatric, and mortality outcomes: Results of the international collaboration on mild traumatic brain injury prognosis. *Archives of Physical Medicine and Rehabilitation*, *95*(3 SUPPL).

Cancelliere, C., Coronado, V.G., Taylor, C.A., & Xu, L. (2017). Epidemiology of isolated versus nonisolated mild traumatic brain injury treated in emergency departments in the United States, 2006-2012: Sociodemographic characteristics. *Journal of Head Trauma and Rehabilitation, 32*(4), E37-E46.

Canty, A.L., Fleming, J., Patterson, F., Green, H.J., Man, D., & Shum, D. H. (2014). Evaluation of a virtual reality prospective memory task for use with individuals with severe traumatic brain injury. *Neuropsychological Rehabilitation, 24*(2), 238-265.

Carelli, L., Morganti, F., Weiss, P. L., Kizony, R., & Riva, G. (2008). A virtual reality paradigm for the assessment and rehabilitation of executive function deficits post stroke: Feasibility study. *Virtual Rehabilitation, 2008*, 99–104.

Cassidy, J.D., Carroll, L., Cote, P., Holm, & Nygren, A. (2004). Mild traumatic brain injury after traffic collisions: A population-based inception cohort study. *Journal of Rehabilitative Medicine, Suppl. 43*, 15-21.

Cattelani, R., Tansi, F., Lombardi, F., & Mazzucchi, A. (2002). Competitive re-employment after severe TBI: Clinical, cognitive and behavioral predicative variables. Brain Injury, 16, 51-64.

Chan, R. C. K., Hoosain, R., & Lee, T. M. C. (2002). Reliability and validity of the Cantonese version of the Test of Everyday Attention among normal Hong Kong Chinese: A preliminary report. *Clinical Rehabilitation, 16*, 900-909.

Chan, R. C. K. (2002). Attention deficits in patients with persisting postconcussive complaints: A general deficit or specific component deficit? *Journal of Clinical and Experimental Neuropsychology, 24*, 1081–1093.

Chan, R. C. K. (2005). Sustained attention in patients with mild traumatic brain injury. *Clinical Rehabilitation*, *19*(2), 188-193.

Chan, R. C. K., Lai, M. K., & Robertson, I. H. (2006). Latent structure of the test of everyday attention in a non-clinical Chinese sample. *Archives of Clinical Neuropsychology, 21,* 477-485.

Chan, R. C. K., Robertson, I. H., & Crawford, J. R. (2003). An application of individual subtest scores calculation in the Cantonese version of the test of everyday attention. *Psychological Reports, 93*, 1275-1282.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181-197.

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*, 217-227

Chen, A., Bushmeneva, K., Zajorski, B., Colantonio, A., Parsons, D., & Wodchis, W. P. (2012). Direct cost associated with acquired brain injury in Ontario. *BMC Neurology, 12*, 1-12.

Christensen, B. K., Colella, B, Inness, E, Hebert, D., Monette, Bayley, M., et al. (2008). Recovery of cognitive function after traumatic brain injury: A multilevel modeling analysis of Canadian outcomes. *Archives of Physical Medicine and Rehabilitation, 89,* 3-15.

Cifu, D., Hurley, R., Peterson, M., Cornis-Pop, M., Rickli, P.A., Ruff, R. L., et al. (2009). VA/DoD clinical practice guideline for management of concussion/mild traumatic brain injury. *Journal of Rehabilitation Research & Development, 46,* CPO1-CP68.

Cifu, D. X., Keyser-Marcus, L., Lopez, E., Wehman, P., Kreutzer, J. S., Englander, J., et al. (1997). Acute predictors of successful return to work I year after traumatic brain injury: A multicentre analysis. *Archives of Physical Medicine and Rehabilitation. 78,* 125-131.

Cicerone, K.D. (1996). Attention deficits and dual task demands after mild traumatic brain injury. *Brain Injury, 10*(2), 79-89.

Cicerone, K. D. (1997). Clinical sensitivity of four measures of attention to mild traumatic brain injury. *Clinical Neuropsychologist*, *11*(3), 266–272.

CIHI (Canadian Institute for Health Information). (2007). *The Burden of Neurological Diseases, Disorders and Injuries in Canada*. Ottawa, ON: CIHI.

Cipresso, P., Albani, G., Serino, S., Pedroli, E., Pallavicini, F., Mauro, A., et al. (2014). Virtual multiple errands test (VMET): a virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Frontiers in Behavioral Neuroscience, 8*, 405.

Clarke, L. A., Genat, R. C., & Anderson, J. F. I. (2012). Long-term cognitive complaint and post-concussive symptoms following mild traumatic brain injury: The role of cognitive and affective factors. *Brain Injury*, *26*(3), 298-307.

Clinton, W.J. (1993). AZQuotes.com. Retrieved August 1, 2018, from: https://www.azquotes.com/quote/590873

Cohen, J. (1988). *Statistical power analyses for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Colantonio, A., Mroczek, D., Patel, J., Lewko, J., Fergenbaum, J., Brison, R. (2010). Examining occupational traumatic brain injury in Ontario. *Canadian Journal of Public Health, 101*(Suppl.1), S58-S62.

Colantonio, A., Parsons, D., Vander Laan, R., & Zagorski, B. (2009). *ABI dataset pilot project phase 1 report; Report to the Ontario Neurotrauma Foundation*. Toronto, ON: Ontario Neurotrauma Foundation.

Colantonio, A., Ratcliff, G., Chase, S., Kelsey, S., Escobar, M., & Vernich, L. (2004). Long term outcomes after moderate to severe traumatic brain injury. *Disability and Rehabilitation*, *26*(5), 253–261.

Conway, A.R.A., Tuholski, S.W., Shisler, R.J., & Engle, R.W. The effect of memory load on negative priming: An individual differences investigation. *Memory & Cognition 1999, 27*, 1042–1050.

Corrigan, J. D., Bogner, J. A., Mysiw, W. J., Clinchot, D., & Fugate, L. (1997). Systematic bias in outcome studies of persons with traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 78*(2), 132–137.

Corrigan, J. D., Lineberry, L. A., Komaroff, E., Langlois, J. A., Selassie, A. W., & Wood, K. D. (2007). Employment after traumatic brain injury: Differences between men and women. *Archives of Physical Medicine and Rehabilitation, 88*, 1400–1409.

Corrigan, J.D., Selassie, A.W., & Orman, J.A. (2010). The epidemiology of traumatic brain injury. *Journal of Head Trauma Rehabilitation, 25*(2), 72-80.

Comprehensive Meta-Analysis (Version 3) [Computer software]. (2016). Englewood, NJ: Biostat. Available from http://www.meta-analysis.com

Connelly, J., Chell, S., Tennant, A., Rigby, A.S., & Airey, C.M. (2006). Modelling 5-year functional outcome in a major traumatic injury survivor cohort. *Disability & Rehabilitation, 28,* 629-636.

Cooper, H., & Hedges, L. V. (1994). *The Handbook of Research Synthesis*. Russell Sage Foundation, New York.

Costas, R., Carvalho, L., & de Aragon, D. (2000). Virtual city for cognitive rehabilitation. In P. Sharkey, A. Cesarani, L. Pugnetti, & A. Rizzo (Eds.), Proceedings of the Third ICD-VRAT (pp. 305–313). Reading, UK: University of Reading.

Cremona-Meteyard, S. L., & Geffen, G. M. (1994). Event-related potential indices of visual attention following moderate to severe closed head injury. *Brain Injury, 8*(6), 541-558.

Crepeau, F., & Scherzer, P. (1993). Predictors and indicators of work status after traumatic brain injury: A meta-analysis. *Neuropsychological Rehabilitation, 3*(1), 5-35.

Cuberos-Urbano, G., Caracuel, A., Vilar-Lopez, R., Valls-Serrano, C., Bateman, A. & Verdejo-Garcia, A. (2013). Ecological validity of the multiple errands test using predictive models of dysexecutive problems in everyday life. *Journal of Clinical and Experimental Neuropsychology, 35*(3), 329-336.

Culbertson, W.C., & Zillmer, E.A. (2001). *Tower of London: Drexel University (TOL-DX): Test manual*. Toronto, Canada: Multi Health Systems.

Dahm, J., & Ponsford, J. (2015). Long-term employment outcomes following traumatic brain injury and orthopaedic trauma: A ten-year prospective study. *Journal of Rehabilitation Medicine, 47*, 932-940.

Damasio, A. R. (1995). Toward a neurobiology of emotion and feeling: Operational concepts and hypotheses. *The Neuroscientist, 1*(1), 19–25.

Dawson, D.R., Anderson, N., Burgess, P., Cooper, E., Krpan, & Stuss, D. T. (2009). Further development of the multiple errands test: Standardized scoring, reliability, and ecological validity for the Baycrest version. *Archives of Physical Medicine and Rehabilitation, 90,* S41-S51.

De Monte, V. E., Geffen, G. M., May, C. R., & McFarland, K. (2004). Double cross-validation and improved sensitivity of the rapid screen of mild traumatic brain injury. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society, 26*(5), 628-644.

Dean, P.J.A., & Sterr, A. (2013). Long-term effects of mild traumatic brain injury on cognitive performance. *Frontiers in Human Neuroscience, 7*, 30.

Deary I.J., Ebmeier K.P., MacLeod K.M., & Dougall, N. (1994). PASAT performance and the pattern of ^{99m}Tc-exametazinme in brain estimated with single photon emission tomography. *Biological Psychology, 38*, 1-18.

DeLand, N., Vanier, M., Lambert, J., & Provost, J. (1992). A study on focused attention in severely head-injured patients. *Proceedings of the Conference on Attention: Theoretical and Clinical Perspectives*. Baycrest Centre, Toronto, Canada.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. (2000). *California Verbal Learning Test – Second Edition (CVLT-II).* San Antonia, TX: The Psychological Corporation.

DeLuca, J. (2008). Information processing speed; How fast, how slow, and how come? In: DeLuca, J., & Kalmar, J. (Eds.), *Information processing speed in clinical populations*, (pp. 265-273). New York, NY: Taylor & Francis.

Demery, J.A., Larson, M.J., Dixit, N.K., Bauer, R.M., & Perlstein, W.M. (2010). Operating characteristics of executive functioning tests following traumatic brain injury. *Clinical Neuropsychologist, 24*, 1292-308.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*, 279–281.

Devitt, R., Colantonio, A., Dawson, D., Teare, G., Ratcliff, G., & Chase, S. (2006). Prediction of long-term occupational performance outcomes for adults after moderate to severe traumatic brain injury. *Disability and Rehabilitation, 28*, 547–559.

DiCenso, A., Bayley, L., & Haynes, B. (2009). Accessing preappraised evidence: Fine-tuning the 5S model into a 6S model. *Annals of Internal Medicine, 151*(6), JC3-2.

Dikmen, S., Machamer, J., & Temkin, N. (2001). Mild head injury: Facts and artifacts. *Journal of Clinical and Experimental Neuropsychology, 23*(6), 729–738.

Dikmen, S., Machamer, J., & Temkin, N. (2009). Neurobehavioral consequences of traumatic brain injury. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (3rd ed., pp. 597–617). New York, NY: Oxford University Press.

Dikmen, S., Machamer, J., & Temkin, N. (2017). Mild traumatic brain injury: Longitudinal study of cognition, functional status, and post-traumatic symptoms. *Journal of Neurotrauma*, *34*(8), 1524-1530.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology, 9*, 80–90.

Dikmen, S., McLean, A., & Temkin, N. (1986). Neuropsychological and psychosocial consequences of minor head injury. *Journal of Neurology, Neurosurgery & Psychiatry, 49*, 1227–1232.

