Measurement invariance testing of the MMPI-2 when used with patients suffering a traumatic brain injury

Nathan Raoul Alkemade

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

January, 2013

Melbourne School of Psychological Sciences
The Faculty of Medicine, Dentistry and Health Sciences

Abstract

The MMPI-2 is one of the most widely used tests of personality and psychopathology in both clinical and research settings (Archer & Newsom, 2000; Butcher, Rouse & Steven, 1996; Smith, Gorske, Wiggins & Little, 2010). It is suggested the neurological damage from a traumatic brain injury (TBI) can falsely inflate MMPI-2 profiles. Following this theory, the Gass (1991) correction procedure removes 14 items from the MMPI-2. Widespread use of the correction procedure continues despite conflicting results from replication studies. In this study measurement invariance analysis was completed separately on MMPI-2 scales Hs1, Hy3 and Sc8 to assess the Gass correction procedure. A TBI sample (n=254) and a sample generated from the MMPI-2 normative data (n=2600) was used for measurement invariance testing. In measurement invariance Test 1 (baseline model test) all residuals and one item loading for each factor were held to equality. In Test 2 (strict invariance test), all parameters were held to equality across groups. The requirement of invariance is that CFI decrease by equal to or less than .002 (Meade, Johnson and Braddy, 2008). If a model failed the test of strict invariance then a partial invariance model was defined using the backwards elimination procedure. Practical impact analysis was completed using the Millsap and Kwok (2004) procedure to assess the clinical effect from a failure to establish strict invariance.

Prior to measurement invariance testing, exploratory factor analysis and confirmatory factor analysis were employed to define a factor model in Hs1, Hy3 and Sc8. A 4-Factor model was selected as best representing the 32 items from Hs1. In Test 1 the model produced reasonable fit indices (RMSEA = .023, CFI = .947, TLI = .943). In Test 2 the decrease in CFI was above the threshold of invariance (RMSEA = .025, CFI = .932, TLI = .930). A partial invariance model was defined with the parameters for four items freed (RMSEA = .023, CFI = .946, TLI = .945). All items passed the tests of no practical impact. In Hy3 a 4-Factor model was selected as best representing the 40 items from this scale not previously analysed in Hs1. In Test 1 the model produced reasonable fit indices (RMSEA .026, CFI = .921, TLI =.915). In Test 2 the decrease in CFI was above the threshold of invariance (RMSEA .043, CFI = .767, TLI =.761). A partial invariance model was defined with the parameters for 20 items freed (RMSEA .026, CFI = .919, TLI =.915). Items 161 and 185 failed the tests of no practical impact. In Sc8 a 5-factor model was selected as best representing the 68 items not previously analysed in Hs1 or Hy3. In Test 1

the model produced reasonable fit indices (RMSEA .015, CFI = .934, TLI = .932). In Test 2 the decrease in CFI was above the threshold of invariance (RMSEA .023, CFI = .838, TLI = .837). A partial invariance model was defined with the parameters for 28 items freed (RMSEA .015, CFI = .932, TLI = .930). Items 17, 92, 190, 278, 281, 291 and 303 failed the tests of no practical impact. Eleven of the 14 items from the Gass correction procedure passed the test of strict invariance, with the other three passing the tests of no practical impact. This finding fails to support continued use of the Gass correction procedure. Additionally the finding is contrary to the hypothesis that neurological content will bias MMPI-2 profiles in specific populations, such as the traumatic brain injury. However, some items were failed the tests of no practical impact and were identified as concerning. The implications from these findings are discussed.

Declaration

This is to certify that

i) the thesis comprises only my original work towards the degree of Master of Psychology (Clinical Psychology)/Doctor of Philosophy except where indicated in the Preface,

ii) due acknowledgement has been made in the text to all other material used,

iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee.

Date / /

Acknowledgements

Firstly thank you to my supervisor Associate Professor Stephen Bowden. Without his assistance this dissertation would not have been completed. There were the countless responses to my queries, answers to frantic phone calls, and replies to borderline incoherent emails that were always dealt with professionally and respectfully. There were many times during the process of completing this dissertation that Stephen's calming confidence and insightful guidance were crucial. Finally, the respect I received from Stephen, whether it was discussing my opinions on topics or my occasional disagreement over the best analytic approach, gave me much confidence in my capabilities as a fledgling researcher. Thank you for everything Stephen.

