The WAIS Coding subtest as an embedded performance validity measure in cases of traumatic brain injury

Arta Krievina

A thesis submitted to Auckland University of Technology in partial fulfillment of the requirements for the degree of Master of Health Science in Psychology

2020

School of Public Health and Psychosocial Studies

Faculty of Health and Environmental Sciences

Primary Supervisor: Doctor Susan Mahon

Mentor: Doctor James Webb

Abstract

Most recent estimates have indicated that around 69 million people worldwide sustain traumatic brain injury (TBI) every year. Often following TBI, a neuropsychological assessment is conducted in order to determine the existence and the extent of the cognitive impairment. A good effort is required from the examinee to obtain valid results on neuropsychological tests. However, extensive literature has demonstrated that not everybody performs to the best of their ability and that some may feign or exaggerate their cognitive deficits. As a result, performance validity testing (PVT) research is considered to be crucial and is currently one of the most dominant themes in the field of neuropsychology.

The aim of the current study was to explore the efficacy of a number of embedded validity indicators derived from the WAIS Coding subtest in a mixed traumatic brain injury sample (*n* = 650), namely: presence of a Coding Error; the Number of Coding Errors; the Coding ACSS; a derived Coding Combination score (i.e. the sum of adding the Coding Error and Coding ACSS). The study also examined if these Coding embedded validity indicators are valid in the range of traumatic brain injury severities or if they are biased against those with more significant brain injuries.

Results of logistic regression analyses revealed that all Coding embedded effort variables were significantly predictive of low effort. Furthermore, findings indicated that the injury severity affects performance on all Coding embedded validity indicators. Results showed an inverse relationship between presence of a Coding Error and injury severity such that individuals with more severe brain injuries were less likely to make Coding Errors than those affected by mild traumatic brain injuries. The Coding ACSS and the Coding Combination score were sensitive to TBI severity, such that those with more severe brain injuries obtained lower scores on the ACSS and higher scores on the Coding Combination score, indicating a more impaired performance. Based on these findings, a ROC analysis was employed to identify recommended cut-off scores for each severity group with an aim to minimise false positive errors. Furthermore, a positive predictive value (PPV) and negative predictive value (NPV) were reported for 20%, 30%, 40%, and 50% base rates.

Overall, results demonstrated excellent specificity rates, but low and variable sensitivity rates. Employing a Coding Error cut-off score of > 0 for all severity groups resulted in 97% specificity and 38% sensitivity. Based on the analysis of the Coding ACSS, in cases of mTBI a cut-score of ≤ 5 resulted in 91% specificity and 56% sensitivity; for cases of Moderate TBI employing a cut-score of ≤ 4 resulted in 94% specificity and 33% sensitivity; in cases of Severe and Very Severe TBI employing a cut-score of ≤ 3 resulted in 94% specificity and 57% sensitivity. Results showed that the Coding Combination score achieved higher sensitivity than the Coding ACSS for the same level of specificity. Based on the findings in this study, in cases of mild TBI, a Coding Combination cut-score of > 14 resulted in 91% specificity and 63% sensitivity; in cases of Moderate TBI, employing a cut-score of > 15 resulted in 90% specificity and 40% sensitivity; in cases of Severe and Very Severe TBI, employing a cut-score of > 16 resulted in 93% specificity and 70% sensitivity.

Overall, the findings in this study show that the Coding embedded validity indicators show good potential. However, due to low and variable sensitivity levels, they should be used in a combination with other PVTs in order to determine performance validity during neuropsychological evaluation and are likely to be used as complementary or confirmatory of other, more sensitive, standalone PVTs.

Table of Contents

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed:

Date: 06/11/2019

Acknowledgments

The completion of this thesis would not have been possible without help, support, and guidance from some exceptional individuals.

Firstly, I would like to thank both my supervisors Dr Susan Mahon and Dr James Webb. Susan has provided me with amazing support, not only throughout the thesis process, but also during my studies this year. Susan, your passion for rehabilitation psychology and your work ethic has been absolutely inspirational. I hope I will have an opportunity to work with you on many more projects. It has also been an honour and privilege to work with James, and to be able to research such a fascinating topic. I wish to say thank you for sharing your knowledge and experience with me, I am very grateful for your incredible support and guidance. James, I hope our paths will cross again during my future studies or on a professional level.

I also wish to thank my loving and supportive family, I am so grateful for your never-ending encouragement and belief in me. You have all endured hours on end of me only talking about my thesis without ever complaining or even just trying to change the topic.

I will never forget how much support you all have provided for me.

Ethical Approval

Ethical Approval was obtained from the Auckland University of Technology Ethics Committee (AUTEC) on 7th May 2019 with the approval number 19/107 (see Appendix A) for research undertaken in this thesis.

Chapter 1 Introduction

It has been estimated that around 69 million people sustain traumatic brain injuries (TBI) worldwide every year (Dewan et al., 2019), and of those around 36,000 people in New Zealand (NZ) sustain TBI every year, caused by recreation or sport, falls, using machinery, accidents while driving, and assaults (Accident Compensation Corporation, 2018). Often following TBI, a neuropsychological assessment is conducted in order to determine the existence and extent of the cognitive impairment, and if it is sufficient for disability or insurance payments (Herrera-Guzmán, Peña-Casanova, Lara, Gudayol-Ferré, & Böhm, 2004; Webb, Batchelor, Meares, Taylor, & Marsh, 2012). Neuropsychological data is often the only objective evidence of impairment, particularly, in cases of mild brain injuries where positive findings on neuroimaging tests and positive neurological signs are seldom evident (Inman & Berry, 2002). However, test results are only helpful if they are accurate, reliable, and valid (Barker-Collo & Fernando, 2015; Willis, Farrer, & Bigler, 2011). Knowledge about the validity of the obtained data is a prerequisite for drawing conclusions, rendering diagnosis, and making treatment recommendations following a neuropsychological evaluation (Lezak, Howieson, Bigler, & Tranel, 2012). As a result, performance validity assessment is an essential part of a complete neuropsychological evaluation (Boone, 2007b).

According to Dandachi-FitzGerald, Ponds, and Merten (2013), “one of the major tasks for neuropsychologists is to determine the objective presence of cognitive deficiencies” (p 771). Neuropsychologists measure brain function and cognitive deficits indirectly using standardised tests (Green & Merten, 2013). Historically, it was thought that differences in cognitive functioning were accurately revealed in the cognitive test scores due to the assumption that individuals, in most cases, tried to the best of their ability on given tasks (Green & Merten, 2013). Consequently, scores in the impaired range were assumed to reflect low abilities and interpreted as signs of cognitive impairment and deficits in brain function (Green & Merten, 2013). However, more recently, research has demonstrated that the assumption of people trying to the best of their ability is often false, and that low scores on neuropsychological tests do not always reflect an individual’s true abilities (Green & Merten, 2013; Green, Rohling, Lees-Haley, & Allen Iii, 2001; Stevens, Friedel, Mehren, & Merten, 2008). Some individuals give distorted or erroneous responses during their assessment, which are not representative of their true cognitive functioning (Lezak et al., 2012). Thus, test scores are said be affected by poor effort leading to invalid results, and as a result, effort testing has now been recognised as an essential part of a neuropsychological assessment (Bush et al., 2005; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). Numerous studies have also provided evidence that clinical judgment alone is not accurate enough to determine validity on the neuropsychological tests (Bianchini, Mathias, & Greve, 2001; Dandachi-FitzGerald, Merckelbach, & Ponds, 2017; Green et al., 2001), therefore objective and specialised performance validity tests (PVTs) are necessary to assess one's performance (van Gorp et al., 1999).

Consequently, currently performance validity research is considered to be one of the most dominant themes in the field of neuropsychology (Bigler, 2014; Inman & Berry, 2002). Due to the importance of developing new measures and validating existing measures, the present study was designed to explore the efficacy of a number of embedded validity indicators derived from the WAIS Coding subtest. This research was conducted with an aim to provide clinicians with additional objective tools when assessing performance validity during a neuropsychological evaluation. Furthermore, this would contribute to the identification of those individuals who are feigning or exaggerating their cognitive impairment, and provide support to credible individuals with TBIs to receive access to scarce rehabilitation resources and much-needed support.

The objectives of this thesis were to:

1. Explore the efficacy of a number of embedded effort measures derived from the WAIS-III Digit Symbol-Coding and the WAIS-IV Coding subtests in the assessment of performance validity among individuals with traumatic brain injuries
2. Investigate if Coding errors, the Coding Age-Corrected Scaled Score (ACSS), and the Combination of both are predictive of effort
3. Examine if the three Coding embedded validity indicators are valid in the range of traumatic brain injury severities or if they show bias against those with more significant brain injuries
4. Examine if demographic characteristics are associated with making errors on the WAIS Coding subtest
5. Explore if there is a greater tendency to make more errors on the WAIS III or the WAIS IV Coding subtests

Chapter 2 Literature Review

2.1 Traumatic Brain Injury (TBI)

2.1.1 Overview

Traumatic Brain Injury (TBI) has been defined as "a traumatically induced physiological disruption of brain function and/or structure resulting from the application of a biomechanical force to the head, rapid acceleration and/or deceleration, or blast related forces" (Bigler & Maxwell, 2013, p. 4). TBI is said to be a significant universal public health concern (Dewan et al., 2019; Dikmen et al., 2009; McAllister, 2008), and the leading cause of death in children and adults under the age of 35 years old (World Health Organization, 2006).

Following TBI, it is important to assess and determine the severity of injury ranging from mild traumatic brain injury (mTBI) to extremely severe injury (Tate, 2012). The purpose of classification of severity is to assist in the management of the acute stage, predict the potential for recovery, and, lastly, to identify the overall consequences of the injury incorporating the initial classification, any complications, and presentation in the post-acute phase (Teasdale, 1995). The injury severity is commonly determined based on the Glasgow Coma Scale (GCS) score, the length of post-traumatic amnesia (PTA) (i.e. loss of memory of events after the injury), and the duration of loss of consciousness (LOC) (Blyth & Bazarian, 2010; Levin & Diaz-Arrastia, 2015; Tate, 2012).

Depending on the GCS score, brain injuries are classified as severe (score of 3 to 8), moderate (score of 9 to 12), and mild (score of 13 to 15) (Tate, 2012). The duration of PTA is commonly used alongside the GCS; for example, PTA less than 5 minutes is classified as a very mild injury; PTA ranging from 5 to 60 minutes is classified as a mild brain injury; PTA ranging from 1 to 24 hours is classified as moderate injury; PTA lasting from 1 to 7 days is classified as severe injury; PTA ranging from 1 to 4 weeks is classified as a very severe injury; and PTA lasting for more than a month is considered as an extremely severe injury (Tate, 2012). Finally, loss of consciousness (LOC) is also used as a severity index for a brain injury, such that LOC ranging from 0 to 30 minutes is associated with mild injury, LOC from 30 minutes to 24 hours is associated with a moderate injury, and LOC for more than 24 hours is associated with a severe brain injury (Blyth & Bazarian, 2010).

An extensive literature review by the World Health Organization (WHO) Collaborating Centre for Neurotrauma Task Force on Mild Traumatic Brain Injury found that research studies did not use the same criteria to classify mild traumatic brain injury (mTBI) (Carroll, Cassidy, Holm, Kraus, & Coronado, 2004). As a result, authors proposed an operational definition and diagnostic criteria for mTBI, which states: “mTBI is an acute brain injury resulting from mechanical energy to the head from external physical forces” (p. 115). The diagnostic criteria for clinical identification includes:

> (i) 1 or more of the following: confusion or disorientation, loss of consciousness for 30 minutes or less, post-traumatic amnesia for less than 24 hours, and/or other transient neurological abnormalities such as focal signs, seizure, and intracranial lesion not requiring surgery; (ii) Glasgow Coma Scale score of 13-15 after 30 minutes post-injury or later upon presentation for healthcare. These manifestations of mTBI must not be due to drugs, alcohol, medication, caused by other injuries or treatment for other injuries (e.g. systemic injuries, facial injuries or intubation), caused by other problems (e.g. psychological trauma, language barrier or coexisting medical conditions) or caused by penetrating craniocerebral injury (Carroll, Cassidy, Holm, et al., 2004, p. 115).

2.1.2 Prevalence

Previously, it was estimated that approximately 10 million people worldwide sustain a traumatic brain injury every year (Hyder, Wunderlich, Puvanachandra, Gururaj, & Kobusingye, 2007), however more recent global estimates indicate that the actual number is closer to 69 million of new cases of traumatic brain injury per year (Dewan et al., 2019). These discrepancies in prevalence rates have been attributed to the fact that only a small number of individuals with mild TBI (mTBI) receive a medical treatment or are admitted to hospital (Feigin et al., 2013), so consequently, a large proportion of injuries go undetected and unreported.

The majority of TBIs are sustained in road traffic accidents (60%) followed by falls (20-30%), violence (10%), and finally in workplace and sports-related accidents (10%) (Hyder et al., 2007). Feigin et al. (2013) examined TBI incidence and outcomes study in NZ and found that incidence rates in NZ are significantly higher than in other high-income countries, with 790 new cases of TBI per 100,000 people per year in comparison to 47-453/100,000 new cases in Europe, and 51-618/100,000 cases in North America. Furthermore, the authors found higher rates of TBI in children and young adults (under the age of 35 – almost 70%), a higher percentage of TBI due to falls with an estimate of 38%, and also a higher percentage of TBI due to assaults with an estimate of 17%, in comparison to the worldwide incidence rates of 40-60%, 13-28%, and 11-14% respectively (Feigin et al., 2013). Results of this research also showed higher rates of TBI in Maori people than in people of European origin, however reasons for this discrepancy are still unclear with one possible contributing factor of increased levels of interpersonal violence in this population (Feigin et al., 2013). Other studies have also reported racial and ethnic differences in the risk of sustaining TBI, for example, in the US, African American, American Indian, and Alaskan Native males have a 4 times higher death rate from TBI than white males (Langlois et al., 2003).

2.1.3 Sequelae of TBI

Following TBI, individuals may experience permanent or temporary impairment in cognitive, physical, behavioural, emotional, and psychosocial functioning (Freire et al., 2011). The most commonly reported difficulties are changes in cognition, which can affect many aspects of an individual's daily life, including independent living, family life, social functioning, and also vocational activities (McAllister, 2008). Persistent cognitive deficits are said to be more associated with penetrating, moderate, and severe TBIs (Dikmen et al., 2009). Symptoms may include difficulties in memory, attention, processing speed, speech and language and visual functions, learning, and also executive functioning including inhibition, self-monitoring, cognitive flexibility, planning, organising, abstract reasoning, and judgment (Fleminger, 2009; McGee, Alekseeva, Chernyshev, & Minagar, 2016; Riggio, 2011).

Changes in emotional and behavioural regulation are often described as "changes in personality" by survivors and their family members (McAllister, 2008). Individuals affected by TBI may experience difficulties with poor impulse control, evident by verbal utterances, physical actions, snap decisions, and poor judgment; irritability and aggression, evident by behaviours ranging from verbal outbursts to aggressive assaultive behaviour; and emotional lability, visible by exaggerated changes in mood and displays of emotional expression (McAllister, 2008; Riggio, 2011). In some cases these symptoms may be complicated by and also attributed to the emotional response to the injury, including stress of the injury, fear of disability, and due to acquired cognitive limitations (Riggio, 2011).

In addition, individuals may also experience somatic symptoms following TBI, such as sleep disturbance, fatigue, dizziness, vertigo, headaches, visual disturbances, nausea, sensitivity to light and sound, hearing loss, and seizures (Riggio, 2011).

Deficits following TBI can be complicated by lack of awareness or lack of insight into experienced difficulties (McAllister, 2008). Lack of insight can further affect the recovery process due associated lowered motivation to engage in rehabilitation (Alderman, 2003).

Psychiatric disorders are prevalent in the TBI population, including mood and anxiety disorders, substance abuse, and psychotic disorders (Fleminger, 2009; Lathif, Phipps, Alton, & Sharma, 2014). These may complicate the recovery and rehabilitation process (Kim et al., 2007). Research by Gould, Ponsford, Johnston, and Schönberger (2011) found that the majority of post-TBI psychiatric patients diagnosed represent the continuation of pre-injury psychiatric disorders, however, 45.8% of their sample developed a novel post-injury disorder. Furthermore, anxiety, depression, and apathy may also be manifestations of the frustration as a result of cognitive deficits and associated disability (Riggio, 2011).

Overall, an individual may experience a variety of temporary or permanent deficits following TBI, and according to Gallagher, McLeod, and McMillan (2019), every individual will have a unique set of difficulties.

2.2 Performance Validity Testing

2.2.1 Overview

Currently, validity testing research is considered to be one of the most dominant themes in the field of neuropsychology (Bigler, 2014; Inman & Berry, 2002). Performance validity testing has become increasingly important due to the growth of forensic evaluations, for example, in neuropsychological evaluations related to litigation, administrative proceedings, consultations to attorneys and courts, and disability determinations (Inman & Berry, 2002; Lezak et al., 2012). As a result, much of the initial and current research has focused on validity testing and malingering issues (Lezak et al., 2012).

One of the main objectives for neuropsychological assessment is to "identify the degree of preservation or impairment of an individual's capacities and to determine the real cognitive state of a particular function" (Herrera-Guzmán et al., 2004, p. 385).

Neuropsychologists are responsible for determining the validity of these assessments including the validity of the test performance and the information obtained during a neuropsychological evaluation (Bush et al., 2005). Furthermore, Bush et al. (2005) suggested that determinations about the validity should be the primary step when interpreting data from neuropsychological evaluation. Numerous studies have demonstrated that cognitive test results can be significantly affected by cognitive underperformance (i.e. not performing to one's true abilities) (Dandachi-FitzGerald et al., 2013). Identification of these invalid or non-credible test performances is essential, otherwise assessment may lead to an inappropriate diagnosis, incorrect identification of severity of experienced difficulties, psychological iatrogenesis, unjustified referrals, treatments, and accommodation, increased healthcare utilisation, and unnecessary spending of healthcare resources (Lippa, 2018; Olsen, Schroeder, Heinrichs, & Martin, 2019). Consequently, adequate validity assessments are crucial in order to increase confidence in obtained neuropsychological assessment results and the clinician's recommendations which have been based on these results (Bush et al., 2005).

2.2.2 Performance validity assessment

Neuropsychological assessment results can potentially be affected by symptom exaggeration, which has been defined as "the conscious or unconscious tendency of a person to under-rate their abilities and/or to overstate their limitations and symptoms" (Bathard, 2016, p. 5); or poor effort or performance, which has been defined as "the level of cognitive and behavioural engagement in a task, and is associated with test performance" (Bathard, 2016, p. 5).

Previously, the expression *symptom validity* had been applied to validity issues in general (Egeland, Andersson, Sundseth, & Schanke, 2015). However, Larrabee (2012) proposed using two descriptive terms when assessing the response bias or validity of a neuropsychological evaluation. It was suggested to use the term symptom validity tests (SVTs) for assessing the accuracy of symptomatic complaints on self-report measures, and the term performance validity tests (PVTs) for measures assessing the validity of a person's performance on an actual task (Larrabee, 2012). In other words, SVTs assess over-reporting of fabrication of symptoms within self-report measures and structured interviews, whereas PVTs assess underperformance on neuropsychological tests (Lippa, 2018). The distinction is made between modes of information collection as somatic symptoms and psychological complaints are typically obtained from self-reports, and cognitive functioning is assessed using performance tests (Egeland et al., 2015). This distinction is also supported by research on malingering, for example, Ruocco et al. (2008) who found dissociation between feigned neuropsychological deficits and exaggerated psychiatric symptoms. Furthermore, researchers reported that individuals who malingered on performance tests did not frequently exaggerate their psychological symptoms.