DMW (2004). Available from: http://www.dmw.ca/index_frame.html.

Doctor, J.N., Castro, J., Temkin, N.R., Fraser, R.T., Machamer, J.E., & Dikmen, S. S. (2005). Workers' risk of unemployment after traumatic brain injury: A normed comparison. *Journal of the International Neuropsychological Society, 11*, 747-52.

Drake, A. I., Gray, N., Yoder, S., Pramuka, M., & Llewellyn, M. (2000). Factors predicting return to work following mild traumatic brain injury: A discriminant analysis. *Journal of Head Trauma Rehabilitation*, *15*(5), 1103-1112.

Dvir, O., Avni, N., Beer, B., Tamir, A., Naveh, Y., & Pelzan, B. (2003). Performance profile of the Behavioral Assessment of the Dysexecutive Syndrome in a healthy Israeli population. *Israeli Journal of Occupational Therapy, 12*, 207–223.

Elkind, J.S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of preliminary results. *CyberPsychology & Behavior, 4*(4), 489-496.

Ellis, Paul, D. (2010). *The essential guide to effect sizes*. Cambridge, UK: University Press.

Epic Games. (2002). Unreal Tournament 2003 3D engine [Software]. Available from: https://www.epicgames.com/unrealtournament/forums/past-unreal-tournament-games/unreal-tournament-2003-2004

Erlanger, D.M., Kutner, K.C., Barth, J.T., & Barnes, R. (1999). Neuropsychology of sports-related head injury: Dementia pugilistica to post-concussion syndrome. *The Clinical Neuropsychologist, 13*, 193–209.

Erez, A.B., Rothschild, E., Katz, N., Tuchner, M., & Hartman-Maeir, A. (2009). Executive functioning, awareness, and participation in daily life after mild traumatic brain injury: A preliminary study. *American Journal of Occupational Therapy, 63*(5), 634-640.

Evans, J. J., Chua, S. E., McKenna, P. J. & Wilson, B. A. (1997). Assessment of the Dysexecutive syndrome in schizophrenia. *Psychological Medicine, 27,* 635-646.

Evans, M.K., Lepkowski, J.M., Powe, N.R., LaVeist, T., Kuczmarski, M.F., & Zonderman, A. B. (2010). Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS): Overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethnicity & Disease, 20(3)*, 267–275.

Ewing-Cobbs, L., & Barnes, M. (2002). Linguistic outcomes following traumatic brain injury in children. *Seminars in Pediatric Neurology, 9*, 209–217.

Ezrachi, O., Ben-Yishay, Y., Kay, T., Diller, L., & Rattok, J. (1991). Predicting employment in traumatic brain injury following neuropsychological rehabilitation. Journal *of Head Trauma Rehabilitation, 6*(3), 71-84.

Farias, S. T., Harrell, E., Neumann, C., & Houtz, A. (2003). The relationship between neuropsychological performance and daily functioning with Alzheimer's disease: Ecological validity of neuropsychological tests. *Archives of Clinical Neuropsychology, 18*(6), 655-672.

Farley, K. L., Higginson, C. I., Sherman, M. F., & MacDougall, E. (2011). The ecological validity of clinical tests of visuospatial function in community-dwelling older adults. *Archives of Clinical Neuropsychology, 26*(8), 723-738.

Faul, M., Xu. L., Wald, M. M., & Coronado, V. G. (2010). *Traumatic Brain Injury in the United States: Emergency Department Visits, Hospitalizations and Deaths 2002–2006*. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Friedland, J.F., & Dawson, D.R. (2001). Function after motor vehicle accidents: A prospective study of mild head injury and posttraumatic stress. *Journal of Nervous and Mental Disorders, 189*, 426-434.

Fleminger, S., & Ponsford, J. (2005). Long term outcome after traumatic brain injury. *British Medical Journal*, *331*, 1419–1420.

Forslund, M.V., Roe, C., Perrin, P.B., Sigurdardottir, S., Lu, J., et al. (2017). The trajectories of overall disability in the first 5 years after moderate and severe traumatic brain injury. *Brain Injury, 31*(3), 329-335.

Fortin, S., Godbout, L., & Braun, C. M. J. (2003). Cognitive structure of executive deficits in frontally lesioned head trauma patients performing activities of daily living. *Cortex, 39*, 273–291.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychology. In: Sbordone, R. J., & Long, C. J. (Eds.), *Ecological Validity of Neuropsychological Testing*, (pp. 91-112). Delray Beach, FL: GR Press/St. Lucie Press.

Frencham, K. A. R., Fox, A. M., & Mayberry, T. M. (2005). Neuropsychological studies of mild traumatic brain injury: A meta-analytic review of research since 1995. *Journal of Clinical and Experimental Neuropsychology, 27*, 334–351.

Fröehlich, T. (1997). Das Virtuelle Ozeanarium, Zeitschrift Thema Forschung. *Darmstadt: Technische Universitat Darmstadt*, pp. 50–57.

Frost, R.B., Farrer, T.J., Primosch, M., & Hedges, D.W. (2013). Prevalence of traumatic brain injury in the general adult population: A meta-analysis. *Neuro-epidemiology, 40*, 154-159.

Fu, T.S., Jing, R., Fu, W.W., & Cusimano, M.D. (2016). Epidemiological trends of traumatic brain injury identified in the emergency department in a publicly-insured population, 2002-2010. *PLoS ONE, 11*(1), 1-13.

Garden, N., & Sullivan, K. A. (2010). An examination of the base rates of post-concussion symptoms: The influence of demographics and depression. *Applied Neuropsychology: Adult*, *17*(1), 1-7.

Gasquoine, P. G. (1997). Postconcussion symptoms. *Neuropsychology Review*, *7*(2), 77-85.

Gennarelli, T. A. (1986). Mechanisms and pathophysiology of cerebral concussion. *Journal of Head Trauma, 1*(2), 23-29.

Gennarelli, T. A., Thibault, L. E., Adams, J., Graham, D., Thompson, C. J., & Marcincin, R. P. (1982). Diffuse axonal injury and traumatic coma in the primate. *Annals of Neurology*, *12*(6), 564-574.

Gennarelli, T. A., Thibault, L. E., & Graham, D. I. (1998). Diffuse axonal injury: an important form of traumatic brain damage. *Neuroscientist*, *4*, 202-215.

Gallagher, A. G., McClure, N., McGuigan, J., Crothers, I., & Browning, J. (1999). Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy, 31*, 310–313.

Ghawami, H., Sadeghi, S., Raghibi, M., & Rahimi-Movaghar, V.R. (2017). Executive functioning of complicated-mild to moderate traumatic brain injury patients with frontal contusions. *Applied Neuropsychology: Adult*, *24*(4), 299-307.

Goldman-Sachs. (January 13, 2016). *Virtual and augmented reality understanding the race for the next computing platform*. Retrieved from http://www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf

Goldstein, G. (1996). Functional considerations in neuropsychology. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 75-89). Delray Beach, FL: GR Press/St. Lucie Press.

Goleburn, C.R., & Golden, C.J. (2001). Traumatic brain injury outcome in older adults: a critical review of the literature. *Journal of Clinical Geropsychology, 7,* 161–187.

Gollaher, M., High, W., Sherer, M., Bergloff, P., Boake, C., Young, M. E., et al. (1998). Prediction of employment outcome one to three years following traumatic brain injury. *Brain Injury, 12*, 255–263.

Google Cardboard, Google Inc. (2018). *Google Cardboard*. Retrieved from: https://vr.google.com/cardboard/

Gosselin, N., Bottari, C., Chen, J.K., Huntgeburth, S.C., De Beaumont, L., Petrides, M., et al. (2012). Evaluating the cognitive consequences of mild traumatic brain injury and concussion by using electrophysiology. *Neurosurgical Focus, 33*, E7: 1-7.

Gouveia, P. A. R., Brucki, S. M. D., Malheiros, S. M. F., & Bueno, O. F. A. (2007). Disorders in planning and strategy application in frontal lobe lesion patients. *Brain and Cognition, 63,* 240–246.

Grauwmeijer, E., Heijenbrok-Kal, M.H., Haitsma, I.K., & Ribbers, G.M. (2012). A prospective study on employment outcome 3 years after moderate to severe traumatic brain injury. *Archives of Physical Medicine Rehabilitation, 93,* 993–999.

Green, J. D. W., Hodges, J. R., & Baddeley, A. D. (1995). Autobiographical memory and executive functioning in early dementia of the Alzheimer's type. *Neuropsychologia, 33,* 1647–1670.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6,* 218–224.

Guilmette, T. J. (2008). Prediction of vocational functioning from neuropsychological data. In I. Z. Schultz, & R. J. Gatchel (Eds.), *Handbook of complex occupational disability claims* (pp. 303–314). US: Springer.

Guilmette, T. J., & Kastner, M. P. (1996). The prediction of vocational functioning from neuropsychological data. In R. Sbordone & C. Long (Eds.), *Ecological Validity of Neuropsychological Testing* (pp. 387-411). Boca Raton, FL: St. Lucie Press.

Haboubi, N.H.J, Long, J., Koshy, M., & Ward, A.B. (2001). Short-term sequelae of minor head injury (6 years experience of minor head injury clinic). *Disability and Rehabilitation, 23*(14), 635-638.

Hanks, R. A., Rapport, L. J., Millis, S. R., & Deshpande, S. A. (1999). Measures of executive functioning as predictors of functional ability and social integration in a rehabilitation sample. *Archives of Physical Medicine and Rehabilitation, 80*, 1030–1037.

Hanlon, R.E., Demery, J.A., Martinovich, Z., & Kelly, J.P. (1999). Effects of acute injury on neurophysical status and vocational outcome following mild traumatic brain injury. *Brain Injury*, *13*(11), 873-887.

Hammond, K.R. (1998). Ecological validity: Then and now. Unpublished manuscript available electronically at: http:00brunswik.org0notes0essay2.html

Harman-Smith, Y.E., Mathias, J.L., Bowden, S.C., Rosenfeld, J.V., & Bigler, E.D. (2013). Wechsler Adult Intelligence Scale-Third Edition profiles and their relationship to self-reported outcome following traumatic brain injury. *Journal of Clinical & Experimental Neuropsychology: Official Journal of the International Neuropsychological Society, 35*, 785-98.

Hart, T., Millis, S., Novack, T., Englander, J., Fidler-Sheppard, R., & Bell, K. R. (2003). The relationship between neuropsychologic function and level of caregiver supervision at 1 year after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *84*(2), 221-230.

Hartikainen, K. M., Wäljas, M., Isoviita, T., Dastidar, P., Liimatainen, S., Solbakk, A., . . . Ohman, J. (2010). Persistent symptoms in mild to moderate traumatic brain injury associated with executive dysfunction. *Journal of Clinical and Experimental Neuropsychology*, *32*(7), 767-774.

Hayden, M.E. (1997). Mild traumatic brain injury: A primer for understanding its impact on employee return to work. *AAOHN Journal, 45*(12), 635-645.

Heaton, R.K., Chelune, G.J., Talley, J., Gary, G., & Glenn Curtiss. (1993*). Wisconsin Card Sorting Test Manual: Revised and Expanded*. U.S.A., Psychological Assessment Resources.

Heaton, R. K. & Pendleton, M. G. (1981) Use of neuropsychological tests to predict adult patient's everyday functioning. *Journal of Consulting and Clinical Psychology, 49*, 807–821.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology*, *18*(4), 621-628.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. (1998). *Fixed- and random-effects models in meta-analysis*. Psychological Methods, 3, 486–504.

Hibbard, M. R., Uysal, S., Kepler, K., Bogdany, J., & Silver, J. (1998). Axis I psychopathology in individuals with traumatic brain injury. J*ournal of Head Trauma and Rehabilitation, 13,* 24-39.