Prior to recent validity assessment research, clinical judgment was used to determine performance validity on neuropsychological tests, which was based on behavioural observations during testing such as a lack of engagement, or when the test performance did not fit with the clinical history or the presentation (Bigler, 2014). However, studies have showed poor levels of accuracy in detecting malingering, based on clinical judgment (Bianchini et al., 2001; Green et al., 2001). For example, a recent study by Dandachi-FitzGerald et al. (2017) explored the ability of 31 experienced neuropsychologists to predict non-credible symptom presentation of 203 hospital outpatients. Their predictions were matched with actual results from two validity measures. Findings showed that neuropsychologists were able to correctly identify 76% of non-credible cases, resulting in misclassification in 24% of cases. Importantly, findings demonstrated that neuropsychologists predicted problematic response validity for 51 patients based on their clinical judgment, however, the actual results of the assessment revealed that 35 of those patients passed both validity measures. Findings from Dandachi-FitzGerald et al. (2017) study demonstrate the importance for clinicians to use objective measures when assessing performance validity. Furthermore, a survey conducted by Dandachi-FitzGerald et al. (2013) found that the majority of clinicians in Western Europe continue to rely on their clinical judgment, despite demonstrating technical knowledge about performance validity. Similarly, a survey by Barker-Collo and Fernando (2015) found that clinical judgment and SVTs are the most commonly used validity assessments amongst NZ psychologists. According to Dandachi-FitzGerald et al. (2013), neuropsychologists tend to overestimate their ability to detect a feigned performance.

Van Gorp et al. (1999) investigated if the performance pattern on neuropsychological tests can be used as a reliable indicator to identify individuals who are not performing to the best of their ability. Results of the study supported the view that specialised performance validity tests are a crucial and necessary component of neuropsychological evaluation in order to identify individuals who are not performing to the best of their ability.

As a result, performance validity testing is said to bring a greater objectivity in a neuropsychological evaluation (Bigler, 2014), improving clinicians' ability to determine the credibility of reported impairments (Bathard, 2016).

2.2.3 Performance validity tests and effort

Slick and Sherman (2013) defined effort as "the amount of mental and/or physical energy expended in performing a task" (p.60). In the PVT literature the term 'effort' is associated with the effort to perform well on tests. Thus effort measures have been designed to assess the degree to which an individual uses effort to perform well on a neuropsychological test (Bush et al., 2005; Slick & Sherman, 2013). An individual who fails a PVT is said to demonstrate poor effort to do well on the task, and an individual passing the test shows good effort to perform well (Slick & Sherman, 2013). It is important to note that malingering "is defined by the goal to which effort is directed and not the level of effort expended" (Slick & Sherman, 2013, p.60), meaning that an individual may expend high levels of effort towards feigning cognitive impairment with a goal of obtaining material gain, but poor level of effort towards performing well on the neuropsychological test.

Good effort is required in order to obtain valid results on standard psychological tests (Stevens et al., 2008). This is due to the fact that individuals included in normative samples are thought to perform to the best of their ability, because their participation is voluntary and often compensated with a payment (Stevens et al., 2008); there is no material advantage to be gained by showing poor effort (Stevens et al., 2008). This, however, is not representative of clinical or forensic settings, where the test performance below published norms may be due to lack of effort, rather than actual cognitive impairment (Stevens et al., 2008). For example, Green et al. (2001) examined 30,736 individual test scores from neurological, psychiatric, and medical patients as well as from individuals with TBI, and found that 53% of variance in these scores was explained by effort. Furthermore, the study found that only 4% of variance was attributed to age and 11% of variance was explained by years of education. Researchers also found an association between a reduction in effort and decline in neuropsychological test scores, demonstrating the relationship between effort and test scores. In addition, another study conducted by Green (2007) found that in the sample of 1,307 patients with a variety of disorders, including mTBI, stroke, and multiple sclerosis, failure in PVTs accounted for more of a variance in the test scores than did the severity of the brain injury. Similarly, Stevens et al. (2008) found that the effect of effort on neuropsychological test results was larger than that of a substantial brain injury. These studies demonstrate the significant impact of effort on neuropsychological assessment test results.

2.2.4 Position statements

The importance of using PVTs during neuropsychological assessment is emphasised in the position paper by the National Academy of Neuropsychology, which states, "the assessment of symptom validity is an essential part of a neuropsychological evaluation. The clinician should be prepared to justify a decision not to assess symptom validity as part of a neuropsychological evaluation" (Bush et al., 2005, p. 421). Furthermore, in NZ, the Accident Compensation Corporation (ACC) guidelines for clinicians using psychometric tests state: "Please consider the validity of symptoms for any assessment where there is a potential benefit to be gained from the client malingering their symptom presentation. The assessor needs to provide comment regarding this as part of their assessment." (Accident Compensation Corporation, 2013, p. 6). In addition, guidelines, provided by the NZ Psychologists Board for the use of psychometric tests, also includes guidelines for the inclusion of PVTs (New Zealand Psychologists Board, 2015). This demonstrates the importance in developing effective effort measures, which clinicians could use as part of a neuropsychological evaluation.

2.3 Malingering

2.3.1 Definition and classification

Slick, Sherman, and Iverson (1999) provided a definition for malingering in the context of neuropsychological assessment:

> Malingering of Neurocognitive Dysfunction (MND) is the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility. Substantial material gain includes money, goods, or services of nontrivial value (e.g., financial compensation for personal injury). Formal duties are actions that people are legally obligated to perform (e.g., prison, military, or public service, or child support payments or other financial obligations). Formal responsibilities are those that involve accountability or liability in legal proceedings (e.g., competency to stand trial) (p. 552).

Authors also proposed diagnostic categories for MND with a set of criteria for each category. Slick and Sherman (2013), suggest this criteria is the most commonly used diagnostic standard in neuropsychological research, however, it was also noted that the use in clinical settings is unknown.

According to the fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-5) “the essential feature of malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs” (American Psychiatric Association, 2013, p.726). DSM-5 also advises that malingering should be strongly suspected in a medicolegal context, where an individual is referred by an attorney or when an individual self-refers while litigation or criminal charges are pending (American Psychiatric Association, 2013). Furthermore, if there is a marked discrepancy between reported stress or disability and the objective findings and observations, lack of cooperation during the diagnostic evaluation, and presence of antisocial personality disorder, malingering should be strongly suspected (American Psychiatric Association, 2013).

2.3.2 Prevalence

Estimated prevalence rates of symptom exaggeration depend on a referral type, setting, and diagnosis (Barker-Collo & Fernando, 2015). Research by Bianchini, Curtis, and Greve (2006) found that malingering and symptom exaggeration is particularly associated with financial gain through compensation seeking and litigation for disability in a TBI population. Authors also found that individuals with mild TBI (mTBI) were more likely to show low effort than those affected by moderate to severe injuries when financial incentive was present, demonstrating a dose-response relationship. Furthermore, the findings indicated that higher levels of non-credible test performances were associated with a potential for obtaining a larger sum of money from a legal procedure.

Numerous studies have demonstrated that for most of those affected by mTBI, post-concussive symptoms gradually resolve and the person returns to their baseline functioning within three to 12-months (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Carroll, Cassidy, Holm, et al., 2004). However, a significant number of people continue to experience post-concussive symptoms beyond this period including headaches, fatigue, cognitive difficulties and other symptoms that are not specific to mTBI (Cassidy et al., 2014). Iverson (2005) suggested that recovery from mTBI could potentially be affected by pre-existing psychiatric problems and substance abuse, poor general health, orthopedic injuries, and comorbid difficulties, such as chronic pain, depression, substance abuse, life stresses, unemployment, and involvement in litigation. As a result, clinicians should take caution when evaluating individuals who have suffered mTBI. A large systematic review found that involvement in litigation and compensation particularly is a strong predictor of persistent symptoms post mTBI (Carroll, Cassidy, Peloso, et al., 2004). Similarly, Chafetz (2008) investigated malingering base rates among social security disability claimants alleging low cognitive functioning. The sample consisted of 317 referrals from the Disability Determinations Service (DDS) in the United States. Results showed that, depending on the administered PVT, 45.8% and 59.7% of those in the compensation-seeking group met the criteria for at least probable malingering based on Slick et al. (1999) classification. Furthermore, the DDS budget in 2004 was $80.33 billion (Chafetz, Abrahams, & Kohlmaier, 2007), thus these studies suggest that potentially large amounts of funds are spent on possibly fraudulent claims of disability (Chafetz, 2008; Chafetz et al., 2007; Chafetzl & Underhill, 2013). Furthermore, Larrabee (2007) reported that more than 50% of people seeking social security disability meet the criteria for malingering, demonstrating a potential impact on the economy; feigning of cognitive deficits following mTBI in US Military veterans has been estimated to costs between $136-235 million per year (Denning & Shura, 2019).

A study conducted by Ardolf, Denney, and Houston (2007) explored base rates for malingering within criminal forensic neuropsychological settings. The sample consisted of 105 criminal defendants frequently referred for neuropsychological assessment. Results of the study showed that, based on the classification for MND by Slick et al. (1999), that 54.3% of individuals met the criteria for probable and definite MND, with 21.9% meeting the criteria for definite MND. Authors concluded that base rates for malingering are higher within criminal forensic settings than in civil forensic settings.

Mittenberg, Patton, Canyock, and Condit (2002) conducted a study with active members of the American Board of Clinical Neuropsychology with an aim to estimate base rates of malingering. Estimates were based on 33,531 cases, and authors found that 29% of personal injury, 30% of disability, 19% of criminal, and 8% of medical cases involved probable malingering and symptom exaggeration. Results also demonstrated that 39% of all mTBI claims were associated with probable malingering. Furthermore, 35% of fibromyalgia/chronic fatigue, 31% of chronic pain, 27% of neurotoxic, and 22% of electrical injury claims were also associated with malingering. Thus, research results show that the mTBI population has the highest number of individuals suspected of feigning cognitive impairment. Interestingly, authors also explored strategies of malingering, for example, analysis revealed that in 65% of all cases reported, the severity of cognitive impairment was inconsistent with the condition, in 57% of cases individuals scored below cut-off scores on PVTs, in 56% individuals demonstrated discrepancies between records, self-report, and observed behaviour, 45% showed implausible changes in test performances across repeated examinations and 38% on validity scales on objective personality tests. However, Young (2015) argues that these results may not represent the actual base rates due to the fact that clinicians responding to the survey were not provided with definitions of malingering or exaggeration, and also base rates provided are for a classification of probable malingering and not rates for definite malingering.

Bianchini et al. (2006) explored a relationship between potential monetary compensation and failure on PVTs and SVTs in TBI. The sample consisted of 332 patients who were divided into 3 groups based on their incentive to perform poorly on neuropsychological assessment: no incentive, limited incentive as provided by state law, and high incentive as provided by federal law. Research used five validated PVTs and SVTs, and results demonstrated that 18% of mTBI in the limited incentive group and 33% of those with mTBI in a high incentive group met the criteria for MND based on the classification by Slick et al. (1999). Furthermore, findings showed that mTBI as well as moderate-severe TBI groups had an increased failure rate when an incentive was present, however, individuals with mTBI were substantially more likely to fail specific measures than individuals with moderate-severe TBI. Overall, researchers concluded that monetary compensation associated with workers compensation claims is a significant motivator for exaggerated and malingered impairments following a brain injury. Consistently, Webb et al. (2012) found that compensation seeking was predictive of PVT failure in a large sample of 555 TBI patients. As a result, clinicians should be particularly cautious in cases where individuals are presenting with high incentive and have suffered mTBI (Bianchini et al., 2006).

Similarly, Stevens et al. (2008) found that in their sample of 233 patients referred from the German Workers' Compensation Board and from claimants in personal injury litigation, 44.8% demonstrated insufficient effort on a number of SVTs and PVTs.

2.3.3 Other variables

It is important to note that there may be other factors aside from intentional feigning, that may contribute to performance validity (Cottingham & Boone, 2014). For example, patients with somatoform disorders are said to be more likely to fail PVTs than those without (Cottingham & Boone, 2014). For example, Boone and Lu (1999) found that two thirds of their patients with cognitive complaints and evidence of somatisation/conversion personality configurations failed PVTs during neuropsychological evaluation. Similarly, N. Kim et al. (2010) found that in their PVT study, a quarter of the credible patients showing poor effort also had a diagnosis of somatoform disorder. It has been suggested that unconscious symptom fabrication, which has been associated with somatoform conditions, is similar to conscious symptom feigning, which has been associated with malingering (Boone, 2007a). Cottingham and Boone (2014) proposed to "conceptualize feigning along a continuum, running from intentional, "conscious" feigning on one end (malingering) and unintentional "unconscious" feigning at the other extreme (somatoform conditions)" (p. 383). Authors explained that at times it could be difficult to determine if an individual is engaging in response bias intentionally or unintentionally.

Research has shown that other psychological variables are associated with PVT failure. For example, Webb et al. (2012) found that self-reported mood disorder and displayed psychotic illness was predictive of effort test failure in a sample of TBI patients. Similar findings have also been reported by Dandachi-Fitzgerald, Ponds, Peters, and Merckelbach (2011), where results showed that in a non-litigant mixed psychiatric patient sample, almost 34% failed PVTs and SVTs. Gfeller and Roskos (2013) also found an association between poor effort and increased self-report of symptoms on neuropsychiatric measures in a sample of military veterans with TBI. By contrast, research by Schroeder and Marshall (2011) found that in the majority of cases, presence of a psychiatric disorder was not associated with failure on PVTs. Results of the study showed that 74% of patients with psychotic disorders and 78% patients with other psychiatric disorders did not fail any administered PVT. Furthermore, 93% of patients with psychotic disorders and 95% with diagnosis of another psychiatric disorder failed less than two PVTs. Due to these inconsistent findings, Webb et al. (2012) suggested that there is a "complex association" (p.1378) between psychological factors and failure on effort measures.

A number of studies have also investigated demographic variables and their association with failure on PVTs. For example, Webb et al. (2012) found that age was related to failure on effort measures when examined independently, however in a multivariable model, age did not appear to be a significant predictor. Likewise, Stevens et al. (2008) found no relationship between age and PVT failure. Webb et al. (2012) also examined ethnicity, and the results showed that apart from the status of being foreign-born, ethnicity itself was not a significant predictor of effort test failure. Similarly, Erdodi, Nussbaum, Sagar, Abeare, and Schwartz (2017) found that limited language proficiency was associated with high rates of false-positives on PVTs. The study revealed that those with English as their second language were more likely to fail PVTs with high verbal mediation but not PVTs with low verbal mediation. Consequently, the authors suggested that PVT failures should be interpreted with a caution for individuals with a different linguistic background. Furthermore, research by Stevens et al. (2008) and Webb et al. (2012) found that education was a significant predictor of effort, such that less years in education were associated with increased likelihood of PVT failure.

Studies have also explored the relationship between severity of brain injury and PVT failure. For example, Green et al. (2001) found that the lowest rates of effort test failure were among patients with the most severe cognitive impairment within their sample. Interestingly, results showed that individuals with mTBI scored significantly lower on PVTs than did individuals with severe brain injuries. Similar findings were also reported by Stevens et al. (2008) and Webb et al. (2012). Interestingly, Boone and Lu (2003) presented two case studies with patients in litigation who had sustained severe brain injuries. Both patients presented with non-credible performance evidenced by failing multiple PVTs and SVTs, inconsistencies in test performance across a number of evaluations, and also inconsistencies between test scores and activities of daily living noted through surveillance videotapes. Consequently, Boone and Lu (2003) suggested that PVTs should be administered to all patients who have a motive to feign a cognitive impairment and not just to those with mild and questionable brain injury.

2.4 Effort Measures

2.4.1 Overview

There are two types of performance validity tests: free-standing/standalone tests, which have been specifically designed to assess response bias, and embedded tests, which enable assessing response bias from a metric derived standard cognitive tests (Cottingham & Boone, 2014). Both types of measures use a cut-score approach to the test results, where a performance inside of a certain cut-score is determined as a “pass” or “valid,” and a performance outside the cut-score is considered as a “failure” and “invalid” test results (Bigler, 2014). Free-standing performance validity tests have been designed to have a low true difficulty level, but a high face difficulty level, consequently inviting noncredible examinees to perform poorly and motivated test-takers to perform well regardless of their deficits (Inman & Berry, 2002). For example, the Test of Memory Malingering (TOMM) is one of the most frequently used free-standing PVTs in NZ (Barker-Collo & Fernando, 2015). The TOMM consists of a presentation of 50 drawings of common objects followed by two trials of a forced-choice task (Chafetz, 2011). Failure on the task is determined by incorrect response on more than five items on the second trial (Chafetz, 2011). Commonly used embedded effort measures include the Reliable Digit Span (RDS), the Digit Span Age-corrected scaled score (ACSS), the forced-choice trial of the California Verbal Learning Test – II (CVLT-II), and Wisconsin Card Sorting Test (WCST) embedded measures including Failure to Maintain Set (FMS), Total Errors, Perseverative Responses, Nonperseverative Errors, Categories Completed, and Trials to First Category (Cottingham & Boone, 2014; Erdodi et al., 2018; Glassmire, Wood, Ta, Kinney, & Nitch, 2019; N. Kim et al., 2010; Whiteside, Caraher, Hahn-Ketter, Gaasedelen, & Basso, 2018).

2.4.2 Advantages and Limitations

A number of disadvantages have been identified when using free-standing PVTs, such as associated costs with additional administration time and test materials, for example, free-standing PVTs may add further 5-45 minutes to the neuropsychological evaluation time, (Inman & Berry, 2002; Slick et al., 1999). Furthermore, it has been suggested that some of the more sophisticated malingerers may escape detection by identifying the free-standing measures, particularly if being coached by their attorneys (Erdodi et al., 2018). Inman and Berry (2002) found that 75% of attorneys spend around 15-60 minutes preparing their clients for a neuropsychological evaluation. In addition, according to a research by Essig, Mittenberg, Petersen, Stauman, and Cooper (2001), almost 50% of attorneys considered informing their clients of PVTs as part of their duty. As a result, it has been said that more sophisticated PVTs need to be developed, which will also provide clinicians with greater variety of measures to assess performance validity (Wetter & Corrigan, 1995).

Embedded measures or embedded validity indicators (EVIs) are said to provide a practical alternative to free-standing PVTs (Whiteside et al., 2018). EVIs have several advantages including:

- EVIs provide information on cognitive functioning as well as performance validity, serving "double duty" (Erdodi et al., 2018).
- EVIs do not add additional time or expense to the assessment process (Inman & Berry, 2002).

- EVIs are less likely to be susceptible to coaching and education due to being designed as neurocognitive tests rather than as effort measures (N. Kim et al., 2010).
- EVIs allow for a continuous monitoring of effort throughout the neuropsychological evaluation (Erdodi & Lichtenstein, 2017).
- EVIs are flexible as they span across a number of sensory modalities, and allow for effort testing within a variety of cognitive domains, whereas free-standing effort measures tend to be based more on a forced choice recognition memory (Cottingham & Boone, 2014; Erdodi & Lichtenstein, 2017). This follows recommendations by Boone (2009) to assess effort across a variety of cognitive domains as individuals may select specific domains in which to feign or exaggerate their impairments, and not others.

Embedded PVTs are said to be “better suited to meet complex demands of changing patient characteristics, developmental issues, and population-specific factors” (Erdodi & Lichtenstein, 2017, p.1030). For example, previously it has been suggested that most individuals feigning cognitive deficits following mTBI tend to target verbal memory tasks on which to simulate their impairments (Greve, Ord, Curtis, Bianchini, & Brennan, 2007). However, Cottingham and Boone (2014) investigated 135 case studies exploring performance patterns of non-credible individuals on PVTs claiming mTBI. Consistent with Boone’s (2009) recommendations, results of this study showed that individuals also feign cognitive deficits within other cognitive domains, for example, attention impairment, language impairment, processing speed and sensory impairment. Authors demonstrated the importance of using multiple PVTs assessing performance validity within a variety of cognitive domains.

2.4.3 Definitions

The effectiveness of a specific diagnostic method can be statistically determined with reference to base rates that are associated with a particular condition (Mittenberg et al., 2002). The term ‘base rate’ is defined as the “prevalence of an event, such as a symptom, sign or disorder, within a given population” (McCaffrey, Pavlav, O’Bryant, & Labarge, 2003, p. 1). Base rates of a certain condition are required for the calculation of positive predictive value (PPV) and negative predictive value (NPV), which enable estimation of predictive accuracy of testing tools (O'Bryant & Lucas, 2006). PPV refers to the likelihood of having a specific condition given a positive finding on a testing instrument, and NPV refers to the likelihood of not having a specific condition given a negative outcome on the testing instrument (McCaffrey et al., 2003).

Furthermore, calculations of sensitivity and specificity are also required in order to estimate predictive accuracy of a testing tool (O'Bryant & Lucas, 2006). According to Zasler and Bender (2019), clinicians need to have a good understanding of sensitivity and specificity in order to be able to interpret performance validity test results. According to Cottingham and Boone (2014):

> Sensitivity refers to the ability of a measure to correctly classify individuals with a particular condition. In the case of PVTs, it reflects the percentage of non-credible individuals detected as non-credible (true positives). Specificity, on the other hand, refers to the measure’s ability to correctly identify patients who do not have the condition. In the case of PVTs, specificity indicates the percentage of credible individuals who are correctly classified as credible (p. 375).

Sensitivity and specificity are said to be interrelated in an inverse fashion, which means that by increasing sensitivity, specificity typically decreases (Cottingham & Boone, 2014). Low sensitivity means that the identified cut-score results in a large number of false negative errors, which means that some individuals with suboptimal effort go undetected (Cottingham & Boone, 2014). However, low specificity means that the cut-score results in large number of false positive errors, meaning that a credible individual may be classified as non-credible (Greve & Bianchini, 2004). Incorrect diagnosis of malingering can have significant financial, occupational and personal consequences on an individual (Greve & Bianchini, 2004). Due to the importance of protecting credible individuals from being misidentified, cut-scores are selected to maximise specificity (Greve & Bianchini, 2004). According to Greve and Bianchini (2004), the general rule is to set specificity at ≥ 90%, which means that only 10% of credible individuals are misidentified as non-credible. Cottingham and Boone (2014) suggested that a range of cut-scores should be identified, and associated specificity and sensitivity reported.

Following these guidelines, clinicians should use PVTs that have demonstrated high sensitivity as well as satisfactory specificity (Greve & Bianchini, 2004). According to the literature, a number of previously researched effort measures have achieved ≥ 70% sensitivity while maintaining ≥ 90 specificity, for example, Warrington Recognition Memory Test-Words (Cottingham & Boone, 2014), Digit Symbol Recognition (M. S.

Kim et al., 2010), Dot Counting (N. Kim et al., 2010) and others (Boone, Lu, & Herzberg, 2002). Embedded measures are considered to be less sensitive than free-standing tests (Cottingham & Boone, 2014). Miele, Gunner, Lynch, and McCaffrey (2012) suggested that it might be due to the fact that results from embedded measures result in a larger range of scores from credible participants, leading to the cut-scores needing to be set at a low level to increase specificity.

However, according to Inman and Berry (2002), acceptable sensitivity and specificity can be obtained (84% and 94%) if a "two-failure rule" is used. Many other studies have supported this finding, for example, Larrabee (2003) found that failure on two PVTs increases specificity to 95% or more. Similarly, research by Chafetz (2011) demonstrated an increase of specificity to 96%, 97% or 99% depending on the PVT combination. Additionally, research has also demonstrated that failure on three or more PVTs is essentially conclusive of invalid performance, for example, this was demonstrated in research by Chafetz (2011) and Victor, Boone, Serpa, Buehler, and Ziegler (2009). Importantly, this means that there is a very small chance of mislabeling an honest individual as non-credible when multiple PVTs are used during a neuropsychological evaluation (Chafetz, 2011).

2.4.4 Multiple effort measures

The use of a "two-failure rule" can also be seen in a current practice where neuropsychologists commonly conclude that the test results are invalid if an individual fails two or more PVTs (Martin, Schroeder, & Odland, 2015). Victor et al. (2009) suggested that one performance validity measure cannot capture all non-credible individuals due to the variety of feigning strategies used, leading to recommendations of using multiple PVTs during the assessment process. Their research also found that 41% of individuals in their credible group failed at least one effort measure, demonstrating that it is a common occurrence among real-world clinical patients. According to Cottingham and Boone (2014), using multiple PVTs increases the classification accuracy as failing multiple measures by chance is not common. Boone (2009) also suggested that effort can fluctuate, thus it should be continuously assessed throughout an assessment. For example, her study found that only 16.4% of 146 non-credible individuals failed all four PVTs, and 35.6% failed two or less out of four administered effort measures, demonstrating that negative response bias is not static and therefore needs a continuous assessment. Similarly, Larrabee (2012), proposed that an individual may perform well on some effort measures but not on others, therefore administering several different measures is crucial when assessing performance validity. Likewise, Meyers, Volbrecht, Axelrod, and Reinsch-Boothby (2011) suggested that effort measure failure “is not an all or none phenomenon” (p. 14), and that individuals are selective on their approach to malingering and what cognitive deficits they are trying to feign or exaggerate. However, clinicians should be cautious not to use multiple PVTs that are strongly correlated with each other, due to the fact, that in this situation failure on the second PVT does not add any additional information and should not increase examiners confidence of invalid performance (Rosenfeld, Sands, & Van Gorp, 2000).

Furthermore, Cottingham and Boone (2014) argued that multiple PVTs should be used during a neuropsychological evaluation because no single measure has 100% sensitivity and specificity. As a result, definite conclusions about the validity of the performance cannot be made. For example, results of a single PVT failure cannot be used as evidence for non-credible performance due to cut-offs being set to allow for a small percentage of credible individuals to fail a PVT, unless the specificity has been set at 100% (Cottingham & Boone, 2014). Equally, it cannot be concluded that an individual is performing true to their abilities when a passing score is obtained on a single measure, as that individual may represent the small percentage of non-credible individuals who pass the PVT (i.e. false negative finding) (Cottingham & Boone, 2014). Researchers demonstrated a number of case studies, where specificity increased with a use of additional PVTs, for example, in one case specificity increased from 59% when using one measure to 100% when failure occurred on six PVTs. In another case the use of a single free-standing PVT was associated with a 41% rate of false-positive identification, but the specificity increased to 100% with failures on four PVTs (Cottingham & Boone, 2014).

More recent research has investigated the potential efficacy of logistic regression derived PVTs in which several individual embedded effort measure scores have been combined (Zasler & Bender, 2019). It has been suggested that using logistically derived PVTs will minimise the possibility of examinees avoiding detection of poor effort, as these measures utilise algorithms that are based on multiple cut-scores from several individual measures (Whiteside et al., 2018; Whiteside, Gaasedelen, et al., 2015; Whiteside, Kogan, et al., 2015).

For example, Whiteside et al. (2018) investigated classification accuracy of individual and combined executive functioning embedded validity measures. The study examined scores from the Trail Making Test B, the Wisconsin Card Sorting Test, and the Stroop

Color Word Test to assess whether individuals with mTBI demonstrated credible performance. Results demonstrated that the logistic regression method achieved similar overall classification accuracy to the best performing individual PVT, which was the embedded measure from the Stroop Color Word Test, and higher sensitivity than any individual measure. When specificity was selected at 90% the Stroop Color Word Test embedded measure individually achieved sensitivity of 43%, however sensitivity for logistically derived PVT was 57%. These results demonstrated that a combined PVT performed better at identifying true positive PVT failures without elevating the false positive rate.

Similarly, Whiteside et al. (2018) investigated language processing measures individually and as a logistically derived combined embedded PVT. Researchers examined scores from the Boston Naming Test and the Verbal Fluency Test, and results showed that when specificity was set at 90%, the individual measures had unacceptable classification accuracy and low sensitivity. However, when individual measures were combined in logistic regression derived PVT, it achieved acceptable classification accuracy although sensitivity remained low.

Furthermore, Whiteside, Kogan, et al. (2015) explored the effectiveness of logistically derived PVTs using measures from multiple cognitive domains, namely: attention, language, and verbal memory. Researchers used scores from California Verbal Learning Test – II (CVLT-II), Brief Test of Attention (BTA), Boston Naming Test – II (BNT-II), and Animal Fluency (AF) obtained from a sample of TBI patients. Results showed that none of the individual measures achieved excellent classification accuracy, and sensitivity ranged from 15% for the BNT to 48% for the CVLT-II when specificity was set at 90%. However, the logistically derived PVT achieved sensitivity of 54% for the same level of specificity. These studies have demonstrated the benefits of logistic regression derived PVTs gaining increased sensitivity while still maintaining specificity at 90% when compared to individual embedded measures.

2.5 The WAIS III Digit Symbol-Coding and the WAIS IV Coding subtests

2.5.1 Overview

The Wechsler Adult Intelligence Scale (WAIS) is considered as a gold standard in assessing intellectual functioning (Whiteside, Gaasedelen, et al., 2015) and is a core component of most neuropsychological evaluations (Erdodi, Abeare, et al., 2017; N. Kim et al., 2010). As a result, researchers have been investigating various WAIS subtests as potential embedded effort measures with good outcomes (N. Kim et al., 2010). A number of subtests are already being used in an assessment of effort, for example, the Reliable Digit Span, the Digit Span, and the Vocabulary minus Digit Span, (Glassmire et al., 2019).

2.5.2 Processing Speed Index

The Processing Speed Index (PSI) "reflects the mental and motor speed with which a person can solve nonverbal problems" (Groth-Marnat & Wright, 2016, p.190). PSI also assesses an individual's planning and organisation skills, motor control, and coordination of visual and motor abilities (Groth-Marnat & Wright, 2016). PSI has also been related to motivation, and is said to be the most sensitive index to cognitive difficulties associated with TBI, dementia, ADHD, and learning disabilities (Groth-Marnat & Wright, 2016).

PSI on the WAIS III and the WAIS IV consists of two subtests: Symbol Search and (Digit-Symbol) Coding (Erdodi, Lichtenstein, et al., 2017). For both subtests examinees are required to process and respond to visually presented information as quickly as possible, thus scores are strongly influenced by an individual's speed of cognitive processing (Glassmire et al., 2019).

2.5.3 Design changes between the WAIS III and the WAIS IV

A large proportion of studies investigating embedded effort measures have used earlier versions of the WAIS such as the WAIS-R and the WAIS III (Erdodi, Lichtenstein, et al., 2017). However, there have been significant changes in the most recent version of the WAIS (WAIS IV) and some of the subtests have been redesigned and re-normed (Erdodi, Lichtenstein, et al., 2017). As a result, this may lead to inaccurate findings if previously validated cut-off scores and algorithms are used (Erdodi, Lichtenstein, et al., 2017; Loring & Bauer, 2010).

According to the WAIS IV technical and interpretive manual, there have been a number of changes made to the Coding subtest from the previous version (Wechsler, Psychological, & PsychCorp, 2008). For example, the technical and interpretive manual by Wechsler et al. (2008) states that in order to allow for additional practice and ensuring that an examinee has been exposed to all nine number-symbol pairings before the actual test, the number of sample items has been increased from four to six. Two of the symbols have been kept from the WAIS III, however, in the WAIS IV those have been paired with different numbers on the key. Furthermore, more complex symbols have been replaced by new symbols. Due to a recommendation of equal item difficulty for processing speed measures, the design has been changed in order that each number appears twice within each row, leading to an increase in the total number of test items from 133 to 135. Finally, in order to reduce visual acuity and fine motor demands, the symbols, numbers, and the boxes for recording responses have been increased in size.

2.5.4 The Coding subtest and performance validity research

Coding is one of the core Processing Speed subtests, where an examinee is asked to copy symbols that have been paired with numbers using a key within a specified time limit (Wechsler et al., 2008). It has to be noted, that the Coding is known to be very sensitive to any type of organic and functional impairment, thus individuals with TBI and mood disorders such as depression tend to obtain lower scores (Groth-Marnat & Wright, 2016). Furthermore, according to Groth-Marnat and Wright (2016), the Coding score is likely to be the lowest out of all the subtests, even for individuals with minimal brain damage. Therefore, caution should be taken when the Coding scores are analysed for the purpose of assessing effort.

The first study exploring the Coding subtest as an embedded validity indicator (EVI) was conducted by Trueblood (1994). The research found that the WAIS-R Digit Symbol age-corrected scaled score (ACSS) of < 5 achieved 100% specificity but only 33% of sensitivity in a sample of credible and non-credible mTBI patients (N. Kim et al., 2010) . Interestingly, Trueblood (1994) also reported that non-credible examinees had a tendency to complete the Digit Symbol subtest with multiple errors, particularly, reversals and substitutions were noted. However, N. Kim et al. (2010) argued that the study has limited generalisability because the control group included mTBI patients, and literature shows that the majority of individuals with mTBI make a full recovery, and therefore cognitive functioning of this group was probably similar to normal population.

Inman and Berry (2002) examined a number of tests including the WAIS-R Digit Symbol with an aim to cross-validate previous research findings. The authors argued that cross-validation is particularly important with embedded measures, due to the fact that scores on these tests are more variable than on free-standing effort measures. Their sample consisted of individuals with mTBI, who were asked to exaggerate their impairment on the neuropsychological assessment, mTBI controls, college student controls without head injury, and college students without head injury who were also asked to feign cognitive impairment. In regards to the Digit Symbol subtest, the study cross-validated a scaled score cut-score of < 5, as previously suggested by Trueblood (1994). Results demonstrated that none of the participants in the control groups scored in a range indicative of low effort on the Digit Symbol subtest. Furthermore, the study showed that both groups, which were asked to feign their cognitive impairment, obtained significantly lower scores than both honest responding groups. Furthermore, the authors of the research found that the cut-score of < 5 for the Digit Symbol subtest achieved excellent specificity of 100%, which is consistent with the findings by Trueblood (1994), but an unacceptable level of sensitivity of 2%.

Etherton, Bianchini, Heinly, and Greve (2007) investigated the WAIS III PSI subtests as potential EVIs in the context of chronic pain, due to the fact that cognitive difficulties have been associated with chronic pain. Thus, the aim of the study was to explore the PSI ability to detect response bias in pain patients (Etherton et al., 2007). The sample consisted of 82 undergraduate students assigned to a variety of study conditions. This included the following: a control group; a simulator group, with participants simulating pain related memory difficulties while completing the PSI subtests; and a cold pain group, with participants receiving a cold-pressor procedure during the PSI subtests. Also included was a procedural distraction condition, where the procedure was the same as in the cold pain group, but in this case, the cold water was replaced with warm water. Results of the study showed that the simulator group scored significantly lower than any other group on both PSI measures. Findings also revealed that, when compared to the control group and the procedural distraction group, the cold pain group obtained significantly lower scores on the Digit Symbol subtest, but not on the Symbol Search or the PSI. Researchers concluded that performance on the Digit Symbol test might be affected by pain as an isolated variable. In addition, Etherton et al. (2007) conducted another study examining the PSI performance of non-credible chronic pain patients, credible chronic pain patients, and credible neurological patients, including individuals with moderate-severe brain injuries and individuals with memory disorders. Credibility of the participants was determined based on Bianchini, Greve, and Glynn's (2005) criteria for MPRD. TBI patients included in the study showed no evidence of poor effort. Results of this study showed, that the non-credible pain group participants obtained significantly lower scores than the other three groups on both PSI subtests. There were no significant differences in performance amongst other groups on the Digit Symbol subtest. According to the researchers, the Digit Symbol subtest achieved 81% sensitivity rate for credible pain patients, specificity of 87% for non-credible chronic pain patients, and inadequate specificity of 62% for TBI patients for a cut- score of < 5. Overall, findings revealed that the performance of the non-credible group on the PSI measures was similar to the performance of the simulator group in the first study. Thus, Etherton et al. (2007) concluded that Digit Symbol, Symbol Search subtests, and an overall PSI score may be used to detect feigned cognitive deficits in chronic pain patients. However caution should be taken as some credible pain patients showed impairment in cognitive processing speed at a level that is seen in TBI patients (Etherton et al., 2007).

Chafetz et al. (2007) conducted a study with an aim to develop a Malingering Rating Scale for the Psychological Consultative Examination (PCE), which provides evidence to the Disability Determinations Service (DDS) in the United States. Referrals often include individuals with low intellectual functioning. For example, 76% of referrals had received assistance from Special Education services, and 85.8% had not completed 12th grade education (Chafetz et al., 2007). The authors explained that this sample allowed for an opportunity to investigate effort in low functioning adults and children. The sample consisted of 317 DDS referrals claiming low cognitive functioning. Participants credibility was determined using Slick et al. (1999) diagnostic criteria for MND. Items for the Malingering Rating Scale were chosen based on the first author's observations during neuropsychological examinations in cases where an individual scored strikingly low on the TOMM (Chafetz et al., 2007). For example, items for examination included: simple arithmetic and sequencing tasks, missed personal information such as age, Ganser-like incorrect answers (i.e. consistently responding close to the correct answer), highly improbable answers, and claiming improbable pathology. Items from the WAIS included: a number of subtests with missed items before the start, if only one or no subtests achieved a scaled score of more than five, the Reliable Digit Span score of less than six, the Vocabulary or the Picture Arrangement score higher by three or more points than the Digit Span, and, finally, if two or more errors were made on the Coding subtest. Interestingly, Chafetz et al. (2007) noted that if the Coding errors included both horizontal and vertical reversals, then it could be considered to be a strong indicator of poor effort. This is due to the fact that deficits in perceptual functioning do not occur both ways (Chafetz et al., 2007). Overall, authors found that individuals with IQ less than 70 were still able to demonstrate good effort achieving perfect or near perfect scores on various PVTs. In terms of the Malingering Rating Scale, results showed that Ganser-like answers, variables of missed items before the start, and Coding errors were the best predictors of effort. In regards to the Coding errors, results of the study revealed that making more than two errors achieved sensitivity of 56% when specificity was set at 89%. When compared to other measures derived from the Coding subtest, these findings propose that the Coding error rate may have the highest sensitivity as a predictor of low effort.

Chafetz (2008) conducted a further validation study for the DDS Malingering Rating Scale using the same sample. Items for further investigation included: simple arithmetic and sequences, Ganser-like incorrect responses, a number of subtests where items were missed before the start, and Coding errors. Results demonstrated that overall the rating scale accounted for 65% of variance and the highest scoring individual variable, accounting for 51% of variance, was the number of subtests in which items were missed before the start. In regards to the Coding subtest, errors accounted for 33% of variance in the adult sample and 54% in the child sample. This study also confirmed previous findings, that failure of effort measures is a result of poor effort and not IQ *per se* in this low functioning population (Chafetz, 2008).

Chafetz (2011) investigated if using multiple items from the previously validated DDS Malingering Rating Scale reduces the rate of false-positives within the DDS claimant population. The author chose three out of eleven items from the DDS Malingering Rating Scale including sequence errors, Ganser-like errors, and Coding errors. This was due to the fact that in Chafetz et al. (2007) these items demonstrated moderate to strong correlations with other previously validated PVTs (Chafetz, 2011). The same sample of DDS referrals from the previous two studies was used. Results of the study indicated that with a failure on all three effort measures, there is a very small chance of misclassifying a credible individual as malingering. In terms of the Coding subtest, the study found that a cut-off score of ≥ 2 errors achieved specificity of 97% and associated sensitivity of 45%, and when the cut-off score was lowered to > 0 errors, then sensitivity increased to 58% with specificity at 92%. These results appear to be consistent with the findings of Chafetz et al. (2007).

Furthermore, Kim and colleagues examined the efficacy of the WAIS III Digit Symbol-Coding subtest in identifying response bias in credible and non-credible neuropsychology clinic patients (N. Kim et al., 2010). More specifically, researchers investigated if adding an additional timed recognition trial following the standard administration could increase the effectiveness of the Digit Symbol-Coding subtest as an effort measure. The sample consisted of 171 participants; their credibility was determined based on the criteria for

MND created by Slick et al. (1999). The additional recognition trial consisted of four multiple-choice options for each number in the Digit Symbol subtest (N. Kim et al., 2010). Furthermore, as part of the recognition trial, authors of the research also included 180° rotation of the correct symbol with an aim to improve the recognition trial sensitivity. According to Kim and colleagues, this was based on a previously reported proposal by Binder (1992) that feigning may be associated with rotational errors on the Digit Symbol-Coding subtest (N. Kim et al., 2010). Thus, the authors examined the Digit Symbol-Coding ACSS, the Digit Symbol raw score, and three variables from the recognition task including correct responses, time, and rotation errors. Results of this study showed that the non-credible group performed significantly worse than the credible group on all Digit Symbol variables. Interestingly, results also showed that the recognition task scores were more sensitive than standard Digit Symbol-Coding scores in detecting poor effort. For example, when specificity was set at ≥ 89%, sensitivity for the ACSS was 18% and the cut-off for the raw score was associated with 40% sensitivity. However, recognition task total for correctly identified symbols cut-score was associated with 58.5% sensitivity, and 49.2% sensitivity for the recognition trial time score. Based on the findings N. Kim et al. (2010) developed an equation, which included ACSS and variables from the recognition trial, and this score was associated with 79.9% sensitivity and 88.7% specificity. With regards to the rotational errors, results showed that non-credible participants were significantly more likely to select these incorrect responses, however, the rotational errors alone were associated with only 32.9% sensitivity at ≥ 90% specificity. N. Kim et al. (2010) concluded that the additional recognition trial has the potential of being an effective effort measure, but the standard scores on the Digit Symbol subtest (meaning the raw score and the ACSS), are inadequate in determining poor effort on neuropsychological tests. Researchers did mention that the recognition trial task has been specifically developed for the WAIS III Digit Symbol-Coding subtest, and a new version will have to be developed for use with the WAIS IV Coding. It is important to note that the sample in this study included a wide range of clinical patients who were referred to a clinical neuropsychology clinic, therefore sensitivity and specificity rates may not be applicable to other populations such as in the homogeneous TBI sample.