Higgins, J.P.T., & Thompson, S.G., (2002). Quantifying heterogeneity in a meta-analysis. *Statistics and*

Medicine, *21*, 1539-1558.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. BMJ, 327, 557–560.

Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology, 15*, 185–204.

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in U.S. warriors returning from Iraq. *New England Journal of Medicine*, *358*(5), 453–463.

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

Holt, D.V., Rodewald, K., Rentrop, M., Funke, J., Weisbrod, M., & Kaiser, S. (2011). The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of International Neuropsychological Society, 17*, 327 – 335.

Hue, P., Delannay, B., & Berland, J.-C. (1997). Virtual reality training simulator for long time flight. In: Seidel, R.J., & Chantelier, P.R. (eds.), *Virtual reality, training's future?* New York: Plenum Press, pp. 69–76.

Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods*, *11*(2), 193-206.

Human Resources and Development Canada. (2006). *National Occupation Classification 2006*. Ottawa: The Government of Canada. Retrieved from: http://noc.esdc.gc.ca/English/noc/welcome.aspx?ver=06

Humphrey, M., & Oddy, M. (1980). Return to workafterheadinjury: a review of post-war studies. *Injury*, *12*(2), 107-114.

Humphreys, I., Wood, R.L., Phillips, C.J., & Macey, S. (2013). The costs of traumatic brain injury: A literature review. *Clinicoeconomics & Outcomes Research, 5*, 281–287.

Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, California: Sage.

IBM Corp. Released 2013. *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Ip, R. Y., Dornan, J., & Schentag, C. (1995). Traumatic brain injury: Factors predicting return to work or school. *Brain Injury*, *9*(5), 517–532.

Iverson, G. L. (2006). Sensitivity of computerized neuropsychological screening in depressed university students. *The Clinical Neuropsychologist, 20*, 695-701.

Iverson, G.L. (2010). Mild traumatic brain injury meta-analyses can obscure individual differences. *Brain Injury, 24*(10), 1246-1255.

Iverson, G. L., & McCracken, L. M. (1997). 'Postconcussive' symptoms in persons with chronic pain. *Brain Injury, 11*(11), 783-790.

Jansari, A.S., Devlin, A., Agnew, R., Akesson, K., Murphy, L., & Leadbetter, T. (2014). Ecological assessment of executive functions: A new virtual reality paradigm. *Brain Impairment, 15*(2), 71-87.

Jeon, W., Clamann, W., Kaber, D.B., & Currie, N.J. (2013). Assessing goal-directed three-dimensional movements in a virtual reality block design task. *IEEE International Conference on Systems, Man, and Cybernetics, Manchester*. 3739-3744.

Johansson, B., Berglund, P., & Ronnback, L. (2009). Mental fatigue and impaired information processing after mild and moderate traumatic brain injury. *Brain Injury, 23* (13-14), 1027-1040.

Johansson, U., & Bernsprang, B. (2001). Predicting return to work after brain injury using occupational therapy assessments. *Disability and Rehabilitation, 23*(11), 474-480.

Josman, N., Klinger, E., & Kizony, R. (2008). Performance within the virtual action planning supermarket (VAP-S): An executive function profile of three different populations suffering from deficits in the central nervous system. In *Proceedings of the 7th ICDVRAT.*

Jourdan, C., Bosserelle, V., Azerad, S., Ghout, I., Bayen, E., Aegerter, P., et al. (2013). Predictive factors for 1-year outcome of a cohort of patients with severe traumatic brain injury (TBI): Results from the PariS-TBI study. *Brain Injury, 27*, 1000–1007.

Jovanovski, D., Zakzanis, K. K., Campbell, Z., Erb, S., & Nussbaum, D. (2012). Development of a novel, ecologically oriented virtual reality measure of executive function: The Multitasking in the City Test. *Applied Neuropsychology, 19*(3), 171-182.

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

Kay, T., Harrington, D.E., Adams, R., Anderson, T., Berrol, S., Cicerone, K., et al. (1993). Definition of mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 8*(3), 85-87.

Kayani, N.A., Homan, S., Yun, S., & Zhu, B.P. (2009). Health and economic burden of traumatic brain injury: Missouri, 2001–2005. *Public Health Reports*, *124*(4), 551–560.

Kang, Y.J., Ku, J., Han, K., Kim, S.I., Yu, T.W., Lee, J. H., et al. (2008). Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. *CyberPsychology, 11*(3), 329-339.

Kennedy, J.E., Leal, F.O., Lewis, J.D., Cullen, M.A., & Amador, R.R. (2010). Posttraumatic stress symptoms in OIF/OEF service members with blast-related and non-blast-related mild TBI. *Neurorehabilitation, 26*(3), 223–231.

Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherman, F. S. (1987). The Functional Independence Measure: A new tool for rehabilitation. In: M. G. Eisenberg R. C. Grzesiak (Eds.), *Advances in clinical rehabilitation, Vol 2*. (pp. 6-18) New York, NY: Springer.

Ketchum, J.M., Almaz Getachew, M., Krch, D., Banos, J.H., Kolakowsky-Hayner, S.A., Lequerica, A., et al. (2012). Early predictors of employment outcomes 1 year post traumatic brain injury in a population of Hispanic individuals. *NeuroRehabilitation, 30*, 13–22.

Kewman, D.G., Yanus, B., & Kirsch, N. (1988). Assessment of distractibility in auditory comprehension after traumatic brain injury. *Brain Injury, 2*(2), 131-137.

Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the california verbal learning test and the wisconsin card sorting test. *Archives of Clinical Neuropsychology*, *13*(6), 523-534.

Kinsella, G. J. (2008). Traumatic brain injury and processing speed. In: DeLuca, J., & Kalmar, J. (Eds.), *Information processing speed in clinical populations*, (pp. 173-194). New York, NY: Taylor & Francis.

Kisser, J., Waldstein, S.R., Evans, M.K, & Zonderman, A.B. (2017). Lifetime prevalence of traumatic brain injury in a demographically diverse community sample. *Brain Injury, 31*(5), 620-623.

Kleiven, S., Peloso, P.M., & von Holst, H. (2003). The epidemiology of head injuries in Sweden from 1987 to 2000. *Injury Control and Safety Promotion, 10*, 173–180.

Klinger, E., Chemin, I., Lebreton, S., & Marie, R.M. (2006). Virtual action planning in Parkinson's disease: A control study. *CyberPsychology & Behavior, 9*(3), 342–347.

Klonoff, P.S., Talley, M.C., Dawson, L.K., Myles, S.M., Watt, L.M., Gehrels, J. A., et al. (2007). The relationship of cognitive retraining to neurological patients' work and school status. *Brain Injury, 21*(11), 1097-1107.

Knerr, B.W., Breaux, R., Goldberg, S.L., & Thurman, R.A. (2002). National defense. In K.S. Hale, & K.M. Stanney (Eds.) *Handbook of virtual environments: Design, implementation, and applications* (pp. 857-872). Mahwah, NJ: CRC Press.

Koenig, S. T., Crucian, G. P., Dalrymple-Alford, J. C., & Dünser, A. (2010, August). Assessing navigation in real and virtual environments: A validation study. Paper presented at the P. Sharkey, & J. Sanchez (Eds.), *Proceedings of the International Conference Series on Disability, Virtual Reality and Associated Technologies*, Vina del Mar, Chile.

Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Wisconsin Card Sorting Test–64 card version professional manual.* Odessa, FL: Psychological Assessment Resources.

Konrad, C., Geburek, A.J., Rist, F., Blumenroth, H., Fischer, B., Husstedt, I., et al. (2011). Long-term cognitive and emotional consequences of mild traumatic brain injury. *Psychological Medicine, 41*, 1197-211.

Krause, J.F., McArthur, D.L., Silverman, T.A., & Jayaraman, M. (1996). Epidemiology of brain injury. In R.K. Narayan et al. (Eds), *Neurotrauma* (pp. 13-30). New York, NY: McGraw-Hill.

Kraus, M.F., Susmaras, T., Caughlin, B.P., Walker, C.J., Sweeney, J.A., & Little, D.M. (2007). White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study. *Brain, 130*, 2508-19.

Kvavilashvili, L. & Ellis, J. (2004). Ecological validity and the real-life laboratory controversy in memory research: A critical and historical review. *History & Philosophy of Psychology*, *6*, 59–80.

Ku, J., Mraz, R., Baker, N., Zakzanis, K. K., Lee, J. H., Kim, I. Y., et al. (2003). A data glove with tactile feedback for fMRI of virtual reality experiments. *Cyberpsychology and Behaviour, 6*(5), 497-508.

Lam, C. S., Priddy, D. A., & Johnson, P. (1991). Neuropsychological indicators of employability following traumatic brain injury. *Rehabilitation Counseling Bulletin, 35*, 68–75.

Lamberts, K.F., Evans, J.J., & Spikman, J.M. (2010). A real-life, ecologically valid test of executive functioning: The executive secretarial task. *Journal of Clinical and Experimental Neuropsychology, 32*(1), 56-65.

Lange, B., Chang, C. Y., Suma, E., Newman, B., Rizzo, A. S., & Bolas., M. (2011). Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect Sensor. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine Biological Society, 2011*, 1831-1834.

Lange, R.T., Iverson, G.L., Zakrzewski, M.J., Ethel-King, P.E., & Franzen, M.D. (2005). Interpreting the Trail Making Test Following Traumatic Brain Injury: Comparison of Traditional Time Scores and Derived Indices. *Journal of Clinical and Experimental Neuropsychology, 27*(7), 897-906.

Langlois, J.A., Rutland-Brown, W., & Thomas, K.E. (2004). *Traumatic brain injury in the United States: Emergency department visits, hospitalizations, and deaths.* Atlanta, GA: US Department Health and

Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Langlois, J.A., Rutland-Brown, W., & Wald, M.M. (2006). The epidemiology and impact of traumatic brain injury. *Journal of Head Trauma Rehabilitation, 21*(5), 375-378.

Larson, E.B., Feignon, M., Gagliardo, P., & Dvorkin, A.Y. (2014). Virtual reality and cognitive rehabilitation: A review of the current outcome research. *NeuroRehabilitation, 34*, 759-772.

LeBlanc, J., Hayden, M., & Paulman, R. (2000) A comparison of neuropsychological and situational assessment for predicting employability after closed head injury. *The Journal of Head Trauma Rehabilitation, 15*, 1022 – 1040.

Lees-Haley, P. R., & Brown, R. S. (1993). Neuropsychological complaint base rates of 170 personal injury claimants. *Archives of Clinical Neuropsychology, 8*(3), 203–209.

Lengenfelder, J., Schultheis, T. M., Al-Shihabi, T., Mourant, R., DeLuca, J. (2002). Divided attention and driving: a pilot study using virtual reality technology. *Journal Head Trauma Rehabilitation, 17*(1), 26-37.

Leon-Carrion, J., Dominguez-Morales, M.R., Martin, J.M.B., & Murillo-Cabezas, F. (2006). Epidemiology of traumatic brain injury and subarachnoid hemorrhage. *Pituitary, 8*, 197-202.

Levin, H. S., Hanten, G., Zhang, L., Swank, P. R., Ewing-Cobbs, L., Dennis, M., . . . Hunter, J. V. (2004). Changes in working memory after traumatic brain injury in children. *Neuropsychology*, *18*(2), 240-247.

Levin, H. S., Mattis, S., Ruff, R. M., Eisenberg, H. M., Marshall, L. F., Tabaddor, K., . . . Frankowski, R. F. (1987). Neurobehavioural outcome following minor head injury: a three-center study. *Journal of Neurosurgery*, (66), 234-243.

Levine, B., Dawson, D., Schwartz, M.L., Boutet, I., & Stuss, D.T. (2000). Assessment of strategic self-regulation in traumatic brain injury: its relationship to injury severity and psychosocial outcome. *Neuropsychology, 14*, 481 – 500.

Levit, M.A., Sutton, M., Goldman, J., Mikhail, M., Christopher, T. (1994). Cognitive dysfunction from minor head trauma. *American Journal of Emergency Medicine*, *12*, 172 -175.

Lezak, M. D. (1976). *Neuropsychological assessment*. Oxford, United Kingdom: Oxford University Press.

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

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis (Vol. 49)*. Thousand Oaks, CA: Sage.

Long, C. J. (1996). Neuropsychological tests: A look at our past and the impact that ecological issues may have on our future. In: Sbordone, R. J., and Long, C. J. (eds.), *Ecological Validity of Neuropsychological Testing*, GR Press/St. Lucie Press, Delray Beach, FL, pp. 1–14.

Losoi, H., Silverberg, N. D., Wäljas, M., Turunen, S., RostiOtajarvi, E., Helminen, M., et al. (2016). Recovery from mild traumatic brain injury in previously healthy adults. *Journal of Neurotrauma, 33*(8), 766-776.

Maas, A.I., Menon, D.K., Adelson, P.D., Andelic, N., Bell, M.J., Belli, A., et al. (2017). Traumatic brain injury: Integrated approaches to improve prevention, clinical care, and research. *Lancet Neurology, 16*(12), 987-1048.

Machamer, J., Temkin, N., Fraser, R., Doctor, J.N., & Dikmen, S. (2005). Stability of employment after traumatic brain injury. *Journal of the International Neuropsychological Society, 11,* 807-816.

MacFlynn, G., Montgomery, E.A., Fenton, G.W., & Rutherford, W. (1984). Measurement of reaction time following minor head injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 47,* 1326-1331.

MacKenzie, E. J., Bosse, M. J., Kellam, J. F., Pollak, A. N., Webb, L. X., Swiontkowski, M. F., et al. (2006). Early predictors of long-term work disability after major limb trauma. *Journal of Trauma, 61*, 688–694.

MacLennan, D.L., & MacLennan, D.C. (2008). Assessing readiness for post-secondary education after traumatic brain injury using a simulated college experience. *NeuroRehabilitation, 23*, 521-528.

Maeir, A., Krauss, S., & Katz, N. (2010). Ecological validity of the Multiple Errands Test (MET) on discharge from neurorehabilitation hospital. *Occupation, Participation and Health, 31,* S38–S46.

Madigan, N. K., DeLuca, J., Diamond, B. J., Tramontano, G., & Averill, A. (2000). Speed of information processing in traumatic brain injury: Modality-specific factors. *Journal of Head Trauma Rehabilitation, 15*(3), 943-956.

Mahoney, D.P. (1997). Defensive driving. *Computer Graphics World*, 20, 71.

Makatura, T. J., Lam, C. S., Leahy, B. J., Castillo, M. T., & Kalpakjian, C. Z. (1999). Standardized memory tests and the appraisal of everyday memory. *Brain Injury, 13*, 355–367.

Malec, J. F., Brown, A. W., Leibson, C. L., Flaada, J. T., Mandrekar, J. N., Diehl, N. N. et al., (2007). The mayo classification system for traumatic brain injury severity. *Journal of Neurotrauma, 24*, 1417-1424.

Malec, J.F., Smigielski, J.S., DePompolo, R.W., & Thompson, J.M. (1993). Outcome evaluation and prediction in a comprehensive-integrated post-acute outpatient brain injury rehabilitation program. *Brain Injury, 7*, 15–29.

Malojcic, B., Mubrin, Z., Coric, B., Susnic, & Spilich, G.J. (2008). Consequences of mild traumatic brain injury on information processing assessed with attention and short-term memory tasks. *Journal of Neurotrauma, 25*, 30-37.

Man, D. W., Poon, W. S., & Lam, C. (2013). The effectiveness of artificial intelligent 3-D virtual reality vocational problem-solving training in enhancing employment opportunities for people with traumatic brain injury. *Brain Injury, 27*(9), 1016-1025.

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury, 18*(11), 1067-1081.

Mangels, J.A., Craik, F.I., Levine, B., Schwartz, M.L., & Stuss, D.T. (2002). Effects of divided attention on episodic memory in chronic traumatic brain injury: A function of severity and strategy. *Neuropsychologia, 40*, 2369–2385.

Manera, V., Petit, P.D, Derreumaux, A., Orvieto, I., Romagnoli, M., Lyttle, G., et al. (2015). ‘Kitchen and cooking’, a serious game for mild cognitive impairment and Alzheimer’s disease: A pilot study. *Frontiers in Aging Neuroscience, 7*, 24.

Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H., (2001). The differential assessment of children's attention: The Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *Journal of Child Psychology, and Psychiatry, and Allied Disciplines, 42*(8), 1065-1081.

Maroon, J.C., Lovell, M.R., Norwig, J., Podell, K., & Hartl, R. (2000). Cerebral concussion in athletes: Evaluation and neuropsychological testing. *Neurosurgery, 47*, 659–669.

Maruta, J., Palacios, E. M., Zimmerman, R. D., Ghajar, J., & Mukherjee, P. (2016a). Chronic post-concussion neurocognitive deficits. I. relationship with white matter integrity. *Frontiers in Human Neuroscience, 10*, Art 35.

Maruta, J., Spielman, L. A., Yarusi, B. B., Wang, Y., Silver, J. M., & Ghajar, J. (2016b). Chronic post-concussion neurocognitive deficits. II. relationship with persistent symptoms. *Frontiers in Human Neuroscience, 10*, Art 45.

Masson, F., Thicoipe, M., Makni, T., Aye, P., Erny, P., et al (2003). Epidemiology of traumatic comas: A prospective population-based study. *Brain Injury, 17*, 279–293.

Mathias, J. L., Beall, J. A., & Bigler, E. D. (2004). Neuropsychological and information processing deficits following mild traumatic brain injury. *Journal of the International Neuropsychological Society, 10*, 286-297.

Mathias J.L., Bowden S.C, Bigler E.D, & Rosenfeld, J.V. (2007). Is performance on the Wechsler test of adult reading affected by traumatic brain injury?. *British Journal of Clinical Psychology, 46*, 457-66.

Mathias, J. L., & Coats, J. L. (1999). Emotional and cognitive sequelae to mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *21*, 200-215.

Mathias, J. L., & Wheaton. P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: a meta-analytic review. *Neuropsychology, 21*(2), 212-23.

Marcotte, T. D., & Grant, I. (2010). *Neuropsychology of Everyday Functioning*. New York, NY: Guilford Press.

Mayer, A.R., Yang, Z., Yeo, R.A., Pena, A., Ling, J.M., Mannell, M. V., et al. (2012). A functional MRI study on multimodal selective attention following mild traumatic brain injury. *Brain Imaging and Behavior*, *6*, 343-354.

McCallister, T. W., Flashman, L. A., McDonald, B. C., & Saykin, A. J. (2006). Mechanisms of working memory dysfunction after mild and moderate TBI: Evidence from functional MRI and neurogenetics. *Journal of Neurotrauma*, *23*(10), 1450-1467.

McAllister, T. W., Saykin, A. J., Flashman, L. A., Sparling, M. B., Johnson, S. C., Guerin, S. J., et al. (1999). Brain activation during working memory 1 month after mild traumatic brain injury. *Neurology, 53*, 1300–1308.

McAllister, T. W., Sparling, M. B., Flashman, L. A., & Saykin, A. J. (2001). Neuroimaging findings in mild traumatic brain injury. Journal of Clinical and Experimental Neuropsychology, 23, 775–791.

McCrea, M., Iverson, G.L., McAllister, T.W., Hammeke, T.A., Powell, M.R., Barr, W. B. et al. (2009). An integrated review of recovery after mild traumatic brain injury (MTBI): Implications for clinical management. *The Clinical Neuropsychologist*, *23*, 1368-1390.

McCrea, M., Kelly, J.P., Randolph, C., Cisler, R., & Berger, L. (2002). Immediate neurocognitive effects of concussion. *Neurosurgery, 50,* 1032–1042.

McCrory, P., Meeuwisse, W., Dvořák, J., Aubry, M., Bailes, J., Broglio, S., et al. (2018). Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine, 51*(11), 838–847.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., Milne, A. B., et al. (2001). Using virtual environments in the assessments of executive dysfunction. *Presence, 10*(4), 375-383.

McInnis, K., Friesen, C.L., MacKenzie, D.E., Westwood, D.A., & Boe, S.G. (2017). Mild traumatic brain injury (mTBI) and chronic cognitive impairment: A scoping review. *PLoS ONE, 12*(4), 1-19.

McLay, R. N., Wood, D. P., Webb-Murphy, J. A., Spira, J. L., Wiederhold, M. D., Pyne, J. M., et al. (2011). A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. *Cyberpsychology, Behaviour and Social Networking*, *14*(4), 223-229.

McPheeters, H.L. (1984). Statewide mental health outcome evaluation: A perspective of two southern states. *Community Mental Health Journal*, *20*, 44–55.

Melamed, S., Stem, M., Rahmani, L., Groswasser, Z., & Najenson, T. (1982). Work congruence, behavioral pathology, and rehabilitation status of severe cranio-cerebral injury patients. In E. Lahav (ed.), *Psychosocial Research in Rehabilitation*. Israel: The State of Israel, Ministry of Defence, Publishing House.

Menon, D., Schwab, K., Wright, D.W., & Maas, A.I. (2010). Position statement: Definition of traumatic brain injury. *Archives Physical Medicine and Rehabilitation, 91*, 1637-1640.

Messinis, L., Kosmidis, M.H., Tsakona, I., Georgiou, V., Aretouli, E., & Papthanasopoulos, P. (2007). Ruff 2 and 7 Selective Attention Test: Normative data, discriminant validity and test–retest reliability in Greek adults. *Archives of Clinical Neuropsychology, 22,* 773-785.

Meyers, J. E., & Meyers, K. E. (1995). *Rey complex figure test and recognition trial: Professional manual*. Odessa, FL: Psychological Assessment Resources Inc.

Meyers J.E, & Rohling M.L. (2004). Validation of the Meyers Short Battery on mild TBI patients. *Archives of Clinical Neuropsychology, 19*, 637-51.

Michael, A.P, Stout, J., Roskos, P.T., Bolzenius, J., Gfeller, J., Mogul, D., & Bucholz, R. (2015). Evaluation of Cortical Thickness after Traumatic Brain Injury in Military Veterans. *Journal of Neurotrauma, 32*, 1751-8.

Miekisiak, G., Czyz, M., Tykocki, T., Kaczmarczyk, J., Zaluski, R., & Latka, D. (2016) Traumatic brain injury in Poland from 2009–2012: A national study on incidence. *Brain Injury, 30*(1), 79-82.

Miller, J.B., & Barr, W.B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology, 32*, 541-554.

Mirsky, A.F., Anthony, B.J., Duncan, C.C., Ahearn, M.B., & Kellam, S.G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review, 2*(2), 109-145.

Mittenberg, W., & Strauman, S. (2000). Diagnosis of mild head injury and the postconcussion syndrome. *Journal of Head Trauma Rehabilitation*, *15*(2), 783-791.

Mooney, G., Speed, J., & Sheppard, S. (2005). Factors related to recovery after mild traumatic brain injury. *Brain Injury, 19*(12), 975-987.

Morey, L. C. (1991). *The Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Inventory.