More recently, Erdodi, Lichtenstein, et al. (2017) extended the previous research by exploring the WAIS IV processing speed scores as a potential measure of response bias. The authors examined a number of possibilities including the PSI score, the Coding and the Symbol Search scaled scores, the Coding minus the Symbol Search scaled score difference, and finally, a 2-tailed cut-off on the Coding/Symbol Search raw score ratio. The sample consisted of 205 medical patients referred for a neuropsychological assessment, including patients with psychiatric diagnoses, TBI, epilepsy, stroke, amnesic MCI, cancer, multiple sclerosis, hydrocephalus, and hepatic encephalopathy. Participants' credibility was determined based on a number of failed PVTs, where performance was deemed as invalid with a failure on two or more measures, and a valid performance was considered if there was none or one effort test failure. Results of the study demonstrated that overall PSI with a cut-score of ≤ 79 achieved sensitivity ranging from 23% to 56% with specificity ranging from 92% to 98%. The Coding scaled score of ≤ 5 achieved low and variable sensitivity ranging from 4% to 28% with associated specificity ranging from 94% to 100%, which is consistent with previous research. Findings also indicated that the scaled score difference between the two subtests as well as the raw score ratio also produced low sensitivity ranging between 8% and 12% and between 15% and 24% respectively. The authors of the research concluded that processing speed measures have the potential as PVTs, however, due to their low and variable sensitivity, they should be used in a combination with other PVTs to determine performance credibility. Erdodi, Lichtenstein, et al. (2017) and Glassmire et al. (2019) reported that caution should be taken when assessing patients with moderate and severe brain injuries as processing speed based EVIs produce unacceptable rates of false positive errors.

Glassmire et al. (2019) conducted a study to examine if the findings of Erdodi, Lichtenstein, et al. (2017) are generalisable to forensic psychiatric patients with schizophrenia spectrum disorders (SSDs). The authors indicated that within forensic and disability contexts individuals with SSDs are frequently evaluated, and that within this context there is incentive to exaggerate cognitive impairment. Furthermore, the literature has demonstrated that individuals with SSD display poor ability to sustain task effort, and are also associated with a variety of neurocognitive difficulties including processing speed, and this may have an effect on PVTs (Glassmire et al., 2019). Consequently, the concern is that the effort performance of individuals with SSDs may be classified as invalid due to the symptoms associated with SSD and not due to poor effort leading to high rates of false positives (Glassmire et al., 2019). In order to assess the false positive rates of PSI-based embedded validity indicators, the study included only those individuals with valid efforts based on PVT performance or clinician opinion. In regards to the Coding subtest, results of the study indicated that 49% of SSD patients would have been incorrectly classified as showing poor effort for a ACSS cut-off score of ≤ 5, thus demonstrating high false positive rates. However, study findings also indicated that when other PSI measures were used the results were more promising. For example, the Coding minus the Symbol Search scaled score difference produced 8% false positive rates, and a 2-tailed cut-off on the Coding/Symbol Search raw score ratio was associated with 2% false positive rates. Authors of the research concluded that effort measures based on performance between 2 subtests are more appropriate for the SSD patient population, than those calculated based on performance of individual measures.

In a different study, Erdodi and Lichtenstein (2017) investigated eight WAIS III and WAIS IV subtests. The sample consisted of 312 mixed clinical patients referred for a neuropsychological assessment. Results of the study demonstrated that all subtests were significant predictors of performance validity, however, their sensitivity levels were largely varied. Furthermore, it was revealed that the Coding, the Digit Span, and the Symbol Search subtests achieved the best classification accuracy when compared with two other PVT measures used in the study. In regards to the Coding subtest, ACSS cut-off score ≤ 5 achieved specificity ranging from 89% to 91% with associated sensitivity ranging from 24% to 29%. When the cut-off score was lowered to ≤ 4, then specificity improved to 94%, but it was at a cost to sensitivity, which ranged from 19% to 23%.

2.6 Conclusion

Validity testing research is considered as one of the most dominant themes in the field of modern neuropsychology (Bigler, 2014; Inman & Berry, 2002), and due to this a large amount of research is available. Currently, there is a growing interest in the area of embedded validity indicators due to identified advantages over free-standing PVTs. A large proportion of most recent research has been dedicated to developing and validating new embedded measures, and evidence is emerging that they could potentially be valuable adjunctive measures of effort alongside the free-standing tests.

Some encouraging evidence exists that the WAIS Coding subtests may have a potential as EVIs during neuropsychological testing. Previously the Coding ACSS has been somewhat explored, but to-date only three studies (Chafetz, 2008, 2011; Chafetz et al., 2007) have investigated Coding Errors as a potential effort measure. All three studies showed promising results, however Coding Errors was investigated as part of a larger Malingering Rating Scale using the same sample of individuals claiming intellectual disability. As a result, further validation is necessary by employing different participant samples. Based on previous findings, it is thought that the Coding subtest, particularly, the Coding Error rate has a potential as an effective embedded performance validity test in a population of people affected by traumatic brain injury. Furthermore, the literature has demonstrated that a combination of a number of embedded measures achieves better ability to detect individuals with low effort, and based on these findings it is believed that a combination score of Coding embedded effort variables may also improve its ability to detect individuals with TBI with low effort.

It is hypothesised that:

1. The WAIS Coding Error rates are predictive of low effort.
2. The WAIS Coding ACSS is predictive of low effort.
3. A newly-devised WAIS Coding Combination score (i.e.an arithmetic sum of Coding Errors and the Coding ACSS) is predictive of low effort.
4. The three Coding embedded validity indicators will bias against those who have suffered more significant brain injuries.

Chapter 3 Methodology

3.1 Design

The current study is a quantitative research with a known-groups design, using PVT outcomes obtained during a neuropsychological assessment to determine participants' group membership.

3.2 Ethics Approval

Ethical Approval was obtained from the Auckland University of Technology Ethics Committee (AUTEC) on 7th May 2019 with the approval number 19/107 (see Appendix A).

3.3 Participants

This research used de-identified archival data, provided by Webb Psychology Ltd, a private clinical neuropsychology practice in Auckland, NZ. All participants had been referred to the private practice for the purpose of a neuropsychological assessment following TBI. Data was collected over the period from 1999-2014 inclusive, representing consecutive referrals. The total sample included 650 participants. All participants were over the age of 16 at the time of their assessment. All participants had signed a consent form (see Appendix B) prior to their neuropsychological evaluation, agreeing for their de-identified health information to be used for research purposes. Participants were excluded on the basis of having a pre-existing history of mental retardation or dementia.

3.3.1 Compensation-seeking

Most of the sample (n = 547; 84.2%) were seeking compensation continuance or seeking entitlement to compensation. Compensation was defined as worker's compensation income replacement payments from the NZ state-insurance system (Accident Compensation Corporation) or disability social security benefits. None of the people in the sample were engaged in litigation (litigation for damages is specifically precluded under New Zealand's no-fault accident compensation legislation). A sizeable minority (n = 102; 15.7%) of the total sample were ineligible for compensation (see Table 1). Information was not available for 1 participant (n = 1; 0.1%).

3.3.2 Demographic variables

Men made up 70.3% of the sample and women 29.7% respectively. The mean age was 42.62 years (*SD* = 12.26, range = 16-76). Ethnicity was as follows: European/White (*n* = 499, 76.8%), Maori/Pacific Island (*n* = 99, 15.2%), and Indo/Asian (*n* = 52, 8.0%). The sample had, on average, 11.84 (*SD* = 2.51, range = 2-21) years of education (see Table 1).

Table 1: *Sample demographic characteristics and compensation-seeking status*

Variable	***n***	**%**
Gender		
Male	457	70.3
Female	193	29.7
Ethnicity		
European/White	499	76.8
Maori/Pacific Island	99	15.2
Indo/Asian	52	8.0
Compensation-seeking		
Yes	547	84.2
No	102	15.7
	M	***SD***
Age	42.62	12.26
Education, years	11.84	2.51

Note: Indo/Asian = participants reporting themselves of Indian or Asian ethnicity

3.3.3 Injury variables

Injury severity was determined based on GCS score at the Emergency Department, duration of PTA assessed by the Westmead Post-traumatic Amnesia scale (Shores, Marosszeky, Sandanam, & Batchelor, 1986), and duration of loss of consciousness (LOC). LOC was assessed in accordance with the guidelines of Ruff et al. (2009). Specifically, that the duration of LOC should result from an impact, not other medical causes, and that LOC was determined from collateral reports of witnesses present at the scene (e.g., paramedic) or from hospital medical records, not from self-reporting by the participant.

Severity of mild TBI (mTBI) was defined according to the WHO Collaborating Task Force mTBI diagnostic criteria (Carroll, Cassidy, Holm, et al., 2004). Complicated mTBI was differentiated from mTBI according to Williams, Levin, and Eisenberg (1990). Moderate to severe TBIs were defined using the Teasdale and Jennett (1974) and Teasdale (1995) criteria.

In cases where the PTA information was not available, the onset of continuous memory was established during a clinical interview based on the research of Gronwall and Wrightson (1980). Classification categories consisted of mTBI, mTBI-complicated, moderate, severe, very severe, and extremely severe brain injury (see Table 2).

Table 2: *Injury severity criteria and sample characteristics*

Descriptor	**Criteria**
mTBI	LOC < 30 mins, PTA < 24 hours, GCS 13 – 15 at 30 mins
mTBI-complicated	LOC < 30 mins, PTA < 24 hours, GCS 13 – 15 at 30 mins, visible (on CT brain imaging) intracranial abnormality not requiring surgery
Moderate	PTA 1 – 24 hours, GCS 9 – 12
Severe	PTA 1 – 7 days, GCS 3 – 8
Very severe	PTA 1 – 4 weeks, GCS 3 – 8
Extremely severe	PTA > 4 weeks, GCS 3 – 8

Note: LOC = Loss of consciousness; mTBI = mild traumatic brain injury; PTA = post-traumatic amnesia; GCS = Glasgow Coma Scale.

For the analysis, the TBI severity variable categories included mTBI, Moderate (a combination of mTBI-complicated and moderate brain injury), Severe, and Very Severe (a combination of very severe and extremely severe brain injuries) (see **Error! Reference source not found.**). This was done to ensure that parameter estimates were based on adequate numbers of cases and that there were no empty cells. This is supported by Kashluba, Hanks, Casey, and Millis' (2008) findings that there are few differences in outcomes between mTBI-complicated and moderate TBI injuries. Furthermore, the findings of Webb et al. (2012), indicate that very severe TBI and extremely severe TBI did not differ in their predictive relationship with effort test failure.

Table 3: *Sample injury severity characteristics*

Injury Severity	***n***	**%**
mTBI	360	55.4
Moderate	92	14.2
Severe	87	13.4
Very Severe	111	17.1

Note. mTBI – mild traumatic brain injury

3.4 Performance Validity Measures

3.4.1 Rey Fifteen Item Test

The Rey Fifteen Item Test (FIT) (Rey, 1964) is one of the most commonly used free-standing measures of effort during a neuropsychological evaluation (Barker-Collo & Fernando, 2015; Love, Glassmire, Zanolini, & Wolf, 2014). The FIT takes about five minutes to administer and is comprised of 15 items consisting of numerical digits, letters of the alphabet, lines, and simple shapes (Morse, Douglas-Newman, Mandel, & Swirsky-Sacchetti, 2013). The items are arranged in three columns and five rows, and are presented to the examinee on a single page for 10 seconds (Morse et al., 2013; Strauss, Sherman, & Spreen, 2006). Following the presentation of the stimulus card, examinees are asked to draw the items from memory in any order (Strauss et al., 2006). The number "15" and the word "different" are emphasised during the instructions with an aim for the task to appear more difficult to the examinee (Lezak et al., 2012). In reality, the test is fairly easy and examinees are only required to recall three or four ideas in order to be able to recall most of the items (Lezak et al., 2012; Strauss et al., 2006). This is due to the item redundancy, for example, the letter items consisting of ABC and abc, the number items consisting of 123, and the line items consisting of I II III (Strauss et al., 2006). The total score is the number of correctly recalled items with a maximum score of 15 (Bailey, Soble, & O'Rourke, 2018). In the literature, the recommended cut-off score is nine, thus non-credible performance is suggested if an examinee obtains a score of less than nine (Lezak et al., 2012).

3.4.2 Test of Memory Malingering

The TOMM (Tombaugh, 1996) is a free-standing PVT and has been developed for the identification of suboptimal effort during a neuropsychological evaluation, and is said to be the most studied and validated test of memory malingering (Morse et al., 2013). It is also one of the most commonly used tests to assess effort in clinical practice (Barker-Collo & Fernando, 2015). The TOMM takes about 15 minutes to administer and consists of two learning trials followed by a forced-choice recognition task and an optional delayed Retention Trial (Lezak et al., 2012; Strauss et al., 2006). For each learning trial, the examinee is presented with a series of 50 pictorial stimuli of common objects (target items), for three seconds each with a one second interval (Strauss et al., 2006). The forced-choice task contains 50 picture pairs with one target item and one new object, and examinees are asked to identify the previously seen target pictures (Lezak et al., 2012;

Strauss et al., 2006). For both learning trials the same pictorial stimuli are used; however, they are presented in a different order, and explicit feedback is provided to the examinee on the response correctness (Strauss et al., 2006). The optional Retention Trial can be administered 15 minutes after Trial 2, and contains another 50 forced-choice pairs; however, for this task the target pictures are not re-administered (Strauss et al., 2006). According to Tombaugh (1996) the two learning trials are usually sufficient in identifying non-credible performance, but the delayed retention trial helps to verify results.

Tombaugh (1996) demonstrated that a cut-off score of 45 on Trial 2 or the Retention Trial is associated with sensitivity of 95% and specificity of > 90%, thus the score of less than 45 indicates that the examinee is not putting forth good effort. However, Greve, Bianchini, and Doane (2006) found that the original cut-score of 45 may be too conservative when assessing individuals with TBI, particularly, for those with mTBI. Research demonstrated that a cut-off score of < 47 achieved 49% sensitivity and 95% specificity on the Trial 2, and 61% sensitivity with the same 95% specificity on the Retention Trial. Based on these findings, authors proposed a cut-off score of < 47 to be used to detect low effort in individuals with TBI.

3.4.3 Word Memory Test

The WMT (Green, Allen , & Astner, 1996) is a commonly used free-standing PVT consisting of a number of tasks that have been designed to assess effort by measuring examinees ability to recognise and recall the word pairs (Bailey et al., 2018). The test includes three primary effort measures: immediate recognition (IR), delayed recognition (DR), and consistency (CNS), and also two memory tasks: memory choice (MC), paired associate (PA) recall task, and, finally, a more difficult memory subtest: free recall (FR) (Bailey et al., 2018).

The WMT takes about 20 minutes to administer, and includes a list of 20 semantically linked word pairs, for example, dog-cat (Strauss et al., 2006). The word pair list is presented twice and can be read by the examiner or shown on a computer screen (Strauss et al., 2006). This is followed by the IR forced-choice task where an examinee is presented with 40 new pairs consisting of one word from the original list and a foil word, for example, dog-rat and cat-mouse. The examinee is then asked to select the previously presented original word (Green, Iverson, & Allen, 1999). After a 30-minute delay and without prior warning, the second effort measure, the DR task is administered, in which the examinee is once again presented with 40 word pairs consisting of a word from the original list and a new foil word, for example, dog-rabbit, and the examinee is then asked to identify the original word (Green et al., 1999). The CNS is the third effort measure where consistency of responses between the IR and the DR task is calculated, for example, if all responses for both tasks were exactly the same, then the examinee would obtain a score of 100%, and if only a half of the responses were the same, then the examinee would obtain a score of 50% (Green et al., 1999). The tasks assessing one's memory ability are administered after the DR task, for example, the MC task where the target word is presented in a list of eight other words; the PA recall task, where the examiner presents the first word of a pair from the original list, and the examinee is asked to recall its associated pair; and two FR tasks, short-delay FR and an optional long-delay FR task administered 20 minutes later (Green et al., 1999; Lezak et al., 2012).

According to Green et al. (1996), in the absence of dementia, a recognition performance of less than 82.5% on any of the primary effort measures should be classified as a task failure indicating a non-credible performance, thus an individual is showing suboptimal effort to do well on the task.

3.4.4 Reliable Digit Span

The RDS (Greiffenstein, Baker, & Gola, 1994) is an embedded PVT derived from the WAIS Digit Span (DS) subtest. The DS subtest assesses verbal attention, immediate recall and working memory, and consists of three components where examinees are asked to repeat a series of verbally presented numerical digits (Groth-Marnat & Wright, 2016; Webber & Soble, 2018). The first component, Digit Span Forward, consists of pairs of increasingly longer strings of numbers with each string containing two items (Babikian, Boone, Lu, & Arnold, 2006). For this task, examinees are asked to repeat the series of digits in exactly the same order as they were presented, and the task is discontinued once the examinee fails to complete both items for a given string correctly (Babikian et al., 2006). The total number for the Digit Span Forward is calculated by summing the total number of correctly repeated strings of numbers (Webber & Soble, 2018). Afterwards the examinee is presented with the second component, Digit Span Backward, where the task is to recite a new set of increasingly longer strings of numbers in backward order, and the task is discontinued following incorrect responses for both items in a given string (Babikian et al., 2006). The scoring is identical to that of the Digit Span Forward. The last component of the Digit Span subtest is the Digit Span Sequencing, where examinees are asked to respond to the presented strings of numbers by rearranging and reciting them in numerical order starting with the lowest number (Webber & Soble, 2018). The scoring is identical to that of the previous two components.

The RDS score is calculated by summing the longest string of digits recited (i.e. both items correctly recalled for a given string of numbers) for the Digit Span Forward and the Digit Span Backward (Strauss et al., 2006). Babikian et al. (2006) reported that a cut-off score of ≤6 achieved 45% sensitivity and 93% specificity, and a cut-off score of ≤7 achieved 62% sensitivity and associated specificity of 77%. Consistently, a systematic review and cross-validation study by Schroeder, Twumasi-Ankrah, Baade, and Marshall (2012) found that a cut-off score of ≤7 achieved global (i.e. for clinical as well as non-clinical samples) specificity below 90% and global sensitivity of 48% and 58% depending on the statistical method used for the analysis. However, a cut-off score of ≤6 achieved global specificity of 96% and 97% with associated global sensitivity of 30% and 35%. The study also reported that for patients with TBI including individuals with post-concussive/mild TBI as well as moderate/severe TBI, the RDS cut-off score of ≤6 achieved a specificity rate above 90%, and sensitivity of 38% for the post-concussive/mild TBI group and 26% for the moderate/severe TBI group. Overall, authors concluded that a cut-off score of ≤6 could be effectively used in many clinical populations to assess effort including TBI patients.

3.5 Effort Classification

The data set included outcomes (i.e. pass or fail) on previously validated effort measures including the Rey Fifteen Item Test (FIT), the Test of Memory Malingering (TOMM), the Word Memory Test (WMT), and the Reliable Digit Span (RDS). For the purpose of this research, outcomes on the WAIS Coding subtest were also provided including the Coding ACSS, the number of errors made, and whether the WAIS III or the WAIS IV battery was administered during the neuropsychological evaluation.

Effort was classified dichotomously: either “Valid Effort” or “Low Effort.” In order to determine each participant’s group membership, an effort measure failure rate was used. Thus participants who failed two or more effort tests were assigned to the “Low Effort” group and participants who passed all effort measures were assigned to the “Valid Effort” group. This classification was based on the previously discussed findings of the research of Victor et al. (2009), Cottingham and Boone (2014), and Larrabee (2012).

The outcome of each effort measure (i.e. pass or fail) was determined based on previously proposed cut-off scores, for example, for the TOMM a cut-off score of < 47 was employed based on Greve et al. (2006) research, for the RDS a cut-off score of < 7 was used based on the reviews of Rickards, Cranston, Touradji, and Bechtold (2018) and Schroeder et al. (2012), for the FIT a cut-off score of < 9 was employed based on Boone, Salazar, Lu, Warner-Chacon, and Razani's (2002) findings, and for the WMT on the standard cut-scores defined by Green et al. (1996) was employed.

3.6 Predictive Variables

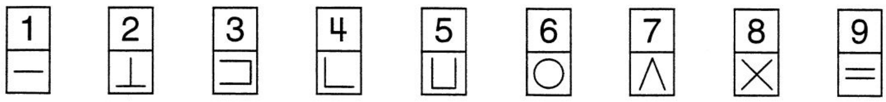
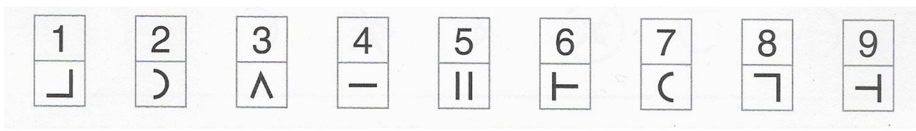
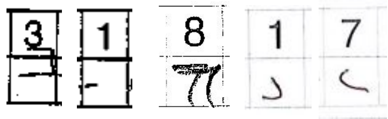
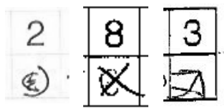
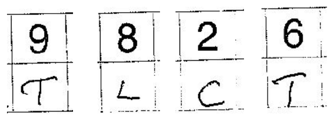
3.6.1 The WAIS Coding subtest

Coding is one of the core WAIS Processing Speed Index subtests which involves processing visually presented information (Etherton et al., 2007). The subtest not only measures processing speed, but also assesses short-term visual memory, learning ability, psychomotor speed, visual perception, visual-motor coordination, ability to follow instructions, visual scanning ability, cognitive flexibility, sequencing ability, attention, concentration, and motivation (Groth-Marnat & Wright, 2016; Wechsler et al., 2008).

The task includes a presentation of a stimulus sheet containing a key, which pairs nine digits with different symbols, and rows of boxes with a digit in the top part and a blank space in the bottom part (Lezak et al., 2012) (see Table 4). Using the key, the examinee is asked to fill in the empty parts of the boxes by drawing a symbol that has been paired with a number on the top part of the box (Wechsler & Psychological, 1997; Wechsler et al., 2008). This is a timed task with a limit of 120 seconds, and the examinee obtains one point for each correctly drawn symbol within the timeframe with a maximum of 133 points for the WAIS III and 135 points for the WAIS IV version (Wechsler & Psychological, 1997; Wechsler et al., 2008). Once the test is completed, the examiner uses the raw score and the WAIS technical and interpretive manual (Wechsler & Psychological, 1997; Wechsler et al., 2008) to calculate the ACSS score. In addition, a number of errors made on the Coding subtest were also calculated for each participant (i.e. a number of incorrectly drawn symbols) for the purpose of this thesis. Coding errors were defined as any substantially incorrect drafting of the symbol. Minor distortions or omissions caused by drafting in haste or psychomotor inaccuracy were not coded as errors unless the symbol became unrecognisable as the correct symbol. Any symbols represented by vertical or lateral mirroring were coded as error, as were symbols representing any other number. Instances that were incorrectly drafted, but were subsequently spontaneously corrected by redrafting (without prompting by the examiner), were not coded as errors (see Table 4).

Thus, three variables derived from the WAIS Coding subtest were the presence of a Coding error (Coding Error), number of Coding errors, and Coding ACSS; each was selected as predictive variables of "Low Effort."

Table 4: *Examples of accepted symbols and symbols coded as errors*

The WAIS III Coding Key Example	
	
The WAIS IV Coding Key Example	
	
Accepted Symbols	
Minor Drafting Inaccuracies	
Spontaneously Corrected Symbols	
Symbols Coded as Errors	
Incorrectly Drafted Symbols	

3.7 Statistical Analysis

This study used IBM SPSS Statistics version 25 to analyse the obtained data.

Descriptive statistics were used to summarise the demographic characteristics, TBI severity, and group membership ("Valid Effort" or "Low Effort") for each participant. Summaries for categorical variables were reported as number of cases (*n*) and percentages, but summaries for continuous variables were reported as means (*M*) and standard deviations (*SD*).

In order to assess if the Coding Error, the Coding ACSS and the Coding Combination score were predictive of low effort, Logistic Regression analyses were employed. This analytic procedure allows for a calculation of the statistical likelihood that an individual belongs to a certain group, and this technique is robust to non-normally distributed sample sets. The statistical analysis was carried out with "Effort" consisting of two categories: "Valid Effort" and "Low Effort" as a dependent variable, and either the Coding Error, Number of Coding Errors, the Coding ACSS, or the Coding Combination score, as an independent variable depending on the hypothesis tested. The Coding ACSS, the Coding Combination score, and the Number of Coding Errors were included as continuous variables and the Coding Error as a dichotomous categorical variable (i.e., presence of > 0 errors, or presence of 0 errors). The significance level was set at $p < 0.05$.

ROC curve analysis was used with an aim to determine the model's ability to differentiate between "Valid Effort" and "Low Effort." This is an essential analysis for a diagnostic test evaluation allowing for a calculation of sensitivity and specificity for every possible cut-score (Hajian-Tilaki, 2013). Furthermore, the area under the ROC curve (AUC) score was reported indicating the model's accuracy for predicting group membership (i.e. how well the Coding Error, Number of Coding Errors, the Coding ACSS, and the Coding Combination Score distinguishes between "Valid Effort" and "Low Effort" groups).

Moreover, in order to assess if Injury Severity impacts Coding performance, the Chi-Square Test and the One-way ANOVA analyses were undertaken. To determine if the Coding Error (dichotomous variable) was associated with injury severity, the Chi-Square Test was used. Further, the One-way ANOVA was used to determine if the Coding ACSS and the Coding Combination score was associated with injury severity.

In order to determine if demographic characteristics were associated with Coding Error, Chi-Square Tests were undertaken for the categorical variables (i.e. Gender, Compensation-seeking status, Ethnicity, or English as a Second Language), and Independent-Sample *t*-tests were used for continuous variables (i.e. Age and Years of Education).

Lastly, the Chi-Square Test was used in order to determine if there was a significant difference between rates of Coding error on the WAIS III and WAIS IV versions of Coding.

Chapter 4 Results

Of 650 participants, the Valid Effort group consisted of 395 (60.8%) participants who passed all administered effort measures, and the Low Effort group consisted of 152 (23.4%) participants who failed ≥ 2 administered PVTs. Of these, 16 (10.5%) participants from the Low Effort group scored below chance on the TOMM or WMT.

Table 5 shows demographic characteristics of the Valid and Low Effort groups. In brief, it was noted that 28.6% (*n* = 110) of male participants and 25.9% (*n* = 42) of female participants failed ≥ 2 effort measures and were assigned to the Low Effort group. A majority of participants (*n* = 426) were of European origin with 24.4% (*n* = 104) classified in the Low Effort group. Furthermore, 39.7% (*n* = 29) of Maori or of Pacific Island ethnicity, and 39.6% (*n* = 19) of Indian or Asian ethnicity were also classified in the Low Effort group. The majority of participants had sustained mTBI (*n* = 314) of whom 36.3% (*n* = 114) fell within the Low Effort group. Furthermore, 19.5 % (*n* = 15) of those with Moderate TBI, 21.9% (*n* = 16) of those with Severe TBI, and only 8.4% (*n* = 7) of those with Very Severe TBI failed ≥ 2 administered PVTs. Additionally, 31.3% (*n* = 143) of compensation-seeking participants and 10.1% (*n* = 9) of those who were not seeking compensation were assigned to the Low Effort group. The mean age in the Low Effort group was 44.1 years (*SD* = 10.97) and the mean age for the Valid Effort group was 42.18 (*SD* = 13.17). Lastly, the mean for years in education in the Low Effort group was 11.36 (*SD* = 2.58) and 12.21 (*SD* = 2.52) in the Valid Effort group.

In addition, the Chi-Square Tests (for categorical variables) and the Independent-Samples *t*-Tests (for continuous variables) were undertaken to explore the relationship between the participants' characteristics and Effort. As Table 5 shows there was a significant positive relationship between Effort and Ethnicity (p = .002), Compensation-Seeking status ($p < .001$), and English as a Second Language ($p < .001$)). Furthermore, there was a significant inverse relationship between Effort and Injury Severity ($p < .001$) and Years of Education (p = .001). The analyses also revealed that there was no statistically significant relationship between Effort and Gender (p = .53), and between Effort and Age (p = .13).

Table 5: *Characteristics of the participants grouped according to effort*

Variable	Valid Effort		Low Effort			Effect Size
	n	%	*n*	%	*p*	(*d*/*φ*)
Gender					.53	.03
Male	275	71.4	110	28.6		
Female	120	74.1	42	25.9		
Ethnicity					.002	.15
European/White	322	75.6	104	24.4		
Maori/Pacific Island	44	60.3	29	39.7		
Indo/Asian	29	60.4	19	39.6		
TBI Severity					<.001	.24
mTBI	200	63.7	114	36.3		
Moderate	62	80.5	15	19.5		
Severe	57	78.1	16	21.9		
Very severe	76	91.6	7	8.4		
Compensation-Seeking					<.001	.18
Yes	314	68.7	143	31.3		
No	80	89.9	9	10.1		
ESL					<.001	.23
Yes	24	42.1	33	57.9		
No	371	75.7	119	24.3		
	M	*SD*	*M*	*SD*		
Age	42.18	13.17	44.01	10.97	.13	.13
Education, years	12.21	2.52	11.36	2.58	.001	.30

Note: p-values are from Independent-Samples *t*-tests or Chi-Square tests. Effect sizes for *t*-tests are Cohen's *d* and Cramer's Phi (*φ*) for Chi-Square tests. Indo/Asian = participants reporting themselves of Indian or Asian ethnicity; TBI = traumatic brain injury; mTBI = mild traumatic brain injury; ESL = English as a Second Language.

4.1 WAIS III and WAIS IV version

The Chi-Square Test was used in order to determine if there is a significant relationship between Coding Error on the WAIS III and Coding Error on the WAIS IV version. Results of the analysis revealed that that there is no significant association between Coding Error (dichotomous variable) on the WAIS III and Coding Error on the WAIS IV, $\chi^2(1, N = 650) = .54, p = .46$.

4.2 Demographic characteristics and Coding Error

In order to determine if demographic characteristics are associated with Coding Error, Chi-Square Tests were undertaken for the categorical variables (i.e. Gender, Compensation-seeking status, Ethnicity, English as a Second Language), and Independent-Sample *t*-tests were used for the continuous variables (i.e. Age and Years of Education) (see Table 6). Because the analyses were exploratory in nature, no statistical correction for multiple analyses (e.g. Bonferroni correction) was undertaken.

Results of the analysis showed that there was no association between Gender and Coding Error, $\chi^2(1, N = 650) = 1.54, p = .21$, and no association between Ethnicity and Coding Error, $\chi^2(2, N = 650) = 5.53, p = .06$. However, results also revealed that there was an association between English as a Second Language and Coding Error, $\chi^2(1, N = 650) = 8.55, p = .003$, and an association between Compensation-Seeking status and Coding Error, $\chi^2(1, N = 649) = 4.65, p = .03$.

The Independent-Sample *t*-tests showed that there was no significant difference between the mean Age for those who made 0 Coding Errors and for those who made > 0 Coding Errors ($t(648) = -1.62, p = .11$). Furthermore, the analysis revealed that there was no significant difference between the mean of years of education for those who made 0 Coding Errors and those who made > 0 Coding Errors ($t(648) = 1.31, p = .19$).

These results showed that Gender, Age, Years of Education, and Ethnicity are unrelated to Coding Error. Furthermore, analysis revealed that there is a statistically significant relationship between Compensation-Seeking status and Coding Error and also English as a Second Language and Coding Error in the direction of those making a Coding error being more likely to be Compensation-seeking and to have English as a second language.

Table 6: *Association between demographic characteristics and Coding Error*

Variable	χ^2/t	*p* value
Gender	1.54	0.21
Age	-1.62	0.11
Years of Education	1.31	0.19
Compensation-Seeking status	4.65	0.03
Ethnicity	5.53	0.06
ESL	8.55	0.003

Note: ESL = English as a second language.

4.3 Coding Embedded Effort Measures as Predictors of Low Effort

4.3.1 Coding Error

Logistic regression analysis was undertaken with Effort as a dependent variable and Coding Error as a dichotomous categorical variable. Results revealed that the Coding Error is a significant predictor of Low Effort ($p < .001$), such that the presence of > 0 errors on the subtest increases the odds of low effort almost 18 times relative to those not making any errors (see Table 7).

4.3.2 Number of Coding Errors

Logistic regression analysis was also conducted with Effort as a dependent variable and Number of Coding Errors as a continuous variable. Results revealed that Number of Coding Errors is a significant predictor of Low Effort ($p < .001$) such that making 1 error on the Coding subtest increases the odds of low effort about 4 times (see Table 7).

4.3.3 Coding ACSS

Logistic regression analysis was undertaken with Effort as a dependent variable and Coding ACSS as a continuous variable. Results revealed that ACSS is also predictive of Low Effort ($p < .001$) such that an increase by 1 scaled score reduces the odds of low effort by 0.65 (or by 35%) (see Table 7).

4.3.4 Coding Combination score

Coding Errors and ACSS scores were arithmetically summed to create a new combined variable (i.e. Coding Combination score) based on the research of Whiteside et al. (2018), Whiteside, Gaasedelen, et al. (2015), and Whiteside, Kogan, et al. (2015). Following the method, a new score for each participant was calculated so that a higher score indicated a more positive performance. As a result, ACSSs (with a minimum of 1 and a maximum score of 19) were reverse coded so that the scores run in the opposite direction, for example, a score of 1 was coded as 19; 2 was coded as 18; and so on until the final score of 19 was coded as 1.

Logistic regression analysis was conducted with Effort as a dependent variable, where the combined score of Number of Coding Errors and ACSS were summed as a continuous variable, with an aim to assess if the new variable is also predictive of low effort. Results showed that the combination of the 2 variables was significantly predictive of Low Effort ($p < .001$) such that an increase of 1 unit of the combination score increased the odds of low effort by 1.55 times (see Table 7).

Table 7: *Logistic regression analyses of the Coding variables as predictors of Low Effort*

Coding Variables	OR (95% CI)	*p* value
Coding Error	17.63 (9.27, 33.54)	<.001
Num Coding Errors	3.89 (2.48, 6.10)	<.001
Coding ACSS	0.65 (0.59, 0.71)	<.001
Coding Combination score	1.55 (1.42, 1.69)	<.001

Note. Num Coding Errors = number of Coding errors; Coding ACSS = Coding age-corrected scaled score; Coding Combination score = number of Coding errors and age corrected scaled score sum.

4.4 Coding Variables and Injury Severity

Previous research has demonstrated that individuals affected by more significant traumatic brain injuries obtain lower scores on the WAIS Processing Speed Index (PSI) subtests (Axelrod, Fichtenberg, Liethen, Czarnota, & Stucky, 2001; Donders & Strong, 2015). Particularly, the Coding subtest is well known to be one of the most sensitive subtests to any type of organic and functional impairment (Groth-Marnat & Wright, 2016). In light of these findings, it is important to assess if Coding Embedded Effort are associated with injury severity.

4.4.1 Coding Error

The analysis revealed that there was a statistically significant relationship between Coding Error and Injury Severity (χ^2 (3, N = 650) = 8.54 p = .04) (see Table 8). This indicates that there is an association between making errors on the Coding subtest and severity of TBI. However, a closer inspection of the expected and the observed counts in the cross tabulation table suggested an inverse relationship between injury severity and Coding Error such that a statistically significant trend that individuals with *more* severe brain injuries appear *less* likely to make errors on the Coding subtest than are those affected by mild TBI.

4.4.2 Coding ACSS

One-way ANOVA analysis was undertaken in order to determine if the Coding ACSS is associated with injury severity. The analysis was conducted excluding those with Low Effort due to the fact that Low Effort may be artificially suppressing the ACSS across the whole sample. Descriptive statistics showed that the mTBI group obtained an average ACSS of 9.28 (SD = 3.02), the Moderate TBI group obtained an average ACSS of 8.34 (SD = 3.09), the Severe TBI group obtained an average ACSS of 8.09 (SD = 2.34), and the Very Severe TBI group obtained an average ACSS of 7.07 (SD = 2.98). Results demonstrated that there was a significant difference between the Injury Severity group means in the direction of lower ACSS associated with more severe TBI ($F(3, 391) = 11.10, p < .001$) (see Table 8). The Tukey's HSD Post Hoc test revealed that there was a significant mean difference between mTBI and Severe TBI ($p = .04$), mTBI and Very Severe TBI ($p < .001$), but not between mTBI and Moderate brain injury ($p = .13$), Moderate and Severe TBI ($p = .97$), Moderate and Very Severe TBI ($p = .57$), and Severe and Very Severe ($p = .52$). Due to the findings, that there was no statistically significant difference between Severe and Very Severe TBI groups, and there was a significant mean difference between mTBI and both Severe and Very Severe groups, those Severe and Very Severe groups were collapsed. Thus, the ROC analyses were run for the three TBI severity groups separately (i.e. mTBI, Moderate TBI, and Severe/Very Severe TBI group combined) with an aim to establish appropriate cut-off scores and associated sensitivity and specificity for each group.

4.4.3 Coding Combination score

One-way ANOVA analysis was conducted to determine whether the Coding Combination score is associated with injury severity. The analysis showed that there is a significant mean difference among the Injury Severity groups for the Coding Combination score ($F(3,646) = 3.32, p = .02$) (see Table 8). The Tukey's HSD Post Hoc test demonstrated that there was a significant difference between mTBI and Very Severe TBI ($p = .02$), but there was no statistically significant difference between mTBI and Moderate TBI ($p = 1.0$), mTBI and Severe TBI ($p = .57$), Moderate and Severe TBI ($p = .75$), Moderate and Very Severe ($p= .11$), and Severe and Very Severe TBI ($p = .63$). As a result, it can be concluded that the Coding Combination is associated with injury severity, and due to these and previous findings of TBI severity affecting the Coding ACSS, ROC analyses were run for 3 injury severity groups separately (i.e. mTBI, Moderate TBI, and Severe/Very Severe TBI group combined) in order to establish appropriate cut-off scores and associated sensitivity and specificity for each severity group.

Table 8: *Relationship between Coding variables and Injury Severity*

Coding Embedded Effort Variables	F/χ^2	*p* value
Coding Error	8.54	.04
Coding ACSS	11.10	<.001
Coding Combination score	3.32	.02

Note. Coding ACSS = Coding age-corrected scaled score; Coding Combination score = number of Coding errors and age corrected scaled score sum.

4.5 Sensitivity and Specificity

4.5.1 Coding Error

ROC analysis was undertaken with Coding Error (dichotomous variable) as a test variable and Effort as a state variable to establish the predictive accuracy and sensitivity and specificity for the cut-score. Results revealed that the Coding Error achieved a low predictive accuracy (AUC = 0.67). The AUC is a combined measure of sensitivity and specificity describing the overall accuracy of a test. As can be seen in Table 9, the Coding Error cut-score was associated with low sensitivity, likely impacting the predictive accuracy resulting in a low AUC score. By contrast, results show that Coding Error was highly specific, with specificity rates of > 95%, indicating low likelihood of false positive errors. As presented in Table 9, results showed that making > 0 errors on the Coding subtest is associated with 96.7% specificity and 37.5 % sensitivity.

4.5.2 Number of Coding Errors

A ROC analysis was undertaken with Number of Coding Errors (continuous variable) as a test variable and Effort as a state variable to establish sensitivity and specificity for various cut-scores. As can be seen in Table 9, Number of Coding Errors showed a low level of predictive accuracy (AUC=0.67). However, as mentioned before, this is likely due to low sensitivity. Overall results showed that making > 0 errors on the Coding subtest is associated with 96.7% specificity and 37.5 % sensitivity. Furthermore, making ≥ 2 errors is associated with almost perfect specificity (98.7%), but reduced sensitivity (19.1%) (see Table 9).