Morris, Z.S., Woodling, S., & Grant. J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of Royal Society of Medicine*, *104*(12), 510-520.

Morrison, M. T., Giles, G. M., Ryan, J. D., Baum, C. M., Dromerick, A. W., Polatajko, H. J., et al. (2013). Multiple errands test-revised (MET-R): A performance-based measure of executive function in people with mild cerebrovascular accident. *American Journal of Occupational Therapy, 67*, 460–468.

Mraz, R., Hong, J., Quintin, G., Staines, W. R., McIlroy, W. E., Zakzanis, K. K., et al. (2003). A platform for combining virtual reality experiments with functional magnetic resonance imaging. *Cyberpsychology and Behaviour, 6*(4), 359-368.

Nakase-Richardson, R., Yablon, S. A., & Sherer, M. (2007). Prospective comparison of acute confusion severity with duration of post-traumatic amnesia in predicting employment outcome after traumatic brain injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 78,* 872–876.

Nelson, H.E. (1982). *National Adult Reading Test (NART)*. Windsor, Bershire, England: The NFER-NELSON Publishing Company.

Nelson, L. A., Yoash-Gantz, R. E., Pickett, T. C., & Campbell, T. A. (2009). Relationship between processing speed and executive functioning performance among OEF/OIF veterans: Implications for postdeployment rehabilitation. *The Journal of Head Trauma Rehabilitation, 24*, 32–40.

NIH. (1998). *Rehabilitation of persons with traumatic brain injury: NIH consensus development conference on rehabilitation of persons with traumatic brain injury. October 26-28, 1998*. Natcher Conference Center Auditorium. Bethesda, Maryland: National Institutes of Health.

NIH. (1999). Rehabilitation of persons with traumatic brain injury. *Journal of the American Medical Association, 282*(10), 974-983.

Nir-Hadad, S. Y., Weiss, P. L., Waizman, A., Schwartz, N., & Kizony, R. (2015). A virtual shopping task for the assessment of executive functions: Validity for people with stroke. Neuropsychological Rehabilitation, 27, 808 – 833.

Nolin, P. (2006). Executive memory dysfunctions following mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *21*(1), 68-75.

Nolin, P., & Heroux, L. (2006). Relations among sociodemographic, neurologic, clinical, and neuropsychologic variables, and vocational status following mild traumatic brain injury: A follow-up study. *Journal of Head Trauma Rehabilitation, 6*, 524-526.

Norcross, J. C., Hedges, M., & Prochaska, J. O. (2002). The face of 2010: A Delphi poll on the future of psychotherapy. *Professional Psychology: Research and Practice*, *33,* 316–322.

Norris, G., & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Concurrent and construct validity. *Neuropsychological Rehabilitation*, *10*, 33-45.

Novakovic-Agopian, T., Chen, A.J., Rom, S., Rossi, A., Abrams, G., D'Esposito, M., et al. (2014). Assessment of subcomponents of executive functioning in ecologically valid settings: The goal processing scale. *Journal of Head Trauma Rehabilitation*, *29*(2), 136-146.

O'Connell, M.J. (2000). Prediction of return to work following traumatic brain injury: intellectual, memory, and demographic variables. *Rehabilitation Psychology, 45*(2), 212–217.

Odgaard, L., Johnsen, S.P., Pedersen, A.R., & Nielsen, J.F. (2016). Return to work after severe traumatic brain injury: A nationwide follow-up study. *Journal of Head Trauma Rehabilitation, 32*(3), E57-E64.

Okonkwo, O. C., Wadley, V. G., Griffith, H. R., Ball, K., & Marson, D. C. (2006). Cognitive correlates of financial abilities in mild cognitive impairment. *Journal of the American Geriatric Society, 54*(11), 1745-1750.

Ommaya, A.K., Fisch, F.J., Mahone, R.M., Corrao, P., & Letcher, F. (1993). Comparative Tolerances for Cerebral Concussion by Head Impact and Whiplash Injury in Primates. In S. E. Backaitis (Ed.), *Biomechanics of Impact Injury and Injury Tolerances of the Head-Neck Complex* (pp. 265-274). Warrendale: Society of Automotive Engineers.

O'Neill, J., Hibbard, M.R., Brown, M., Jaffe, M., Sliwinski, M., Vandergoot, D., et al. (1998). The effect of employment on quality of life and community integration after traumatic brain injury. *Journal of Head Trauma Rehabilitation, 13*, 68–79.

Ord, J. S., Greve, K. W., Bianchini, K. J., & Aguerrevere, L. E. (2010). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin card sorting test. *Journal of Clinical and Experimental Neuropsychology*, *32*(2), 132-140.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157-159.

Osterrieth, P.A. (1944). "Filetest de copie d'une figure complex: Contribution a l'etude de la perception et de la memoire". *Archives de Psychologie, 30,* 286–356.

Overgaard, M., Hoyer, C.B., & Christensen, E.F. (2011). Long-term survival and health-related quality of life 6 to 9 years after trauma. *Journal of Trauma, 71*, 435–441.

Ownsworth, T., & McKenna, K. (2004). Investigation of factors related to employment outcome following traumatic brain injury: A critical review and conceptual model. *Disability and Rehabilitation, 26*(13), 765-784.

Ozen, L.J., & Fernandes, M.A. (2012). Slowing down after a mild traumatic brain injury: A strategy to improve cognitive task performance? *Archives of Clinical Neuropsychology, 27*, 85-100.

Paniak, C., Toller-Lobe, G., Melnyk, A., & Nagy, J. (2000). Prediction of vocational status three to four months after mild traumatic brain injury. *Journal of Musculoskeletal Pain, 8*(1/2), 193–200.

Paré, N., Rabin, L. A., Fogel, J., & Pépin, M. (2009). Mild traumatic brain injury and its sequelae: Characterisation of divided attention deficits. *Neuropsychological Rehabilitation*, *19*(1), 110–137.

Parsons, T. D., Courtney, C., Rizzo, A. A., Armstrong, C., Edwards, J., & Reger, G. (2012). Virtual reality paced serial assessment for neuropsychological assessment of a military cohort. *Studies in Health Technology and Informatics, 173*, 331-337.

Parsons, T.D., McMahan, T., & Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *The Clinical Neuropsychologist, 32*(1), 16-41.

Parsons, T.D., & Kane, R. (2017). Computational neuropsychology: Current and future prospects for interfacing neuropsychology and technology. In R.L. Kane & T.D. Parsons (Eds.), *The role of technology in clinical neuropsychology* (pp. 471-483). New York, NY: Oxford University Press.

Parsons, T. D., & Rizzo, A. A. (2008). Initial validation of a virtual environment for assessment of memory functioning: Virtual reality cognitive performance assessment test. *Cyberpsychology & Behavior, 11*, 17–25.

Perfetti, B., Varanese, S., Mercuri, P., Mancino, E., Saggino, A., & Onofrj, M. (2010). Behavioural assessment of dysexecutive syndrome in Parkinson’s disease without dementia: A comparison with other clinical executive tasks. *Parkinsonism and Related Disorders*, *16*, 46-50.

Pertab, J.L., James, K.M., & Bigler, E.D. (2009). Limitations of mild traumatic brain injury meta-analysis. *Brain Injury, 23*(6), 498-508.

Peterson, S.E., & Posner, M.I. (2012). The attention system of the human brain: 20 years after. *Annual Review in Neuroscience, 35*, 73-89.

Plancher, G., Nicolas, S., & Piolino, P. (2008). Contribution of virtual reality to neuropsychology of memory: Study in aging. *Psychologie & Neuropsychiatrie du Vieillissement, 6*(1), 7-22.

Pollak, Y., & Weiss, P. L., Rizzo, A. A., Weizer, M., Shriki, L., Shalev, R. S., et al. (2009). The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *Journal of Developmental and Behavior Pediatrics, 30*(1), 2-6.

Ponsford, J., Cameron, P., Fitzgerald, M., Grant, M., & Mikocka-Walus, A. (2011). Long-term outcomes after uncomplicated mild traumatic brain injury: a comparison with trauma controls. *Journal of Neurotrauma, 28*, 937-46.

Ponsford, J., Downing, M.G., Olver, J., Ponsford, M., Acher, R., Carty, M., et al. (2014). Longitudinal follow-up of patients with traumatic brain injury: Outcome at two, five, and ten years post-injury. *Journal of Nuerotrauma, 31*, 64-77.

Ponsford, J., Draper, K. & Schoenberger, M. (2008). Functional outcome 10 years after traumatic brain injury: Its relationship with demographic, injury severity, and cognitive and emotional status. *Journal of the International Neuropsychological Society, 14*(2), 233-242.

Ponsford, J., & Kinsella, G. (1992). Attentional deficits following closed head injury. *Journal of Clinical and Experimental Neuropsychology, 14,* 822-838.

Ponsford, J., & Spitz, G. (2015). Stability of employment over the first 3 years following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 30*(3), E1-E11.

Ponsford, J., Willmont, C., Rothwell, A., Cameron, P., Kelly, A., Nelms, R., et al. (2000). Factors influencing outcome following mild traumatic brain injury in adults. *Journal of the International Neuropsychological Society, 6*, 568–579.

Pontifex, M. B., Broglio, E.S., Drollette, M. R., Scudder, C. R., Johnson, P. M., O'Connor, P. M., et al. (2012). The relation of mild traumatic brain injury to chronic lapses of attention. *Research Quarterly for Exercise and Sport, 83*(4), 553-559.

Posner, M.I., & Peterson, S.E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25-42.

Posner, M.I., & Rothbart, M.K. (1992). Attentional mechanisms and conscious experience. In A.D. Milner & M.D. Rugg (Eds.), *The neuropsychology of consciousness* (pp. 91-111). New York: Academic Press.

Potter, D. D., Jory, S. H., Bassett, M. R., Barrett, K., & Mychalkiw, W. (2002). Effect of mild head injury on event-related potential correlates of Stroop task performance. *Journal of the International Neuropsychological Society, 8*, 828–837.

Prigatano, G.P. (1999). *Principles of Neuropsychological Rehabilitation*. New York, NY: Oxford University Press.

Pugnetti, L., Mendozzi, L., Motta, A., Cattaneo, A., Barbieri, E., & Brancotti, A. (1995). Evaluation and retraining of adults' cognitive impairments: which role for virtual reality technology? *Computers in Biology & Medicine*, *25*, 213–227.

Pugnetti, L., Mendozzi, L., Attree, E., Barbarieri, E., Brooks, B. M., Cazzullo, C. L., et al. (1998). Probing memory and executive functions with virtual reality: past and present studies. *CyberPsychology & Behavior, 1*, 151–162.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*(1), 33-65.

Rabin, L.A., Paolillo, E., & Barr, W.B. (2016). Stability in test-usage practices in clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology, 31*, 206-230.

Rabinowitz, A.R., Li, X., McCauley, S.R., Wilde, E.A., Barnes, A., Hanten, G., et al. (2015). Prevalence and Predictors of Poor Recovery from Mild Traumatic Brain Injury. *Journal of Neurotrauma, 32*, 1488-96.

Radanov, B.P., Dvorak, J., & Valach, L. (1992). Cognitive deficits in patients after soft tissue injury of the cervical spine. *Spine, 17*(2), 127–131.

Rand, D., Weiss, P.L, & Katz, N. (2009). Training multitasking in a virtual supermarket: A novel intervention after stroke. *The American Journal of Occupational Therapy, 63*, 535-542.

Rao, N., Rosenthal, M., Cronin-Stubbs, D., Lambert, R., Barnes, P., & Swanson, B. (1990). Return to work after rehabilitation following traumatic brain injury. *Brain Injury, 4*(1), 49-56.