4.5.3 Coding ACSS

Three separate ROC analyses were run to identify the most appropriate cut-scores and their respective sensitivity and specificity rates. As shown in Table 9, the ACSS showed that Coding ACSS had very good predictive accuracy (AUC = 0.83) for the mTBI group and a cut- score of ≤ 5 achieved 90.5% specificity and 56.1% sensitivity. Furthermore, a cut-score of ≤ 4 achieved an excellent specificity of 97% with lower sensitivity of 39.5%. Coding ACSS achieved a good predictive accuracy (AUC = 0.75) for the Moderate TBI group, and results showed that a cut- score of ≤ 4 achieved 93.5% specificity with associated sensitivity of 33.3%, and a cut- score of ≤ 3 was associated with specificity of 96.8% and 26.7% sensitivity. For the combined group of Severe and Very Severe TBIs, the ACSS also showed a very good predictive accuracy (AUC = 0.86). The analysis revealed that a cut-score of ≤ 3 achieved 94% specificity and 56.5% sensitivity, and the cut-score of ≤ 2 was associated with 96.2% specificity but only 17.4% sensitivity.

4.5.4 Coding Combination score

Three separate ROC analyses were run to identify the most appropriate cut-scores and their respective sensitivity and specificity rates. As shown in Table 9, the Coding Combination score (i.e. arithmetic sum of Coding Error and Coding ACSS) showed very good predictive accuracy (AUC=0.85) for the mTBI group and a cut-score of > 14 achieved 90.5% specificity and 63.2% sensitivity. Furthermore, a cut-score of > 15 achieved an excellent specificity of 97% and 44.7% sensitivity, and a cut-score of > 17 achieved a perfect specificity of 100% but was associated with lower sensitivity of 22.8%. The Coding Combination score showed good predictive accuracy (AUC=0.76) for the Moderate TBI group and a cut-score of > 15 achieved 90.3% specificity and 40% sensitivity. In addition, a cut-score of > 16 was associated with 93.5% specificity and 33.3% sensitivity, and a cut-score of 17 achieved a perfect specificity of 100% with 20% sensitivity. Lastly, for the Severe/Very Severe TBI group, the Coding Combination score demonstrated a very good predictive accuracy (AUC=0.88). The cut-score of > 16 achieved 93.2% specificity and 69.6% associated sensitivity, and a cut-score of > 18 achieved excellent specificity of 97% with 21.7% sensitivity.

Table 9: *Sensitivity and Specificity values for the Coding variable cut-scores*

Cut-Score	AUC	Sensitivity	Specificity
Coding Errors	0.67		
> 0		37.5	96.7
≥ 2		19.1	98.7
≥ 3		11.8	99.5
≥ 4		8.6	99.5
Coding ACSS			
mTBI	0.83		
≤ 2		12.3	100.0
≤ 3		21.1	99.5
≤ 4		39.5	97.0
≤ 5		56.1	90.5
Moderate TBI	0.75		
≤ 2		13.3	100.0
≤ 3		26.7	96.8
≤ 4		33.3	93.5
≤ 5		60.0	82.3
Severe/Very Severe TBI	0.86		
≤ 2		17.4	96.2
≤ 3		56.5	94.0
≤ 4		78.3	88.0
≤ 5		82.6	75.9
Coding combination score			
mTBI	0.85		
> 14		63.2	90.5
> 15		44.7	97.0
> 16		29.8	99.5
> 17		22.8	100.0
> 18		15.8	100.0
Moderate TBI	0.76		
> 14		60.0	80.6
> 15		40.0	90.3
> 16		33.3	93.5
> 17		20.0	100.0
> 18		6.7	100.0
Severe/Very Severe TBI	0.88		
> 15		78.3	86.5
> 16		69.6	93.2
> 17		39.1	95.5
> 18		21.7	97.0
> 19		13.0	99.2

Note. Coding ACSS = Coding age-corrected scaled score; TBI = traumatic brain injury; mTBI = mild traumatic brain injury.

4.6 Positive and Negative Predictive Values

Positive predictive value (PPV) and negative predictive value (NPV) were calculated for 20%, 30%, 40%, and 50% base rates based on the methodology described by Glaros and

Kline (1988). As can be seen from Table 10, PPV and NPV was calculated for each identified cut-score for each of the analysed Coding embedded effort measures. This allows for establishing the likelihood that an individual is showing Low Effort given a positive finding on each measure on the Coding scale (PPV), and the likelihood that an individual is showing Good Effort given a negative finding on a Coding measure (NPV). Overall, for all embedded Coding measures, the results show high PPV, particularly for higher base rates, and somewhat lower NPV with higher values associated with lower base rates. For example, it was noted that for a cut-score of > 0 Coding errors, at 40% base rate, achieves 88% probability of correctly identifying Low Effort given that an individual makes > 0 errors (positive finding), and 70% probability of correctly identifying Good Effort given that an individual does not make any errors in the Coding subtest (negative finding). Similarly, at a 40% base rate, a cut-off of ≥ 16 for the Coding combination score, for the mTBI, Moderate TBI, and Severe TBI group, is associated with 89% probability of correctly identifying an individual showing Low Effort given that the individual achieves a score of 16 or more, and 78% probability of correctly identifying an individual with Good Effort given that the individual achieves a score of < 16.

Table 10: *Positive and negative predictive power for 20% 30%, 40%, and 50% base rates*

	20% Base rate		30% Base rate		40% Base rate		50% Base rate	
Cut-off scores	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)
Coding Errors								
> 0	0.74	0.86	0.83	0.78	0.88	0.70	0.92	0.61
≥ 2	0.79	0.83	0.86	0.74	0.91	0.65	0.94	0.55
≥ 3	0.86	0.82	0.91	0.72	0.94	0.63	0.96	0.53
≥ 4	0.81	0.81	0.88	0.72	0.92	0.62	0.95	0.52
Coding ACSS								
mTBI								
≤ 2	1.00	0.82	1.00	0.73	1.00	0.63	1.00	0.53
≤ 3	0.91	0.83	0.95	0.75	0.97	0.65	0.98	0.56
≤ 4	0.77	0.87	0.85	0.79	0.90	0.71	0.93	0.62
≤ 5	0.60	0.89	0.72	0.83	0.80	0.76	0.85	0.67
Moderate TBI								
≤ 2	1.00	0.82	1.00	0.73	1.00	0.63	1.00	0.54
≤ 3	0.68	0.84	0.78	0.75	0.85	0.66	0.89	0.57
≤ 4	0.56	0.85	0.69	0.77	0.77	0.68	0.84	0.58
≤ 5	0.46	0.89	0.59	0.83	0.69	0.76	0.77	0.67
Severe/Very Severe TBI								
≤ 2	0.53	0.82	0.66	0.73	0.75	0.64	0.82	0.54
≤ 3	0.79	0.90	0.80	0.83	0.86	0.76	0.90	0.68
≤ 4	0.62	0.94	0.74	0.90	0.81	0.86	0.87	0.80
≤ 5	0.46	0.95	0.59	0.91	0.70	0.87	0.77	0.81
Coding combination score								
mTBI								
> 14	0.62	0.91	0.74	0.85	0.82	0.79	0.87	0.71
> 15	0.79	0.88	0.86	0.80	0.91	0.72	0.94	0.64
> 16	0.94	0.85	0.96	0.77	0.98	0.68	0.98	0.59
> 17	1.00	0.84	1.00	0.75	1.00	0.66	1.00	0.56
> 18	1.00	0.83	1.00	0.73	1.00	0.64	1.00	0.54
Moderate TBI								
> 14	0.44	0.89	0.57	0.82	0.67	0.75	0.76	0.67
> 15	0.51	0.86	0.64	0.78	0.73	0.69	0.80	0.60
> 16	0.56	0.85	0.69	0.77	0.77	0.68	0.84	0.58
> 17	1.00	0.83	1.00	0.74	1.00	0.65	1.00	0.56
> 18	1.00	0.81	1.00	0.71	1.00	0.62	1.00	0.52
Severe/Very Severe TBI								
> 15	0.59	0.94	0.71	0.90	0.79	0.86	0.85	0.80
> 16	0.72	0.92	0.81	0.88	0.87	0.82	0.91	0.75
> 17	0.68	0.86	0.79	0.79	0.85	0.70	0.90	0.61
> 18	0.64	0.83	0.76	0.74	0.83	0.65	0.88	0.55
> 19	0.80	0.82	0.87	0.73	0.92	0.63	0.94	0.53

Note. Coding ACSS = Coding age-corrected scaled score; TBI = traumatic brain injury; mTBI = mild traumatic brain injury.

Chapter 5 Discussion

The aim of this study was to investigate whether measures derived from the WAIS Coding subtest could be employed as embedded performance validity indicators during neuropsychological assessment; embedded measures included presence of a Coding Error, Number of Coding Errors, the Coding ACSS, and a newly devised Coding Combination score (i.e. an arithmetic sum of Coding Errors and the Coding ACSS). Furthermore, this research examined whether these embedded Coding effort measures are valid in a range of traumatic brain injury severities.

Overall, out of 650 participants referred to the private clinical neuropsychology practice for the purpose of a neuropsychological assessment following TBI, 60.8% passed all administered effort measures, operationalised as affording Valid Effort, and 23.4% failed ≥ 2 previously validated PVTs, operationalised as affording Low Effort. These results also suggested that the base rate for Low Effort in a mixed severity sample of TBI survivors is slightly above 20%. This is somewhat lower in comparison to the previous literature. However, Webb et al. (2012) found similar rates of effort test failure among their traumatic brain injury sample in a New Zealand context. The authors suggested that this may be due to the inclusion of a higher proportion of individuals with more severe TBIs in comparison to other research studies; previous literature has demonstrated PVT failure is more associated with those affected by mild TBI than severe TBI (Green et al., 2001; Stevens et al., 2008; Webb et al., 2012; West, Curtis, Greve, & Bianchini, 2011).

Furthermore, when the mTBI group was explored in isolation, it was noted that 36.3% failed ≥ 2 effort measures. These results are consistent with previous estimates of suboptimal effort in this population, such that estimates range from 30% to 40% among individuals with mTBIs (Larrabee, 2000; Mittenberg et al., 2002). This study found that 19.5% of those with Moderate TBI, 21.9% of those with Severe TBI and only 8.4% of those with Very Severe brain injuries failed ≥ 2 PVTs. These results are also consistent and supportive of previous findings that individuals with the most severe brain injuries are less likely to fail effort measures, and that individuals with mTBI tend to perform significantly lower on PVTs than any other injury severity group (Green et al., 2001; Stevens et al., 2008; Webb et al., 2012).

Furthermore, consistent with the previous literature (Bianchini et al., 2006; Webb et al., 2012), the present study found that a large proportion of those with Compensation-

Seeking status demonstrated Low Effort during a neuropsychological evaluation (31%), in comparison to only 10% of those who were not Compensation-Seeking.

5.1 Coding Errors

This study found that making any Coding error is a statistically significant predictor of Low Effort within the TBI population, such that making more than 0 errors on the WAIS Coding subtest increases the odds of Low Effort almost 18 times. The study findings also indicated that making any Coding Error is a highly specific measure of Low Effort with specificity rates above 95%. High specificity means that the identified cut-scores produce low levels of false positive errors. This is particularly important due to the significant and negative implications of any erroneous conclusion that an individual's cognitive effort during neuropsychological assessment is non-credible (Greve & Bianchini, 2004). However, results also revealed that sensitivity levels for a range of cut-scores remained low, for example, 19% for the cut-score of ≥ 2, and 12% for the cut-score of ≥ 3. This finding is consistent with previous research exploring embedded WAIS PVTs (Erdodi, Abeare, et al., 2017; Inman & Berry, 2002; N. Kim et al., 2010; Trueblood, 1994). Previous studies have showed that embedded PVTs have lower sensitivity than free-standing measures (Cottingham & Boone, 2014). It has been suggested that it may be due to the wide range of scores obtained from credible individuals, therefore the cut-scores need to be set at a lower level in order to increase specificity (Miele et al., 2012).

Results of this study identified that the cut-score of > 0 Coding errors was associated with 97% specificity and 38% sensitivity. This indicates that when a cut-score of > 0 errors is used, only 3% of individuals with Valid Effort will be misclassified as demonstrating Low Effort, and that 38% of individuals with Low Effort will be correctly classified, but 62% of those applying low effort will go undetected. Furthermore, the cut-score of > 0 Coding errors at a 20% base rate, achieved 74% probability of correctly identifying Low Effort given that an individual makes > 0 errors, and 86% probability of correctly identifying Valid Effort given that an individual does not make any errors on the Coding subtest. Results also demonstrated that this cut-score at a 30% base rate, achieved 83% likelihood of correctly identifying Low Effort, and 78% likelihood of identifying Valid Effort.

A cut-score of ≥ 2 errors was associated with 99% specificity, meaning that only 1% of individuals with Valid Effort will be misclassified as not performing to the best of their ability (i.e. showing Low Effort). However, employing this more conservative cut-score, resulted in sensitivity levels reduced to 19%.

Overall, analyses have revealed that the Coding Error sensitivity levels from this study were somewhat consistent with the research of N. Kim et al. (2010) exploring only 180° rotational errors on the WAIS III Digit Symbol-Coding subtest as part of an additional recognition task in a wide neuropsychological sample. However, sensitivity rates for the Coding Error in this study are lower than in three previous studies by Chaftez and colleagues exploring the use of Coding Error as an embedded validity indicator (Chafetz, 2008, 2011; Chafetz et al., 2007). It is possible that this is due to different classification methods used for the known group of Low Effort. The previous studies employed the Test of Memory Malingering (TOMM), the Medical Symptom Validity Test (MSVT), and also the “A” Random Letter Test for the most recent study, thus only the TOMM was consistently used across previous research and the present study. However, the three previous studies applied considerably lower cut-off scores for the TOMM, namely a cut-score of < 18 was used for the definite malingered neurocognitive dysfunction (MND) group, and a cut-score of 18-32 for the probable MND group. The current study employed the cut-score of < 47 to classify an individual to the Low Effort group. This indicates that the sample in previous studies included a higher proportion of those performing below or at chance-level than for the current study, potentially impacting the overall performance on the WAIS Coding subtest as a PVT.

Exploratory analyses of demographic characteristics showed that there was no association between Coding Error and Gender, Age, Years of Education, and Ethnicity. Similarly, other studies have also found no relationship between failure on PVTs and Gender (Inman et al., 1998; Lee, Graham, Sellbom, & Gervais, 2012; Stevens et al., 2008), and between PVT failure and Age (Stevens et al., 2008; Webb et al., 2012). Consistent with the present study, the literature also demonstrates no relationship between Ethnicity and failure on PVTs (Inman et al., 1998; Meyers et al., 2011; Webb et al., 2012). Interestingly, Webb et al. (2012) found that having a foreign-born status was predictive of effort test failure. It may be that this is related to the present study’s finding of an association between English as a Second Language and making errors on the Coding subtest. However, these findings are inconsistent with the study by Erdodi, Nussbaum, et al. (2017) that showed that only PVTs with high verbal mediation are associated with increased failure rates if administered in the non-dominant language. Furthermore, authors found that none of their participants with a different linguistic background failed the WAIS Coding and the Rey

Fifteen Item Test as PVTs, which are considered to be measures with low verbal mediation. It is important to note that the research study by Erdodi, Nussbaum, et al. (2017) examined these measures with healthy individuals, whereas the sample for the current study consisted of individuals with traumatic brain injuries. These findings suggest that outcomes of PVTs, even of those with low verbal mediation, may be impacted by English proficiency among people with TBI. Thus, clinicians should always be cautious when interpreting failure on PVTs administered in English to those with a different linguistic background. Furthermore, inconsistent with the present study, previous research has found a relationship between PVT failure and years of education, such that lower education has been associated with increased likelihood of failing PVTs (Babikian et al., 2006; Greve, Ord, Bianchini, & Curtis, 2009; Stevens et al., 2008; Webb et al., 2012). This finding may be specific to the Coding Error suggesting that years of education are not related to making errors on the Coding Subtest but may be related to failure on other PVTs. Finally, the present study found a significant relationship between Coding Error and Compensation-Seeking status such that those making errors on the Coding subtest are more likely to be Compensation-Seeking, a finding that is consistent with previous literature (Bianchini et al., 2006; Webb et al., 2012).

WAIS Coding performance is known to be inversely correlated with TBI severity such that those with more severe injuries tend to perform more poorly on Coding (Groth-Marnat & Wright, 2016). The present study explored whether the Coding Error, as an embedded PVT, exhibited bias against those with more significant brain injuries. Findings demonstrated an inverse relationship between Coding Error and injury severity such that individuals with more severe brain injuries were less likely to make Coding Errors than those affected by mild traumatic brain injuries. These findings are consistent with previous research showing that those affected by mTBI are more likely to fail effort measures (i.e. showing low effort) than those with more significant brain injuries (Green et al., 2001; Stevens et al., 2008; Webb et al., 2012; West et al., 2011).

Lastly, there have been a number of changes made to the Coding subtest in the most recent version of the WAIS (i.e. the WAIS IV) (Wechsler et al., 2008). The present study aimed to explore whether there was any significant difference between the performance on the WAIS-III and the WAIS-IV Coding subtest in respect of the likelihood of making an error. Findings showed, that there was no association between Coding Errors and the WAIS version. This suggests the identified cut-scores can be used for either version of the Coding subtest.

Overall, the findings from the present study suggest that due to the excellent specificity levels of > 95% and the easily obtainable score, Coding Error has the potential to be a valuable embedded effort measure for clinicians. It has to be noted, that due to low sensitivity, Coding Error should be used in conjunction with other PVTs when assessing effort.

5.2 Coding ACSS

This study also validated the Coding ACSS as an embedded PVT. According to Inman and Berry (2002) cross-validation of embedded effort measures is particularly important due to a greater variability in scores in comparison to free-standing PVTs. The findings were consistent with previous research showing that the Coding ACSS is a statistically significant predictor of Low Effort. However, this study found that the Coding ACSS is sensitive to TBI severity, which is consistent with previous research, showing that those individuals with more severe brain injuries tend to obtain lower scores on the WAIS processing speed subtests (Axelrod et al., 2001; Donders & Strong, 2015). Due to potential bias against those with more significant brain injuries, it was essential to identify cut-scores and their associated sensitivity and specificity rates for each injury-severity group.

The analyses revealed that the Coding ACSS had a very good predictive accuracy of Low Effort among individuals with mTBI, such that AUC for mTBI achieved a score of 0.83. For the mTBI group a cut-off score of ≤ 5 achieved around 91% specificity with somewhat higher sensitivity than previously reported in the literature (56%). Previous research found that for a cut-score of ≤5 sensitivity levels ranged from 4% to 29% at > 90% specificity (Erdodi, Abeare, et al., 2017; Erdodi & Lichtenstein, 2017). Furthermore, the present study found that a cut-off score of ≤ 4 achieved 94% specificity and almost 40% sensitivity; again, these rates are also somewhat higher than those previously reported in the literature (Etherton et al., 2007; Inman & Berry, 2002; N. Kim et al., 2010; Trueblood, 1994). It is possible that variety in sensitivity rates across studies is due to the inclusion of different populations, for example, diverse neuropsychological patients (N. Kim et al., 2010), only individuals with moderate and severe brain injuries (Etherton et al., 2007) or only mild TBIs (Inman & Berry, 2002; Trueblood, 1994), and also a variety of research designs have been utilised. The present study also identified associated PPV and NPV for mTBI, for each cut-score at various base rates, for example, at the 30% base rate, a cut-score of ≤ 5 was associated with 72% likelihood of correctly identifying those with Low Effort and 83% likelihood of correctly identifying those with Valid Effort.

Furthermore, a cut-off score of ≤ 4 was associated with increased likelihood of correctly identifying 85% of mTBI individuals with Low Effort, and 79% likelihood of correctly identifying mTBI individuals with Valid Effort.

Findings of this study revealed that the Coding ACSS had a good predictive accuracy for Low Effort for those with Moderate TBI, with an AUC score of 0.76. Findings revealed that applying a cut-off score of ≤ 5 would not be appropriate for that injury severity group due to inadequate specificity levels, thus allowing for unacceptable rates of false positive errors in that group. However, a cut-off score of ≤ 4 achieved almost 94% specificity and 33% sensitivity. Furthermore, results demonstrated that at a 30% base rate, there is a 69% likelihood of correctly identifying those with Low Effort, and 77% likelihood of correctly identifying those with Valid Effort. Findings also showed that a cut-score of ≤ 3 achieved almost 98% specificity and almost 27% sensitivity. If the same base rate is applied, then there is 78% likelihood that an individual with Moderate TBI is correctly identified with Low Effort given that the person obtains an age corrected scaled score of 4 or less on the Coding subtest, and 75% likelihood of correctly identifying an individual with Valid Effort given that the person obtains a score higher than 4.

Results of this study revealed that the Coding ACSS has a very good predictive accuracy with an AUC score of 0.86 for the Severe and Very Severe TBI group. Findings demonstrated that the cut-scores of ≤ 5 and ≤ 4 would be not be appropriate to apply for this group due to specificity levels < 90%, thus these cut-scores present high levels of false positive errors. However, a cut-score of ≤ 3 achieved 94% specificity and good sensitivity of almost 57%. Furthermore, at a 30% base rate, this cut-score was associated with 80% likelihood of correctly identifying Low Effort given that an individual obtains a score of 3 or less, and 83% probability of correctly identifying Valid Effort if an individual obtains a score of greater than 3. The analysis of this study also revealed that if a cut-score of ≤ 2 is used, then the associated specificity level increased to 96%, but sensitivity reduced to 17%. This means that 83% of individuals with Severe/Very Severe TBI showing Low Effort would go undetected, and 4% of individuals with Valid Effort would be misclassified as demonstrating Low Effort.

Overall, the Coding ACSS shows a very good predictive accuracy for Low Effort as an embedded validity indicator. However, it is important for clinicians to apply the appropriate cut-scores consistent with the injury severity when assessing effort as a means of minimising the risk of making a false positive error with patients affected by severe and very severe TBIs.

5.3 Coding Combination score

Lastly, this study examined if a Coding Combination Score (i.e. an arithmetic sum of Number of Coding Errors and Coding ACSS) is predictive of Low Effort. This was based on recent research suggesting an increased accuracy of identifying Low Effort when multiple individual embedded measures have been combined (Whiteside et al., 2018; Whiteside, Gaasedelen, et al., 2015; Whiteside, Kogan, et al., 2015; Zasler & Bender, 2019). Findings of the present study revealed that the Coding Combination score was significantly predictive of Low Effort. However, results showed that the Coding Combination score was also sensitive to TBI severity, which is consistent with previous findings (Axelrod et al., 2001; Donders & Strong, 2015) and the present study's findings in respect of the Coding ACSS. Due to potential bias against those with more significant brain injuries, it was important to identify cut-off scores and their associated sensitivity and specificity rates for each injury-severity group.

Overall, the Coding Combination score achieved a very good predictive accuracy with the AUC score of 0.85 for the mTBI group, which is somewhat higher than for the Coding ACSS alone. Findings showed that the cut-score of > 14 achieved 91% specificity and high sensitivity of 63%. The Coding ACSS for the same specificity level achieved lower sensitivity, therefore showing an increased ability to identify individuals with Low Effort. Furthermore, at a 30% base rate, this cut-score was associated with 74% likelihood of correctly identifying Low Effort given that an individual obtains a score greater than 14, and 85% likelihood of correctly identifying Valid Effort given that an individual achieves a score of 14 or less. Furthermore, a cut-score of > 15 was associated with excellent specificity of 97% and almost 45% sensitivity, which once again, was higher than for the Coding ACSS at the same specificity level of 97%. The probability of correctly identifying Low Effort was 86%, and 80% likelihood at correctly identifying an individual with Valid Effort at a 30% base rate. Finally, a cut-off score of > 16 achieved almost 30% sensitivity at almost 100% specificity in comparison to 21% sensitivity at almost 100% specificity for the Coding ACSS. The cut-score of > 16 was associated with 96% probability of correctly identifying Low Effort, and 77% likelihood of correctly identifying Valid Effort at a 30% base rate.

Moreover, the Coding Combination score demonstrated good predictive accuracy for the Moderate TBI group with the AUC score of 0.76, which was slightly higher than for the Coding ACSS. The analyses revealed that the cut-off score of > 14 is not appropriate for individuals with Moderate brain injuries due to inadequate specificity level. However, the cut-score of > 15 achieved 90% specificity and 40% sensitivity. Furthermore, at a 30% base rate, it was associated with 64% likelihood of correctly identifying Low Effort given that an individual obtains a score greater than 15, and 78% probability of correctly identifying Valid Effort given that an individual achieves a score less than 15. A score of > 16 was associated with almost 94% specificity and 33% sensitivity, which, once again was greater than sensitivity for the Coding ACSS at the same specificity level. Results of the present study also revealed that the cut-score of > 17 demonstrated 100% specificity with sensitivity at 20%. Due to the perfect specificity, the likelihood of identifying Low Effort was also 100%, and the probability of identifying Valid Effort was 74%, at 30% base rate.

Finally, the Coding Combination score showed an almost excellent predictive accuracy for the Severe and Very Severe TBI group with the AUC score of 0.88. Results also showed that cut-off scores of >14 and >15 would not be appropriate for this group, and should not be applied when assessing individuals with significant TBIs due to a high risk of making false positive errors. However, cut-scores of >16 achieved a very good sensitivity of 70% with associated specificity of 93%, and a cut-off score of >17 was associated with 96% specificity and lower sensitivity of 39%. In addition, the cut-score of >19 achieved almost perfect specificity of 99%, but low sensitivity of 13%. Results also showed that for the cut-score of >16 there was an 81% likelihood of correctly classifying Low Effort and 88% probability of identifying Valid Effort at a 30% base rate. At the same base rate, the cut-score of >17 was associated with reduced likelihood of 79% of correctly identifying Low Effort, and the same probability of 79% of correctly identifying Valid Effort.

Overall, the Coding Combination score showed a very good and excellent predictive accuracy for detecting Low Effort. As noted above, in respect of the Coding ACSS, it is important for clinicians to use appropriate cut-scores depending on injury severity to avoid bias against those with more significant TBIs. Furthermore, consistent with previous research showing that combined measures achieve higher sensitivity than an individual measure, results revealed that the Combination score achieved higher sensitivity levels than the Coding ACSS for the same specificity level (Whiteside et al., 2018; Whiteside, Gaasedelen, et al., 2015; Whiteside, Kogan, et al., 2015). This means that the Coding Combination score possesses a greater ability to detect individuals with Low Effort than the Coding ACSS. Because the Coding Combination score is effectively a combination of two embedded effort measures, these measures are derived from the same WAIS subtest. Clinicians should not rely on effort measures that are strongly correlated as individuals attempting to exaggerate their cognitive symptoms may do so in one domain but not others (Rosenfeld et al., 2000).

1.1. Strengths and Limitations

One of the strengths of this study is the large sample size (*n* = 650), which allows for increased generalisability. In addition, keeping with the current practice guidelines and previous research recommendations for the use of multiple effort measures to determine low effort (Cottingham & Boone, 2014; Larrabee, 2012; Martin et al., 2015; Victor et al., 2009), the present study employed failure on two or more PVTs strategy to establish the Low Effort group membership. Furthermore, consistent with the literature suggesting that some credible individuals with valid effort may still fail one PVT (Cottingham & Boone, 2014; Victor et al., 2009), the current study determined the Valid Effort group membership based on perfect performance (i.e. failure on none of the PVTs). These strategies reduce the possibility of false positive errors, thus allowing for more reliable conclusions.

In addition, the present research identified specific cut-scores for each injury severity group, thereby minimising false positive error rates, particularly for those with more significant brain injuries. Further, sensitivity and specificity rates were provided for each identified cut-score, as well as PPV and NPV, which are often said to be overlooked in research estimating predictive accuracy of testing instruments (O'Bryant & Lucas, 2006).

One of the limitations identified for the present study is the use of the archival sample of referrals to a private clinical neuropsychology practice. As such the sample may include a higher proportion of those with more significant traumatic brain injuries and those that are chronically disabled. However, these findings may possibly extend the generalisability due to the fact that a large part of research assessing effort and PVTs tend to include only individuals affected by mild TBI. Another limitation of the sample was a disproportionate representation of ethnicities such that the majority of the participants were of European origin, and only a minority were of Maori, Pacific Island, Indian and Asian ethnicities. Furthermore, the sample may also be overrepresented by those with the

Compensation-Seeking status; therefore caution should be taken if applying the conclusions to other populations.

The study employed a known-groups design, which has the advantage of clinical relevance due to the inclusion of a real-world clinical sample, however, it is recommended to use multiple designs when developing and validating PVTs, for example, in conjunction with the known-groups design the simulation design would potentially contribute information about the reliability and validity of these embedded PVTs (Heilbronner et al., 2009).

1.2. Future Research

To our knowledge, this is the first study fully investigating the WAIS Coding Error as an embedded validity indicator in a TBI sample. Thus, it is important to cross-validate these findings in order to expand or further support the potential of the Coding Error as an embedded validity indicators. As the Coding Error is a very quick and practical measure of effort, it would be beneficial to explore its efficacy to assess effort in a combination with other well-validated and commonly used PVTs based on other cognitive domains.

Furthermore, this is the first study exploring the Coding Combination score as an embedded validity indicator. Thus, once again, it is important to cross-validate these findings in order to expand or further support the use of the Coding Combination score in the assessment of effort during a neuropsychological evaluation. It would be beneficial to cross-validate these results not only in a traumatic brain injury sample but also to assess its efficacy in other populations. Furthermore, obtaining the Coding Combination score involves score reversal and also some minor additional calculations. This procedure adds potential for calculation error in clinical practice. Some exploration of the reliability of the technique in clinical practice would be appropriate.

1.3. Conclusion

It has been estimated that around 69 million people sustain traumatic brain injuries worldwide every year (Dewan et al., 2019), with around 36,000 suffering traumatic brain injury in New Zealand every year (Accident Compensation Corporation, 2018). Often following TBI, a neuropsychological assessment is conducted in order to determine the existence and the extent of the cognitive impairment, and if the individual is entitled to disability or insurance payments (Herrera-Guzmán et al., 2004; Webb et al., 2012).

However, conclusions, diagnosis, and treatment recommendations can only be made if the results are accurate and valid (Barker-Collo & Fernando, 2015; Lezak et al., 2012; Willis et al., 2011). Therefore, determinations about validity are said to be a primary step when interpreting data from a neuropsychological assessment (Bush et al., 2005). As a result, performance validity testing research is considered to be crucial and is currently one of the most dominant themes in the field of neuropsychology (Bigler, 2014).

Due to the importance of developing new performance validity tests and also validating the existing ones, the current study explored the efficacy of a number of embedded validity indicators derived from the WAIS Coding subtest, namely: Coding Errors; the Coding ACSS; and the Coding Combination score (i.e. arithmetic sum of Coding Error and Coding ACSS). This research was conducted with an aim to assist clinicians in identifying those individuals who may be feigning or exaggerating their cognitive impairment.

It was hypothesised that the WAIS Coding embedded validity indicators explored in this study (the Coding Error, the Coding ACSS, and the Coding Combination score) would be predictive of Low Effort. It was also hypothesised that the three Coding embedded validity indicators would bias against those with more significant brain injuries.

Findings indicated that all Coding embedded effort variables were significantly predictive of Low Effort. Furthermore, the Coding Error was found to be a highly specific measure of effort and also very time efficient, therefore valuable to practicing clinicians. The present study also found that injury severity has an impact on the WAIS Coding performance, therefore clinicians should always be cautious and use appropriate cut-scores based on the injury severity when using Coding embedded validity indicators.

Assuming the base rate of 30%, the recommended cut-off score for the Coding Error as an embedded validity indicator is: > 0 (97% specificity, 38% sensitivity). Based on the analysis of the Coding ACSS, the recommended cut-off scores for mTBI are: ≤ 5 (91% specificity, 56% sensitivity) and ≤ 4 (97% specificity, 40% sensitivity); for Moderate TBI: ≤ 4 (94% specificity, 33% sensitivity) and ≤ 3 (97% specificity, 27% sensitivity); for Severe and Very Severe TBI: ≤ 3 (94% specificity, 57% sensitivity). Furthermore, results revealed that the Coding Combination score achieved higher sensitivity than the Coding ACSS for the same level of specificity. Based on the findings in this study, the recommended cut-off scores for the Coding Combination score for mTBI are: > 14 (91% specificity, 63% sensitivity) and > 15 (97% specificity, 45% sensitivity); for Moderate

TBI: > 15 (90% specificity, 40% sensitivity) and > 16 (94% specificity, 33% sensitivity); for Severe and Very Severe TBI: > 16 (93% specificity, 70% sensitivity) and > 17 (96% specificity, 39% sensitivity).

Overall, the findings in this study show that the Coding embedded validity indicators have good potential as embedded validity indicators. However, due to low and variable sensitivity levels, they should be used in a combination with other PVTs in order to determine performance validity during neuropsychological evaluation and are likely to be used as complementary or confirmatory of other, more sensitive, standalone PVTs.

References

Accident Compensation Corporation. (2013). *Guidelines of the use of psychometric tests*. Auckland, NZ: ACC.

Accident Compensation Corporation. (2018). *Reducing traumatic brain injuries (TBI)*. Retrieved from https://www.acc.co.nz/preventing-injury/traumatic-brain-injury-tbi/

Alderman, N. (2003). Contemporary approaches to the management of irritability and aggression following traumatic brain injury. *Neuropsychological Rehabilitation, 13*(1-2), 211-240. https://doi.org/10.1080/09602010244000327

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *Clinical Neuropsychologist, 21*(6), 899-916. https://doi.org/10.1080/13825580600966391

Axelrod, B. N., Fichtenberg, N. L., Liethen, P. C., Czarnota, M. A., & Stucky, K. (2001). Performance characteristics of postacute traumatic brain injury patients on the WAIS-III and WMS-III. *Clinical Neuropsychologist, 15*(4), 516-520. https://doi.org/10.1076/clin.15.4.516.1884

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *Clinical Neuropsychologist, 20*(1), 145-159. https://doi.org/10.1080/13854040590947362

Bailey, K. C., Soble, J. R., & O'Rourke, J. J. F. (2018). Clinical utility of the Rey 15-Item Test, recognition trial, and error scores for detecting noncredible neuropsychological performance in a mixed clinical sample of veterans. *Clinical Neuropsychologist, 32*(1), 119-131. https://doi.org/10.1080/13854046.2017.1333151

Barker-Collo, S. L., & Fernando, K. (2015). A survey of New Zealand psychologists' practices with respect to the assessment of performance validity. *New Zealand Journal of Psychology, 44*(2), 35-42.

Bathard, H. (2016). *Symptom validity. Literature review update*: ACC.

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. https://doi.org/10.1017/S1355617705050277

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose-response relationship? *The Clinical Neuropsychologist, 20*, 831-847.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *Spine Journal, 5*(4), 404-417. https://doi.org/10.1016/j.spinee.2004.11.016

Bianchini, K. J., Mathias, W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *Clinical Neuropsychology, 15*(1), 19-45.

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*(13-14), 1623-1638. https://doi.org/10.3109/02699052.2014.947627

Bigler, E. D., & Maxwell, W. L. (2013). Overview of traumatic brain injury. In D. B. Arciniegas, N. D. Zasler, R. D. Vanderploeg, & M. S. Jaffee (Eds.), *Management of adults with traumatic brain injury* (pp. 3 - 32). Arlington, VA: American Psychiatric Publishing.

Binder, L. M. (1992). Deception and malingering. In A. Puente & R. J. McCaffrey (Eds.), *Handbook of neuropsychological assessment: A biopsychosocial approach*. New York, NY: Plenum.

Blyth, B. J., & Bazarian, J. J. (2010). Traumatic alterations in consciousness: Traumatic brain injury. *Emergency Medicine Clinics of North America, 28*(3), 571-594. https://doi.org/10.1016/j.emc.2010.03.003

Boone, K. B. (2007a). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY: Guildford Press.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *Clinical Neuropsychologist, 23*(4), 729-741. https://doi.org/10.1080/13854040802427803

Boone, K. B. (Ed.). (2007b). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York, NY: Guilford Press.

Boone, K. B., & Lu, P. (2003). Noncredible cognitive performance in the context of severe brain injury. *Clinical Neuropsychologist, 17*(2), 244-254. https://doi.org/10.1076/clin.17.2.244.16497

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The dot counting manual*. Los Angeles: Western Psychological Services.

Boone, K. B., & Lu, P. H. (1999). Impact of somatoform symptomatology on credibility of cognitive performance. *Clinical Neuropsychologist, 13*(4), 414-419.

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. https://doi.org/10.1076/jcen.24.5.561.1004

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. J., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity, NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419-426.

Carroll, L. J., Cassidy, J. D., Holm, L., Kraus, J., & Coronado, V. G. (2004). Methodological issues and research recommendations for mild traumatic brain injury: The WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine, Supplement* (43), 113-125.

Carroll, L. J., Cassidy, J. D., Peloso, P. M., Borg, J., von Holst, H., Holm, A., et al. (2004). Prognosis for mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitaton Medicine, 43*, 84-105.

Cassidy, J. D., Cancelliere, C., Carroll, L. J., Côté, P., Hincapié, C. A., Holm, L. W., et al. (2014). Systematic review of self-reported prognosis in adults after mild traumatic brain injury: Results of the international collaboration on mild traumatic brain injury prognosis. *Archives of Physical Medicine and Rehabilitation, 95*(3 SUPPL), S132-S151. https://doi.org/10.1016/j.apmr.2013.08.299

Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *Clinical Neuropsychologist, 22*(3), 529-546. https://doi.org/10.1080/13854040701346104

Chafetz, M. D. (2011). Reducing the probability of false positives in malingering detection of social security disability claimants. *Clinical Neuropsychologist, 25*(7), 1239-1252. https://doi.org/10.1080/13854046.2011.586785

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the social security disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*(1), 1-14. https://doi.org/10.1016/j.acn.2006.10.003

Chafetz L., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*(7), 633-639. https://doi.org/10.1093/arclin/act038

Cottingham, M. E., & Boone, K. B. (2014). Malingering in mild traumatic brain injury. In M. Sherer & A. M. Sander (Eds.), *Clinical handbooks in neuropsychology. Handbook on the neuropsychology of traumatic brain injury* (pp. 371-388). New York, NY, US: Springer Science.

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. H. M. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 257-264. https://doi.org/10.1080/13803395.2016.1223278

Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology, 28*(8), 771-783. https://doi.org/10.1093/arclin/act073

Dandachi-Fitzgerald, B., Ponds, R. W. H. M., Peters, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *Clinical Neuropsychologist, 25*(5), 812-828. https://doi.org/10.1080/13854046.2011.583280

Denning, J. H., & Shura, R. D. (2019). Cost of malingering mild traumatic brain injury-related cognitive deficits during compensation and pension evaluations in the veterans benefits administration. *Applied Neuropsychology:Adult, 26*(1), 1-16. https://doi.org/10.1080/23279095.2017.1350684

Dewan, M. C., Rattani, A., Gupta, S., Baticulon, R. E., Hung, Y. C., Punchak, M., et al. (2019). Estimating the global incidence of traumatic brain injury. *Journal of Neurosurgery, 130*(4), 1080-1097. https://doi.org/10.3171/2017.10.JNS17352

Dikmen, S. S., Corrigan, J. D., Levin, H. S., MacHamer, J., Stiers, W., & Weisskopf, M. G. (2009). Cognitive outcome following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 24*(6), 430-438. https://doi.org/10.1097/HTR.0b013e3181c133e9

Donders, J., & Strong, C. H. (2015). Clinical utility of the Wechsler Adult Intelligence Scale–Fourth Edition after traumatic brain injury. *Assessment, 22*(1), 17-22. https://doi.org/10.1177/1073191114530776

Egeland, J., Andersson, S., Sundseth, O. O., & Schanke, A. K. (2015). Types or modes of malingering? A confirmatory factor analysis of performance and symptom validity tests. *Applied Neuropsychology:Adult, 22*(3), 215-226. https://doi.org/10.1080/23279095.2014.910212

Erdodi, L. A., Abeare, C. A., Lichtenstein, J. D., Tyson, B. T., Kucharski, B., Zuccato, B. G., & Roth, R. M. (2017). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment, 29*(2), 148-157. https://doi.org/10.1037/pas0000319

Erdodi, L. A., Hurtubise, J. L., Charron, C., Dunn, A., Enache, A., McDermott, A., & Hirst, R. B. (2018). The D-KEFS trails as performance validity tests. *Psychological Assessment, 30*(8), 1082-1095. https://doi.org/10.1037/pas0000561

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: An emerging paradox of embedded validity indicators. *Clinical Neuropsychologist, 31*(6-7), 1029-1046. https://doi.org/10.1080/13854046.2017.1323119

Erdodi, L. A., Lichtenstein, J. D., Kucharski, B., Zuccato, B. G., Abeare, C. A., Tyson, B. T., & Roth, R. M. (2017). Wechler Adult Intelligence scale – Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychological Assessment, 29*(2), 148-157.