Raskin, S. (1997). The relationship between sexual abuse and mild traumatic brain injury. *Brain Injury, 11*(8), 587-603.

Raskin, S., & Rearick, E. (1996). Verbal fluency in individuals with mild traumatic brain injury. *Neuropsychology, 10*(3), 416-422.

Raspelli, S., Carelli, L., Morganti, F., Poletti, B., Corra, B., Silani, V., et al. (2010). Implementation of the multiple errands test in NeuroVR-supermarket: A possible approach. *Studies in Health, Technology and Informatics, 154*, 115-119.

Raspelli, S., Pallavicini, F., Carelli, L., Morganti, F., Poletti, B., Corra, B., et al. (2011). Validation of a Neuro Virtual Reality-based version of the multiple errands test for the assessment of executive functions. *Studies in Health Technology and Informatics, 167*, 92-97.

Ratcliff, G., Colantonio, A., Escobar, M., Chase, S., & Vernich, L. (2005). Long-term survival following traumatic brain injury. *Disability & Rehabilitation, 27*, 305-314.

Redmill, D.A., McIlwee, A., McNicholl, B., & Templeton, C. (2011). Long term outcomes 12 years after major trauma. *Injury, 37*, 243–246.

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

Reitan, R M. (1979). Manual for administration of neuropsychological test batteries for adults and children. Tuscon, AZ: Reitan Neuropsychological Laboratory.

Rey, A. (1941). L'examen psychologie dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 286-340

Richard, N. M., O'Connor, C., Dey, A., Robertson, I. H., & Levine, B. (2018). External modulation of the sustained attention network in traumatic brain injury. *Neuropsychology, 32*(5), 541-553.

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectrums, 11,* 35– 44.

Rizzo, A., Bowerly, T., Buckwalter, G., Schultheis, T., Matheis, R., Shahabi, C., et al. (2002). Virtual environments for the assessment of attention and memory processes: the virtual classroom and office. In P. Sharkey, C. S. Lanyi, & P. Standen (Eds.), *Proceedings of the 4th International Conference on Disability, Virtual Reality, and Associated Technology* (pp. 3-12). Reading, UK: University of Reading.

Rizzo, A., Buckwater, G., van der Zaag, C., Bowerly, T., Humphrey, L., Neumann, C., et al. (2000). The virtual classroom: a virtual environment for the assessment and rehabilitation of attention deficits. *CyberPsychology and Behavior, 3,* 483-499.

Rizzo, A., & Koenig, S.T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology, 31*(8), 877-899.

Rizzo, A., & Kim, G.J. (2005). A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence, 14*(2), 119-146.

Rizzo, A., Schultheis, M., Kerns, K., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation, 14*(1/2), 207-239.

Robertson, K., & Schmitter-Edgecombe, M. (2017). Naturalistic tasks performed in realistic environments: A review with implications for neuropsychological assessment. *The Clinical Neuropsychologist, 31*(1), 16-42.

Robertson, I.H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Bury St Edmunds, England: Thames Valley Test Company.

Robertson, I.H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The test of everyday attention. *Journal of the International Neuropsychological Society, 2*, 525-534.

Robinson, S.J., & Brewer, G. (2016). Performance on the traditional and the touch screen, tablet versions of the Corsi Block and the Tower of Hanoi tasks. *Computers in Human Behavior, 60*, 29-34.

Rochat, L., Ammann, J., Mayer, E., Annoni, J. M., & Van der Linden, M. (2009). Executive disorders and perceived socio-emotional changes after traumatic brain injury. *Journal of Neuropsychology, 3,* 213-227.

Rogers, J. C., & Holm, M. B. (2016). Functional assessment in mental health: lessons from occupational therapy. *Diologues Clinical Neuroscience, 18*, 145-154.

Rohling, L. M., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005) and Pertab et al. (2009). *The Clinical Neuropsychologist, 25*, 608–623.

Rohling, M. L., Larrabee, G. J., & Millis, S. R. (2012). The "miserable minority" following mild head traumatic brain injury: Who are they and do meta-analysis hide them? *The Clinical Neuropsychologist, 26*(2), 197-213.

Rolls, E.T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex, 10*, 284-294.

Rose, F.D., Attree, E.A., Brooks, B.M., & Johnson, D.A. (1998). Virtual reality in brain damage: A rationale from basic neuroscience. In: G. Riva, B.K. Wiederhold, & E. Molinari (Eds.), Virtual environments in clinical psychology: Scientific and technological challenges in advanced patient-therapist interaction (pp. 233-242). Amsterdam: IOS Press.

Rose, F.D., Brooks, B.M., & Rizzo, A.A. (2005). Virtual reality in brain damage rehabilitation: Review. *CyberPsychology & Behavior, 8*(3)*,* 241-262.

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin, 85,* 188-193.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. (rev. ed.) Beverly Hills, CA: Sage Publications.

Rosenthal, R. (1995). *Writing meta-analytic reviews*. Psychological Bulletin, 118, 183–192.

Ruff, R. (2003). A friendly critique of neuropsychology: Facing the challenges of our future. *Archives of Clinical Neuropsychology, 18*, 847-864.

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

Ruff, R.M., & Allen, C.C. (1996). *Ruff 2 & 7 Selective Attention Test Professional Manual*. Odessa, FL: PAR.

Ruff, R. M., Camenzuli, L., & Mueller, J. (1996). Miserable minority: Emotional risk factors that influence the outcome of a mild traumatic brain injury. *Brain Injury, 10*(8), 551-565.

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

Ruff, R. M., & Jamora, C. W. (2009). Myths and mild traumatic brain injury. *Psychological Injury and the Law*, *2*, 34–42.

Ruff, R. M., Marshall, L.F., Crouch, J., Klauber, M.R., Levin, H.S., Barth, J., et al. (1993). Predictors of outcome following severe head trauma: follow-up data from the Traumatic Coma Data Bank. *Brain Injury, 7*(2), 101-111.

Ruffolo, C.F., Friedland, J.F., Dawson, D.R., Colantonio, A., & Lindsay, P. (1999). Mild traumatic brain injury from motor vehicle accidents: Factors associated with return to work. *Archives of Physical Medicine Rehabilitation, 80,* 392-398.

Ruffolo, L.F., Guilmette, T.J., & Willis, W.G. (2000). Comparison of time and error rates on the trail making test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. *The Clinical Neuropsychologist, 14*(2), 223-230.

Runge, J.W. (1993). The cost of injury. *Emergency Medicine Clinics of North America, 11*(1), 241–253.

Rutland-Brown, W., Langois, J.A., Thomas, K.E., & Xi Y.L. (2006). Incidence of traumatic brain injury in the United States 2003. *Journal of Head Trauma Rehabilitation, 21*(6), 544-548.

Ryan, T. V., Sautter, S. W., Capps, C. F., Meneese, & Barth, J. T. (1992). Utilizing neuropsychological measures to predict vocational outcome in a head trauma population. *Brain Injury, 6,* 175–182.

Saltychev, M., Eskola, M., Tenovuo, O., & Laimi, K. (2013). Return to work after traumatic brain injury: Systematic review. *Brain Injury, 27*(13-14), 1516-1527.

Sato, T. (1996). Type I and type II error in multiple comparisons. *The Journal of Psychology: Interdisciplinary and Applied, 130*(3), 293-302.

Sauzeon, H., N'Kaoua, B., Arvind-Pala, P., Taillade, M., Auriacombe, S., & Guitton, P., (2016). Everyday-like memory for objects in ageing and Alzheimer's disease assessed in a visually complex environment: The role of executive functioning and episodic memory. *Journal of Neuropsychology, 10,* 33-58.

Sbordone, R. J. (1996). Ecological validity: Some critical issues for the neuropsychologist. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 55–41). Delray Beach, FL: GR Press/St. Lucie Press.

Sbordone, R.J. (2001). Limitations of neuropsychological testing to predict the cognitive and behavioral functioning of persons with brain injury in real-world settings. *NeuroRehabilitation, 16,* 199-201.

Sbordone, R.J., Purisch, A.D. (1996). Hazards of blind analysis of neuropsychological test data in assessing cognitive disability: The role of psychological pain and other confounding factors. *NeuroRehabilitation, 7,* 15–26.

Schmidt, M., Trueblood, W., Merwin, M., & Durham, R. L. (1994). How much do "attention" tests tell us? *Archives of Clinical Neuropsychology, 9*(5), 383–394.

Schmitter-Edgecombe, M. (1996). Effects of divided attention on implicit and explicit memory performance following severe closed head injury. *Neuropsychology, 10,* 155–167.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15,* 341-349.

Schulman, J., Sacks, J., & Provenzano, G. (2002). State level estimates of the incidence and economic burden of head injuries stemming from non-universal use of bicycle helmets. *Injury Prevention, 8,* 47-52.

Schultheis, M. T., Himelstein, J., & Rizzo, A.A. (2002). Virtual reality and neuropsychology: Upgrading the current tools. *Journal of Head Trauma Rehabilitation, 17*(5), 378-394.

Schultz, I.Z., Law, A.K., & Cruikshank, L.C. (2016). Prediction of occupational disability from psychological and neuropsychological evidence in forensic context. *International Journal of Law and Psychiatry, 49,* 183-196.

Selassie, A.W., Zaloshnja, E., Langlois, J.A., Miller, T., Jones, P., & Steiner, C. (2008). Incidence of long-term disability following traumatic brain injury hospitalization, United States, 2003. *Journal of Head Trauma Rehabilitation, 23*(2), 123–131.

Serino, S., Morganti, F., Di Stefano, F., & Riva, G. (2015). Detecting early egocentric and allocentric impairments deficits in Alzheimer's disease: an experimental study with virtual reality. *Frontiers in Aging Neuroscience, 7,* 88.

Seymour N. E., Gallagher A. G., Roman S. A., O'Brien, M. K., Bansal, V. K., Anderson, D. K., et al. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery, 236*, 458–63.

Shafi, S., Marquez, D.L.P., Diaz-Arrastia, R., Shipman, K., Carlile, M., Frankel, H., et al. (2007). Racial disparities in long-term functional outcome after traumatic brain injury. *Journal of Trauma, 63*, 1263–1268.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727-741.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *351*, 1405-1412.

Shames, J., Treger, I., Ring.H., & Giaquinto, S. (2007). Return to work following traumatic brain injury: Trends and challenges. *Disability and Rehabilitation, 29*(17), 1387-1395.

Shandera-Ochsner, A.L., Berry, D.T., Harp, J.P., Edmundson, M., Graue, L.O., Roach, A., et al. (2013). Neuropsychological effects of self-reported deployment-related mild TBI and current PTSD in OIF/OEF veterans. *Clinical Neuropsychologist, 27*, 881-907.

Sharp, D.J., Beckmann, C.F., Greenwood, R., Kinnunen, K.M., Bonnelle, V., De Boissezon, X., et al. (2011). Default mode network functional and structural connectivity after traumatic brain injury. *Brain, a Journal of Neurology, 134*, 2233-2247.

Sharp, D.J, & Jenkins, P.O. (2015). Concussion is confusing us all. *Practical Neurology, 15*, 172-186.

Sherer, M., Bergloff, P., Levin, E., High, W. J., Oden, K. E., & Nick, T. G. (1998). Impaired awareness and employment outcome after traumatic brain injury. *Journal of Head Trauma and Rehabilitation, 13*(5), 52–61.

Sherer, M., Madison, C.F., & Hannay, H.J. (2000). A review of outcome after moderate and severe closed head injury with an introduction to life care planning. *Journal of Head Trauma Rehabilitation, 15,* 767–782.

Sherer, M., Novack, T.A., Sander, A.M., Struchen, M.A., Alderson, A., & Thompson, R. N. (2002). Neuropsychological assessment and employment outcome after traumatic brain injury: a review. *The Clinical Neuropsychologist, 16*(2), 157 – 178.

Shum, D.H.K., McFarland, K.A., & Bain, J.D. (1990). Construct validity of eight tests of attention: Comparison of normal and closed head injured samples. *The Clinical Neuropsychologist, 4*, 151-162.

Signoretti, S., Vagnozzi, R., Tavazzi, B., & Lazzarino, G. (2010). Biochemical and neurochemical sequelae following mild traumatic brain injury: Summary of experimental data and clinical implications. *Neurosurgery Focus, 29*(5), E1-E12.

Sigurdardottir, S., Andelic, N., Roe, C., & Schanke, A.K. (2009). Cognitive recovery and predictors of functional outcome 1 year after traumatic brain injury. *Journal of the International Neuropsychological Society, 15*, 740-750.

Silverberg, N. D., & Millis, S. R. (2009). Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society, 15*, 94-102.

Slick, D. (1997). *VSVT, Victoria Symptom Validity Test: Version 1.0, professional manual*. Odessa, FL: Psychological Assessment Resources.

Sloan, S., & Ponsford, J. (1995). Assessment of cognitive dif®culties following TBI. In J. Ponsford, S. Sloan, & P. Snow (Eds.), Traumatic brain injury: Rehabilitation for everyday adaptive living (pp. 65-102). Hove, UK: Lawrence Erlbaum Associates.

Smith, A. (1993). *Symbol Digit Modalities Test*. Los Angeles: Western Psychological Services.

Smith-Knapp, K., Corrigan, J.D., & Arnett, J.A. (1996). Predicting functional independent from neuropsychological tests following traumatic brain injury. *Brain Injury, 10*(9), 651-661.

Soberg, H.L., Roise, O., Bautz-Holter, E., & Finset, A. (2011). Returning to work after severe multiple injuries: Multidimensional functioning and the trajectory from injury to work at 5 years. *Journal of Trauma, 71*, 425–434.

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach*. New York, NY: Guilford.

Solbakk, A. K., Reinvang, I., & Nielsen, C. S. (2000). ERP indices of resource allocation difficulties in mild head injury. *Journal of Clinical and Experimental Neuropsychology, 22*, 743–760.

Solbakk, A. K., Reinvang, I., Neilsen, C., & Sundet, K. (1999). ERP indicators of disturbed attention in mild closed head injury: A frontal lobe syndrome? *Psychophysiology, 36*, 802–817.

Spitz, G., Ponsford, J., & Schonberger, M. (2013). The relationship between cognitive impairment, coping style, and emotional adjustment following traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 28*(2), 116–125.

Spreen, O., & Strauss, I. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Harcourt Brace & Company.

Spooner, D.M., & Pachana, N.A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology, 21*, 327-337.

Steadman-Pare, D., Colantonio, A., Ratcliff, G., Chase, S., & Vernich, L. (2001). Factors associated with perceived quality of life many years after traumatic brain injury. *Journal of Head Trauma Rehabilitation, 16*, 330-42.

Sterr, A., Herron, K. A., Hayward, C., & Montaldi, D. (2006). Are mild head injuries as mild as we think? neurobehavioral concomitants of chronic post-concussion syndrome. *BMC Neurology, 6*, 7.

Steudel, W., Cortbus, F., & Schwerdtfeger, K. (2005). Epidemiology and prevention of fatal head injuries in Germany – trends and the impact of the reunification. *Acta Neurochir (Wien), 147*, 231–242.

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

Stone, R.J. (2002). Applications of virtual environments: An overview. In K.S. Hale, & K.M Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 827-856). Mahwah, NJ: CRC Press.

Storzbach, D., O'Neil, M.E., Roost, S.M., Kowalski, H., Iverson, G.L., Binder, L.M., et al. (2015). Comparing the Neuropsychological Test Performance of Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) Veterans with and without Blast Exposure, Mild Traumatic Brain Injury, and Posttraumatic Stress Symptoms. *Journal of the International Neuropsychological Society*, *21*, 353-63.

Strich, S. J. (1970). Lesions in the Cerebral Hemispheres after Blunt Head Injury. *Journal of Clinical Pathology*, *23*(4), 166-171.

Studerus-Germann, A.M., Engel, D.C., Stienen, M.N., von Ow, D., Hildebrant, G., & Gautschi, O. P. (2017). Three versus seven days to return-to-work after mild traumatic brain injury: a randomized parallel-group trial with neuropsychological assessment. *International Journal of Neuroscience, 127*(10), 1-9.

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 362*, 901- 915.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.

Stuss, D.T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review Psychology, 53*, 401-433.

Stuss, D.T., Stethem, L.L., Hugenholtz, H., Picton, T., Pivik, J., & Richard, M.T. (1989). Reaction time after head injury: Fatigue, divided and focused attention, and consistency of performance. Journal of Neurology, Neurosurgery, and Psychiatry, 52, 742–748.

Suhr, J. A., & Gunstad, J. (2002). Postconcussive symptom report: The relative influence of head injury and depression. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 981-993.

Swan, A.R., Nichols, S., Drake, A., Angeles, A., Diwakar, M., Song, T., et al. (2015). Magnetoencephalography slow-wave detection in patients with mild traumatic brain injury and ongoing symptoms correlated with long-term neuropsychological outcome. *Journal of Neurotrauma, 32*, 1510-1521.

Sweet, J.J., Benson, L.M., Nelson, N.W., & Moberg, P.J. (2015). Initial findings from the American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society of Clinical Neuropsychology (APA Division 40) 2015 TCN professional practice and "salary survey". *The Clinical Neuropsychologist, 29*(3), 308-393.

Tabachnick, B., & Fidell, L. (1996). *Using multivariate statistics*. New York, NY: HarperCollins.

Tagliaferri, F., Compagnone, C., Korsic, M., Servadei, F., & Kraus, J. (2006). A systematic review of brain injury epidemiology in Europe. *Acta Neurochir, 148*(3), 255-268.

Taylor, C.A., Bell, J.M., Breidling, M.J., & Xu, L. (2017). Traumatic brain injury-related emergency department visits, hospitalizations, and deaths-United States, 2007 and 2013. *Surveillance Summaries, 66*(9), 1-16.

Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness. *Lancet, ii,* 81 – 84.

Temkin, N.R., Holubkov, R., Machamer, J.E., Winn, H.R., & Dikmen, S.S. (1995). Classification and regression trees (CART) for prediction of function at 1 year following head trauma. *Journal of Neurosurgery, 82*(5), 764-771.

Temkin, N. R., Machamer, J. E., & Dikmen, S. S. (2003). Correlates of functional status 3–5 years after traumatic brain injury with ct abnormalities. *Journal of Neurotrauma, 20*(3), 229-241.

Thornhill, S., Teasdale, G.M., Murray, G.D., McEwen, J., Roy, C.W., & Penny, K. I. (2000). Disability in young people and adults one year after head injury: Prospective cohort study. *British Medical Journal, 320*, 1631-1635.

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

Thurman, D. J., Sniezek, J. E., Johnson, D., Greenspan, A., & Smith, S.M. (1995). *Guidelines for Surveillance of Central Nervous System Injury.* Atlanta (GA): Centers for Disease Control and Prevention.

Tiersky, L.A., Cicerone, K.D., Natelson, B.H., & DeLuca, D. (1998). Neuropsychological functioning in chronic fatigue syndrome and mild traumatic brain injury: A comparison. *The Clinical Neuropsychologist*, *12*(4), 503-512.

Tipper, S.P., & Baylis, G.C. (1987). Individual differences in selective attention: The relation of priming and interreference to cognitive failure. *Personality and Individual Differences, 8*(5), 667-675.

Tippett, J. W., Lee, J., Zakzanis, K. K., Black, E. S., Mraz, R., & Graham, S. J. (2009). Visually navigating a virtual world with real-world impairments: a study of visually and spatially guided performance in individuals with mild cognitive impairments. *Journal of Clinical and Experimental Neuropsychology, 31*(4), 447-454.

Titov, N., & Knight, R.G. (2005). A computer-based procedure for assessing functional cognitive skills in patients with neurological injuries: The virtual street. *Brain Injury, 19*(5), 315–322.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, NY: Multi Health Systems.

Tranel, D., Hathaway-Nepple, J., & Anderson, S. W. (2007). Impaired behavior on real-world tasks following damage to the ventromedial prefrontal cortex. *Journal of Clinical and Experimental Neuropsychology, 29*, 319–332.

Trepagnier C. G. (1999). Virtual environments for the investigation and rehabilitation of cognitive and perceptual impairments. *Neurorehabilitation, 12,* 63–72.

Valve Corporation. (2018). *Steam*. Retrieved from: https://store.steampowered.com/

van der Ham, I.J., Faber, A.M., Venselaar, M., Kreveld, M.J., & Loffler, M. (2015). Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology, 6, A*637.

van der Leeuw, G., Leveille, S.G., Jones, R.N., Hausdorff, J.M., McLean, R., Kiely, D. K., et al. (2017). Measuring attention in very old adults using the Test of Everyday Attention. *Aging, Neuropsychology, and Cognition, 24*(5), 543-554.

van der Naalt, J., van Zomeren, A. H., Sluiter, W. J., & Minderhoud, J. M. (1999). One year outcome in mild to moderate head injury: The predictive value of acute injury characteristics related to complaints and return to work. *Journal of Neurology, Neurosurgery, and Psychiatry, 66*(2), 207-213.

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

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

Vanderploeg, R. D., Curtiss, G., Duchnick, J.J., & Luis, C.A. (2003). Demographic, medical, and psychiatric factors in work and marital status after mild head injury. *Journal of Head Trauma Rehabilitation, 18*(2), 148-163.

Vanderploeg, R. D., Curtiss, G., Luis, C. A., & Salazar, A. M. (2007). Long-term morbidities following self-reported mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 29*, 585–598.

Vayalakkara, J., Devaraju-Backhaus, S., Bradley, J.D., Simco, E.R., & Golden, C.J. (2000). Abbreviated form of the Wisconsin Card Sort Test. International Journal of Neuroscience, 103, 131–137.

Veeramuthu, V., Narayanan, V., Kuo, T.L., Delano-Wood, L., Chinna, K., Bondi, M.W., et al. (2015). Diffusion Tensor Imaging Parameters in Mild Traumatic Brain Injury and Its Correlation with Early Neuropsychological Impairment: A Longitudinal Study. *Journal of Neurotrauma, 32*, 1497-509.

Virtual reality. (n.d.). Retrieved August 18, 2018, from https://www.merriam-webster.com/dictionary/virtual%20reality

Wäljas, M., Iverson, G. L., Lange, R. T., Liimatainen, S., Hartikainen, K. M., Dastidar, P., et al. (2014). Return to work following mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *29*(5), 443–450.

Watt, S., & Crowe, S. F. (2018). Examining the beneficial effect of neuropsychological assessment on adult patient outcomes: a systematic review. *Clinical Neuropsychologist*, *32*(3), 368–390.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised. Manual.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised. Manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS-III).* (3rd Ed.). San Antonio, TX: Harcourt Assessment.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.

Wehman, P., Brieout, J., Targett, P. (2017). Supported employment for persons with traumatic brain injury: A guide for implementation. In R. Fraiser (Eds.), *Traumatic Brain Injury Rehabilitation*. New York, NY: Taylor and Francis.

Wehman, P., Targett, P., West, M., & Kregel, J. (2005). Productive work and employment for persons with traumatic brain injury: what have we learned after 20 years? *Journal of Head Trauma Rehabilitation, 20*, 115-27.