Erdodi, L. A., Nussbaum, S., Sagar, S., Abeare, C. A., & Schwartz, E. S. (2017). Limited English proficiency increases failure rates on performance validity tests with high verbal mediation. *Psychological Injury and Law, 10*(1), 96-103. https://doi.org/10.1007/s12207-017-9282-x

Essig, S. M., Mittenberg, W., Petersen, R. S., Stauman, S., & Cooper, J. T. (2001). Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology, 16*, 271-291.

Etherton, J. L., Bianchini, K. J., Heinly, M. T., & Greve, K. W. (2007). Pain, malingering, and performance on the WAIS-III Processing Speed Index. *Journal of Clinical and Experimental Neuropsychology, 28*(7), 1218-1237. https://doi.org/10.1080/13803390500346595

Feigin, V. L., Theadom, A., Barker-Collo, S., Starkey, N. J., McPherson, K., Kahan, M., et al. (2013). Incidence of traumatic brain injury in New Zealand: A population-based study. *The Lancet Neurology, 12*(1), 53-64. https://doi.org/10.1016/S1474-4422(12)70262-4

Fleminger, S. (2009). Head injury. In A. S. David, S. Fleminger, M. D. Kopelman, S. Lovestone, & J. D. S. Mellers (Eds.), *Lishman's organic psychiatry* (4th ed., pp. 167-279). Oxford, UK: Blackwell Publishing Ltd.

Freire, F. R., Coelho, F., Lacerda, J. R., Fernando da Silva, M., Goncalves, V. T., Machado, S., et al. (2011). Cognitive rehabilitation following traumatic brain injury. *Dementia & Neuropsychologia, 5*(1), 17-25.

Gallagher, M., McLeod, H. J., & McMillan, T. M. (2019). A systematic review of recommended modifications of CBT for people with cognitive impairments following brain injury. *Neuropsychological Rehabilitation, 29*(1), 1-21. https://doi.org/10.1080/09602011.2016.1258367

Gfeller, J. D., & Roskos, P. T. (2013). A comparison of insufficient effort rates, neuropsychological functioning, and neuropsychiatric symptom reporting in military veterans and civilians civilians with chronic traumatic brain injury. *Behavioral Sciences and the Law, 31*(6), 833-849. https://doi.org/10.1002/bsl.2084

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology, 44*(6), 1013-1023. https://doi.org/10.1002/1097-4679(198811)44:6<1013::AID-JCLP2270440627>3.0.CO;2-Z

Glassmire, D. M., Wood, M. E., Ta, M. T., Kinney, D. I., & Nitch, S. R. (2019). Examining false-positive rates of Wechsler Adult Intelligence Scale (WAIS-IV) processing speed-based embedded validity indicators among individuals with schizophrenia spectrum disorders. *Psychological Assessment, 31*(1), 120-125. https://doi.org/10.1037/pas0000650

Gould, K. R., Ponsford, J. L., Johnston, L., & Schönberger, M. (2011). The nature, frequency and course of psychiatric disorders in the first year after traumatic brain injury: A prospective study. *Psychological Medicine, 41*(10), 2099-2109. https://doi.org/10.1017/S003329171100033X

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America, 18*(1), 43-68. https://doi.org/10.1016/j.pmr.2006.11.002

Green, P., Allen , L., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-adminstered forms, US version 1.1*. Durham, NC: Cognisyst.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13*(10), 813-819. https://doi.org/10.1080/026990599121205

Green, P., & Merten, T. (2013). Noncredible explanations of noncredible performance on symptom validity tests. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering*. New York, NY: Springer Publishing Company, LLC.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen Iii, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*(12), 1045-1060. https://doi.org/10.1080/02699050110088254

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224. https://doi.org/10.1037/1040-3590.6.3.218

Greve, K. W., Bianchini, K., & Doane, B. (2006). Classification accuracy of the test of memory malingering in traumatic brain injury: Results of a known-groups analysis. *Journal of Clinical and Experimental Neuropsychology, 28*(7), 1176-1190. https://doi.org/10.1080/13803390500263550

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cut-offs on psychometric indicators of negative responce bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychologist, 19*(4), 533-541.

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2007). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *Clinical Neuropsychologist, 22*(5), 896-918. https://doi.org/10.1080/13854040701565208

Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain Referred referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation, 90*(7), 1117-1126. https://doi.org/10.1016/j.apmr.2009.01.018

Gronwall, D., & Wrightson, P. (1980). Duration of post-traumatic amnesia after mild head injury. *Journal of Clinical Neuropsychology, 2*(1), 51-60. https://doi.org/10.1080/01688638008403780

Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of psychological assessment* (6th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627-635.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology Consensus Conference statement on the neuropsychological assessment of effort, response bias, and malingering. *Clinical Neuropsychologist, 23*(7), 1093-1129. https://doi.org/10.1080/13854040903155063

Herrera-Guzmán, I., Peña-Casanova, J., Lara, J. P., Gudayol-Ferré, E., & Böhm, P. (2004). Influence of age, sex, and education on the Visual Object and Space Perception Battery (VOSP) in a healthy normal elderly population. *Clinical Neuropsychologist, 18*(3), 385-394. https://doi.org/10.1080/1385404049052421

Hyder, A. A., Wunderlich, C. A., Puvanachandra, P., Gururaj, G., & Kobusingye, A. A. (2007). The impact of traumatic brain injuries: A global perspective. *NeuroRehabilitation, 22*, 341 - 353.

Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering. A comparison of nine neuropsychological tests, four tests of malingering, and behavioural observations. *Archives of Clinical Neuropsychology, 17*, 1-23.

Inman, T. H., Vickery, C. D., Berry, D. T. R., Edwards, C. L., Lamb, D. G., & Smith, G. T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment, 10*(2), 128-139. https://doi.org/10.1037/1040-3590.10.2.128

Iverson, G. L. (2005). Outcome from mild traumatic brain injury. *Current Opinion in Psychiatry, 18*(3), 301-317. https://doi.org/10.1097/01.yco.0000165601.29047.ae

Kashluba, S., Hanks, R. A., Casey, J. E., & Millis, S. R. (2008). Neuropsychologic and functional outcome after complicated mild traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 89*(5), 904-911. https://doi.org/10.1016/j.apmr.2007.12.029

Kim, E., Lauterbach, E. C., Reeve, A., Arciniegas, D. B., Coburn, K. L., Mendez, M. F., et al. (2007). Neuropsychiatric complications of traumatic brain injury: A critical review of the literature (A report by the ANPA Committee on research). *Journal of Neuropsychiatry and Clinical Neurosciences, 19*(2), 106-127. https://doi.org/10.1176/jnp.2007.19.2.106

Kim, M. S., Boone, K. B., Victor, T., Marion, S. D., Amano, S., Cottingham, M. E., et al. (2010). The Warrington Recognition Memory Test for words as a measure of response bias: Total score and response time cutoffs developed on "real World" credible and noncredible subjects. *Archives of Clinical Neuropsychology, 25*(1), 60-70. https://doi.org/10.1093/arclin/acp088

Kim, N., Boone, K. B., Victor, T., Lu, P., Keatinge, C., & Mitchell, C. (2010). Sensitivity and specificity of a digit symbol recognition trial in the identification of response bias. *Archives of Clinical Neuropsychology, 25*(5), 420-428. https://doi.org/10.1093/arclin/acq040

Langlois, J. A., Kegler, S. R., Butler, J. A., Gotsch, K. E., Johnson, R. L., Reichard, A. A., et al. (2003). Traumatic brain injury-related hospital discharges. Results from a 14-state surveillance system, 1997. *MMWR. Surveillance summaries : Morbidity and mortality weekly report. Surveillance summaries/CDC, 52*(4), 1-20.

Larrabee, G. J. (2000). Neuropsychology in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology, 22*(5), 702-707.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clinical Neuropsychologist, 17*(3), 410-425. https://doi.org/10.1076/clin.17.3.410.18089

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York, NY, US: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625-631. https://doi.org/10.1017/S1355617712000240

Lathif, N., Phipps, E., Alton, P., & Sharma, D. T. (2014). Prevalence of psychiatric disorders following brain injury. *British Journal of Medical Practitioners, 7*(3).

Lee, T. T. C., Graham, J. R., Sellbom, M., & Gervais, R. O. (2012). Examining the potential for gender bias in the prediction of symptom validity test failure by MMPI-2 symptom validity scale scores. *Psychological Assessment, 24*(3), 618-627. https://doi.org/10.1037/a0026458

Levin, H. S., & Diaz-Arrastia, R. R. (2015). Diagnosis, prognosis, and clinical management of mild traumatic brain injury. *The Lancet Neurology, 14*(5), 506-517. https://doi.org/10.1016/S1474-4422(15)00002-2

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). Testing for effort, response bias, and malingering. In *Neuropsychological Assessment* (5th ed.). New York, NY: Oxford University Press, Inc.

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *Clinical Neuropsychologist, 32*(3), 391-421. https://doi.org/10.1080/13854046.2017.1406146

Loring, D. W., & Bauer, R. M. (2010). Testing the limits: Cautions and concerns regarding the new Wechsler IQ and Memory scales. *Neurology, 74*(8), 685-690. https://doi.org/10.1212/WNL.0b013e3181d0cd12

Love, C. M., Glassmire, D. M., Zanolini, S. J., & Wolf, A. (2014). Specificity and false positive rates of the test of memory malingering, Rey 15-Item Test, and Rey Word Recognition Test among forensic inpatients with intellectual disabilities. *Assessment, 21*(5), 618-627. https://doi.org/10.1177/1073191114528028

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists validity testing beliefs and practices: A survey of North American professionals. *Clinical Neuropsychologist, 29*(6), 741-776. https://doi.org/10.1080/13854046.2015.1087597

McAllister, T. W. (2008). Neurobehavioral sequelae of traumatic brain injury: Evaluation and management. *World Psychiatry, 7*(1), 3-10. https://doi.org/10.1002/j.2051-5545.2008.tb00139.x

McCaffrey, R. J., Pavlav, A. A., O'Bryant, S. E., & Labarge, A. S. (Eds.). (2003). *Critical issues in neuropsychology. Practitioner's guide to symptom base rates in clinical neuropsychology.* New York, NY: Kluwer Academic/Plenum Publishers.

McGee, J., Alekseeva, N., Chernyshev, O., & Minagar, A. (2016). Traumatic brain injury and behavior. A practical approach. *Neurologic Clinics, 34*(1), 55-68. https://doi.org/10.1016/j.ncl.2015.08.004

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*(1), 8-15. https://doi.org/10.1093/arclin/acq083

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychologist, 27*(1), 10-22.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094-1102.

Morse, C. L., Douglas-Newman, K., Mandel, S., & Swirsky-Sacchetti, T. (2013). Utility of the Rey-15 Recognition Trial to detect invalid performance in a forensic neuropsychological sample. *Clinical Neuropsychologist, 27*(8), 1395-1407. https://doi.org/10.1080/13854046.2013.832385

New Zealand Psychologists Board. (2015). *Guidelines for the use of psychometric tests*. Wellington, NZ: New Zealand Psychologist's Board.

O'Bryant, S. E., & Lucas, J. A. (2006). Estimating the predictive value of the Test of Memory Malingering: An illustrative example for clinicians. *Clinical Neuropsychologist, 20*(3), 533-540. https://doi.org/10.1080/13854040590967568

Olsen, D. H., Schroeder, R. W., Heinrichs, R. J., & Martin, P. K. (2019). Examination of optimal embedded PVTs within the BVMT-R in an outpatient clinical sample. *Clinical Neuropsychologist, 33*(4), 732-742. https://doi.org/10.1080/13854046.2018.1501096

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology:Adult, 25*(3), 219-230. https://doi.org/10.1080/23279095.2017.1278602

Riggio, S. (2011). Traumatic brain injury and its neurobehavioral sequelae. *Neurologic Clinics, 29*(1), 35-47. https://doi.org/10.1016/j.ncl.2010.10.008

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*(4), 349-359. https://doi.org/10.1016/S0887-6177(99)00025-6

Ruocco, A. C., Swirsky-Sacchetti, T., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *Clinical Neuropsychologist, 22*(3), 547-564. https://doi.org/10.1080/13854040701336444

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *Clinical Neuropsychologist, 25*(3), 437-453. https://doi.org/10.1080/13854046.2011.556668

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable Digit Span: A systematic review and cross-validation study. *Assessment, 19*(1), 21-30. https://doi.org/10.1177/1073191111428764

Shores, E. A., Marosszeky, J. E., Sandanam, J., & Batchelor, J. (1986). Preliminary validation of a clinical scale for measuring the duration of post-traumatic amnesia. *Medical Journal of Australia, 144*(11), 569-572.

Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering*. New York, NY: Springer Publishing Company, LLC.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist, 13*(4), 545-561. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research, 157*(1-3), 191-200. https://doi.org/10.1016/j.psychres.2007.01.003

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

Tate, R. L. (2012). Traumatic brain injury. In P. Kennedy (Ed.), *The Oxford Handbook of Rehabilitation Psychology* (pp. 248-272). New York, NY: Oxford University Press, Inc.

Teasdale, G. M. (1995). Head injury. *Journal of Neurology, Neurosurgery, and Psychiatry, 58*(5), 526-539. https://doi.org/10.1136/jnnp.58.5.526

Teasdale, G. M., & Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *The Lancet, 304*(7872), 81-84. https://doi.org/10.1016/S0140-6736(74)91639-0

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, NY: Multi Health Systems.

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology, 16*(4), 597-607. https://doi.org/10.1080/01688639408402671

van Gorp, W. G., Humphrey, L. A., Kalechstein, A., Brumm, V. L., McMullen, W. J., Stoddard, M., & Pachana, N. A. (1999). How well do standard clinical neuropsychological tests identify malingering? A prelimenary analysis. *Journal of Clinical and Experimental Neuropsychology, 21*(2), 245-250.

Victor, T., L.,, Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *Clinical Neuropsychology, 23*(2), 297-313.

Webb, J. W., Batchelor, J., Meares, S., Taylor, A., & Marsh, N. V. (2012). Effort test failure: Toward a predictive model. *The Clinical Neuropsychologist, 26*(8), 1377-1396. https://doi.org/10.1080/13854046.2012.728248

Webber, T. A., & Soble, J. R. (2018). Utility of various WAIS-IV Digit Span indices for identifying noncredible performance validity among cognitively impaired and unimpaired examinees. *Clinical Neuropsychologist, 32*(4), 657-670. https://doi.org/10.1080/13854046.2017.1415374

Wechsler, D., & Psychological, C. (1997). *WAIS-III: Administration and scoring manual: Wechsler Adult Intelligence Scale*. [San Antonio, Tex]: Psychological Corporation.

Wechsler, D., Psychological, C., & PsychCorp. (2008). *WAIS-IV technical and interpretive manual*. San Antonio, Tex.: Pearson.

West, L. K., Curtis, K. L., Greve, K. W., & Bianchini, K. J. (2011). Memory in traumatic brain injury: The effects of injury severity and effort on the Wechsler Memory Scale-III. *Journal of Neuropsychology, 5*(1), 114-125. https://doi.org/10.1348/174866410X521434

Wetter, M. W., & Corrigan, S. K. (1995). Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. *Professional Psychology: Research and Practice, 26*, 474-477.

Whiteside, D. M., Caraher, K., Hahn-Ketter, A., Gaasedelen, O., & Basso, M. R. (2018). Classification accuracy of individual and combined executive functioning embedded performance validity measures in mild traumatic brain injury. *Applied Neuropsychology: Adult*, 1-10. https://doi.org/10.1080/23279095.2018.1443935

Whiteside, D. M., Gaasedelen, O. J., Hahn-Ketter, A. E., Luu, H., Miller, M. L., Persinger, V., et al. (2015). Derivation of a cross-domain embedded performance validity measure in traumatic brain injury. *Clinical Neuropsychologist, 29*(6), 788-803. https://doi.org/10.1080/13854046.2015.1093660

Whiteside, D. M., Kogan, J., Wardin, L., Phillips, D., Franzwa, M. G., Rice, L., et al. (2015). Language-based embedded performance validity measures in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 37*(2), 220-227. https://doi.org/10.1080/13803395.2014.1002758

Williams, D. H., Levin, H. S., & Eisenberg, H. M. (1990). Mild head injury classification. *Neurosurgery, 27*(3), 422-428. https://doi.org/10.1227/00006123-199009000-00014

Willis, P. F., Farrer, T. J., & Bigler, E. D. (2011). Are effort measures sensitive to cognitive impairment? *Military Medicine, 176*(12), 1426-1431.

World Health Organization. (2006). *Neurological disorders: Public health challenges*. Retrieved from https://www.who.int/mental_health/neurology/neurological_disorders_report_web.pdf

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law, 8*(3), 188-199.

Zasler, N. D., & Bender, S. D. (2019). Validity assessment in traumatic brain injury impairment and disability evaluations. *Physical Medicine and Rehabilitation Clinics of North America, 30*(3), 621-636. https://doi.org/10.1016/j.pmr.2019.03.009

Appendices

Appendix A: Letter of Ethical Approval AUT University Ethics Committee (AUTEC)

TE WĀNANGA ARONUI
O TĀMAKI MAKAU RAU

Auckland University of Technology Ethics Committee (AUTEC)

Auckland University of Technology
D-88, Private Bag 92006, Auckland 1142, NZ
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

7 May 2019

Susan Mahon
Faculty of Health and Environmental Sciences

Dear Susan

Re Ethics Application: **19/107 Analysis of WAIS 111 and WAIS IV Coding subsets as an effective embedded effort measure**

Thank you for providing evidence as requested, which satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC).

Your ethics application has been approved for three years until 7 May 2022.

Standard Conditions of Approval

1. A progress report is due annually on the anniversary of the approval date, using form EA2, which is available online through http://www.aut.ac.nz/research/researchethics.
2. A final report is due at the expiration of the approval period, or, upon completion of project, using form EA3, which is available online through http://www.aut.ac.nz/research/researchethics.
3. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form: http://www.aut.ac.nz/research/researchethics.
4. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
5. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.

Please quote the application number and title on all future correspondence related to this project.

AUTEC grants ethical approval only. If you require management approval for access for your research from another institution or organisation then you are responsible for obtaining it. You are reminded that it is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard.

For any enquiries, please contact ethics@aut.ac.nz

Yours sincerely,

Kate O'Connor
Executive Manager
Auckland University of Technology Ethics Committee

Cc: akrievina@googlemail.com

Appendix B: Participant Consent Form provided by Webb Psychology Ltd

Client Registration and Consent Form

Full name *eg Jonathan Smith*		Call me... *eg Johnny*	
Mailing Address		Telephone	Home Work Mobile
Email			
Date of Birth		Ethnicity	
Current Age		Marital status	
Occupation		Partner's name	
OK to leave a message for you at (circle):		Home Work	Mobile With Partner
Medication (include dose if known)			

For ACC clients:
I understand that information gathered may be shared with the ACC (I have signed an ACC Consent Form (ACC167/ACC6300 – this is an ACC authority for the collection and disclosure of information/authority to collect medical and other information).

For all clients:
Please sign below if you consent to undertake psychological assessment and/or treatment.
This involves identification of current problems and concerns, gathering relevant personal information, goal-setting and treatment planning. Psychological questionnaires and cognitive tests may be involved and psychometric data may be collected. During any cognitive testing it is important that you give your best effort to the tasks and your level of effort will be formally assessed.

- I understand that I have the right to a support person when receiving these services.
- I understand that I have the right to make complaint about services and without penalty.
- I understand that I can request access to my clinical information held by Webb Psychology and correct any factual information by contacting the relevant health professional.
- I consent to Webb Psychology storing and using my **de-identified** health information for research purposes so long as complete confidentiality and anonymity is maintained.

Signed: ……………………………………………………….. **Date:** ………………………

Webb Psychology Ltd • 176 Green Lane West Greenlane Auckland NZ • Tel: +64 9 522 5083 • admin@webbpsychology.co.nz • www.webbpsychology.co.nz