Weiss, P. L., Bialik, P., & Kizony, R. (2003). Virtual reality provides leisure time opportunities for young adults with physical and intellectual disabilities. *CyberPsychology & Behavior, 6,* 335-342.

Whelon, B. M., & Murdoc, B. E. (2013). The impact of mild traumatic brain injury (mTBI) on language function: More than meets the eye? *Brain and Language, 99*, 183-184.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., et al. (2016). Verbal Fluency: Language or Executive Function Measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. 2016. Retrieved from http://dx.doi.org/10.1080/23279095.2015.1004574

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594 – 604.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test--Fourth Edition*. Lutz, FL: Psychological Assessment Resources.

Wilkins, A. J., Shallice, T., & McCarthy, R. (1987). Frontal lesions and sustained attention. *Neuropsychologia*, *25,* 359–365.

Willemse-van Son, A. H., Ribbers, G. M., Verhagen, A. P., & Stam, H. J. (2007). Prognostic factors of long-term functioning and productivity after traumatic brain injury: A systematic review of prospective cohort studies. *Clinical Rehabilitation, 21*, 1024–1037.

Wilson, B.A. (1993). Ecological validity of neuropsychological assessment: do neuropsychological indexes predict performance in everyday activities? *Applied & Preventive Psychology, 2,* 209 – 215.

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural assessment of the dysexecutive syndrome: Test manual*. England: Thames Valley Test Company.

Wilson, B.A., Evans, J.J., Emslie, H., Alderman, N., & Burgess, P. 1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation, 8,* 213–228.

Wilson, P. N., Foreman, N., & Stanton, D. (1997). Virtual reality, disability and rehabilitation. *Disability and Rehabilitation, 19*(6), 213-220.

Wither, F. K., Brouwer, W. H., van Zomeren, A. H. (2000). Fitness to drive in older drivers with cognitive impairment. *Journal of the International Neuropsychological Society, 6*(4), 480-490.

Wood, R. L., & Liossi, C. (2007). The relationship between general intellectual ability and performance on ecologically valid executive tests in a severe brain injury sample. *Journal of the International Neuropsychological Society, 13,* 90-98.

Wood, R. L. L., & Rutterford, N. A. (2006). Demographic and cognitive predictors of long-term psychosocial outcome following traumatic brain injury. *Journal of the International Neuropsychological Society*, *12,* 350-358.

Wolf, T.J., Morrison, T., & Matheson, L. (2008). Initial development of a work-related assessment of dysexecutive syndrome: The complex task performance assessment. *Work*, *31,* 221-228.

World Health Organization. (2006). *Neurological disorders: Public health challenges*. Geneva: World Health Organization.

Wrightson, P., & Gronwall, D. (1999). *Mild head injury: A guide to management.* New York, NY: Oxford University Press.

Yamout, B., Issa, Z., Herlopian, A., El Bejjani, M., Khalifa, A., Ghadieh, A.S., et al. (2013). Predictors of quality of life among multiple sclerosis patients: A comprehensive analysis. *European Journal of Neurology, 20,* 756-764.

Yasuda, S., Wehman, P., Targett, P., Cifu, D., & West, M. (2001). Return to work for persons with traumatic brain injury. *American Journal of Physical Medicine & Rehabilitation, 80*(11), 852-864.

Zakzanis, K. K. (1998). Brain is related to behaviour (p < .05)*. *Journal of Clinical and Experimental Neuropsychology, 20*(3), 419-427).

Zakzanis, K. K. (2001). Statistics to tell the truth, the whole truth, and nothing but the truth: Formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology, 16,* 654-667.

Zakzanis, K.K., & Grimes, K. (2016). Relationship among apathy, cognition, and real-world disability after mild traumatic brain injury. *Applied Neuropsychology; Adult*, *24*(6), 559-565.

Zakzanis, K. K. & Jeffay, E. (2011). Cognitive variability in high-functioning individuals: Implications for understanding the neuropsychology of mild traumatic brain injury. *Psychological Reports*, *108*, 290-300.

Zakzanis, K.K., Leach, L., & Kaplan, E. (1999). Mild traumatic brain injury. In *Neuropsychological differential diagnosis* (pp.163-171). Exton, Pennsylvania: Swets & Zeitlinger.

Zakzanis, K. K., McDonald, K., Troyer, A. K. (2011). Component analysis of verbal fluency in patients with mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *33*(7), 785-792.

Zakzanis, K. K., Quintin, G., Graham, S. J., & Mraz, R. (2009). Age and dementia related differences in spatial navigation within an immersive virtual environment. *Medical Science Monitor*, *15*(4), 140-150.

Zaloshnja, E., Miller, T., Langlois, J.A., & Selassie, A.W. (2008). Prevalence of long-term disability from traumatic brain injury in the civilian population of the United States, 2005. *Journal of Head Trauma Rehabilitation, 23*, 394–400.

Ziino, C., & Ponsford, J. (2006). Selective attention deficits and subjective fatigue following traumatic brain injury. *Neuropsychology*, *20*(3), 383-390.

Zillmer, E. A., Spiers, M. V., & Culbertson, W. (2008). *Principles of neuropsychology* (2nd ed.). Belmont, CA: Wadsworth.

Appendix

Office Task Instructions

"I want you to imagine that you work for a courier company and you've just arrived at an office building where you need to deliver some packages. The trouble is, the packages do not always have room numbers on them but instead they have other clues to help. You're job is to deliver the packages, one at a time, to the correct door. To pick up a package simply use the joystick to walk into the courier trolley. Once you pick up a package, you need to go to the door and press the blue button "X" on the gamepad. If you can't see the package very well, press the green button "A" and it will be made more visible. The computer will tell you if you are correct or incorrect. If you are wrong, try again and see if you can figure it out. There is no time on this test. Ready? Let's begin."

Conveyor Belt Task (CBT) Instructions

<u>*Selective Attention*</u>

"You are a quality control specialist at a Tic-Tac-Toe Board making factory and need to ensure that the products all have 9 squares on the board. Only the RIGHT conveyor belt works. Your job is to press the 'V' button on the keyboard whenever a defective product is rolled down the belt. A defective product is one with an octagon shape in it. (draw if necessary)"

For every successful detection, the participant will receive 100 points. For every false detection 50 points will be deducted from their score.

In the difficultly level 1, the participant is distracted by squares. The target shape is an octagon.

In the difficultly level 2, the participant is distracted by squares, parallelograms and rounded squares. The target shape is an octagon.

In the difficultly level 3, the participant is distracted by circles. The target shape is an octagon.

In the difficultly level 4, the participant is distracted by hexagons and circles. The target shape is an octagon.

<u>*Sustained Attention*</u>

"You are a quality control specialist for a toy company and the product sorting machine is malfunctioning."

Difficulty 1: *"Normally, the machine separates the different products onto different conveyor belts. However, the machine is putting every type of product onto the GLOBE conveyor belt. Your job is to press the 'V' key whenever you see something that is NOT a globe rolling down the conveyor belt."*

Difficulty 2: *"In this round, the belt is fixed but the globes seem to be falling over. Your job is to press the 'V' key as soon as you see a globe fall so it can be checked for damage. It's important that you press the 'V' key immediately otherwise it won't be checked for damage."*

Difficulty 3: *"In this round, some of the globes are missing a coat of rust proof paint. Your job is to detect the rusted globes by pressing the 'V' key whenever you see a globe rust. It's important that you press the 'V' key immediately otherwise it won't be checked for damage."*

Difficulty 4: *"In this round, a weight has been installed in the stand of the globe to prevent it from moving or falling over. Some of the globes do not have this installed causing them to rotate slightly. Your job is to press the 'V' key whenever you see a globe rotate slightly. It's important that you press the 'V' key immediately otherwise it won't be checked for damage."*

For every successful detection, the participant will receive 100 points. For every false detection, 50 points will be deducted from their score.

In the difficultly level 1, the participant is distracted by globes. The target shapes are any toys other than a globe.

In the difficultly level 2, the participant is distracted by globes. The target is any globe that falls.

In the difficultly level 3, the participant is distracted by globes. The target is any globe that rusts and turns orange.

In the difficultly level 4, the participant is distracted by globes. The target is any globe that rotates slightly.

Divided Attention

"*You are a quality control specialist at a Tic-Tac-Toe Board making factory and need to ensure that the products all have 9 squares on the board. You are in charge of two conveyor belts. Your job is to press the 'Z' (left conveyor belt) and 'V' (right conveyor belt) button on the keyboard whenever a defective product is rolled down the belt."*

Difficulty level 1 and 2 are identical in administration with the exception of the conveyor belts moving at a faster speed.

<u>*Attention Shift*</u>

"You are a quality control specialist at a Tic-Tac-Toe Board making factory and need to ensure that the products all have 9 squares on the board. You are in charge of two conveyor belts. Unfortunately, an error detecting arrow is also malfunctioning in that it sometimes points to the conveyor belt that has the defect and at times does not. Your job is to press the 'Z' (left conveyor belt) and 'V' (right conveyor belt) button on the keyboard whenever a defective product is rolled down the belt, regardless of which way the arrow is pointing."

As with the Divided Attention module, difficulty level 1 and 2 are identical in administration with the exception of the conveyor belts moving at a faster speed.

Conveyor Belt Task Pilot and Parameters

A pilot set of healthy controls recruited from the Psychology Department's undergraduate research pool were used to determine the parameters surrounding difficulty levels. It was determined that manipulating the speed of stimulus presentation, frequency of the target object, and size of the target object moderated perceived difficulty of the task. Formal statistical analysis was not completed and since the task was entirely novel, the initial values were estimated and adjusted on a trial-by-trial basis. Table 13 displays the variables and values that constitute each difficulty level of the CBT. A brief description of the variables follows.

Trial – The number of trials administered.

Speed – The speed at which the stimuli are presented at. The units are in an ordinal scale, where smaller values indicate faster presentations of the stimuli.

Starting Point – The number of points that are given at the start of the test.

Correct Reward – The number of points that are given for each correct response.

Error Debit – the number of points deducted after each incorrect response.

Frequency – The frequency of the presentation of the target stimuli. Higher values represent higher frequency.

Appearance % - the percentage of target stimuli shown across all trials.

Separation Time – the inter-trial interval between the presentation of the target stimuli only. Higher values represent longer inter-trial intervals.

Limitation – Time interval after the target stimuli has appeared for responses to be counted. Higher values represent longer durations.

Bias Amount % – In the Attention Shift subtest, an arrow is presented in the middle of the two conveyor belts. This variable sets the percentage whereby the arrow is pointing to the conveyor belt that has the target stimuli.

Table 13.

Difficulty level parameters of the Conveyor Belt Task subtests.

	Selective Attention				Sustained Attention				Divided Attention		Attention Shift	
Difficulty Level	1	2	3	4	1	2	3	4	1	2	1	2
Trials	30	30	30	30	50	50	50	50	30	30	30	30
Speed	4	3	2.5	2	2	3	2	2	3	2	4	2
Starting Point	0	0	0	0	0	0	0	0	0	0	0	0
Correct Reward	100	100	100	100	100	100	100	100	100	100	100	100
Error Debit	50	50	50	50	50	50	50	50	50	50	50	50
Frequency	2	2	2	2	0.5	1	1	1	2	1	2	1
Appearance %	0.5	0.5	0.5	0.5	0.2	0.2	0.2	0.2	N/A	N/A	N/A	N/A
Separation Time	N/A	N/A	N/A	N/A	0.5	0.5	0.5	0.25	N/A	N/A	N/A	N/A
Limitation	N/A	N/A	N/A	N/A	1	1	1	1	N/A	N/A	N/A	N/A
Bias Amount %	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.8	0.8

Abbreviation: SD (standard deviation)