Secondly, thank you to my wife Jenny and my son Louis. You both provided me plenty of time to selfishly work on my thesis. Also you both gave me a reminder of what are the truly important things in life. I will always remember bowing my head because some analysis went awry, only to see Louis on the ground smiling up at me. Suddenly data analysis was irrelevant in my world, and consequently less stressful. Jenny, a walk down the park or a cup of coffee with you was the boost I often needed, thanks. I also thank you for the many discussions, often about topics with which you were initially unfamiliar. You were consistently a source of invaluable feedback.

Finally to my student cohort, a big thanks. To be able to share this experience with a group of dedicated, intelligent and passionate people was one of the true joys of the experience.

Table of Contents

Table of Figures

Table of Tables

Overview of Study

The devastating effects of traumatic brain injury (TBI) to the individual, their loved ones and the costs to the community are for the most part unrecognised. In Australia 2% of the population live with a TBI which impedes their daily functioning, and worldwide TBI is the leading cause of disability in people under the age of 40 (Brain Injury Association of Queensland, 2006; WHO, 2006). In addition to the physical, emotional, occupational and cognitive challenges' faced by a person with a TBI, the literature also points to a marked increase in the risk of psychopathology in people with a TBI. Research suggests suffering a TBI is associated with an increased risk for diagnosis of various personality disorders, depressive disorders, anxiety disorders, substance disorders and psychotic disorders (Arciniegas, Harris, & Brousseau, 2003; Koponen et al., 2002; Silver, Kramer, Greenwald, & Weissman, 2001).

Importantly for individuals with a TBI, therapy has been shown to successfully treat psychological disorders and alleviate the burden of suffering (Dawson, Schwartz, Winocur, & Stuss, 2007; Folzer, 2001; Seel et al., 2003). To best utilise limited resources, understanding the mechanisms which underlie symptom expression is necessary to apply appropriate therapy. One of the principal challenges facing clinicians working with the TBI population is disentangling the neurological symptoms from the psychological. Of particular concern is that neurological symptoms may be mistakenly interpreted as psychological. This concern has led to the proposal that some diagnostic instruments require alterations when used with the TBI population (Alfano, Paniak, & Finlayson, 1993; Gass, 1991). Both Gass and Alfano et al. propose removing items from the Minnesota Multiphasic Personality Inventory 2nd edition (MMPI-2). The Gass (1991) correction procedure will be discussed in more detail in Chapter Two.

The MMPI-2 is one of the most frequently used measures of personality and psychopathology in both clinical and research environments (Archer & Newsom, 2000; Butcher, Rouse & Steven, 1996; Smith, Gorske, Wiggins & Little, 2010). The MMPI-2 comprises a 567 item inventory which requires True or False response to statements. The items were selected using a criterion-keying approach which importantly requires no theoretical basis for item selection. This procedure resulted in the inclusion of items which may be interpreted as referring to diverse conditions, including neurologic abnormalities. The inclusion of such items has led to the

suggestion that a person suffering a TBI may endorse these items due to the neurological symptoms of the injury and not psychopathology that is the primary target of assessment. This alternate reason for endorsement appears a reasonable proposition and has led to the generation of various correction procedures. Correction procedures have been proposed for persons suffering multiple sclerosis, cerebrovascular disease, spinal cord injury (Barncord & Wanlass, 2000; Gass, 1992; Kendall, Edinger, & Eberly, 1978; Meyerink, Reitan, & Selz, 1988). However the Gass (1991) correction procedure proposed for persons suffering a TBI is the most widely adopted and is the focus of this thesis.

Despite the widespread use of the correction procedure there are researchers who question the applicability of this approach. The questions fall into two broad categories, theoretical and empirical. Empirically the validity studies of the Gass (1991) correction procedure have produced conflicting results. Some results support the correction procedure (Gass & Wald, 1997; Rayls, Mittenberg, William, & Theroux, 1997). In contrast other studies are less supportive and query whether the correction procedure produces any incremental validity in the MMPI-2 (Arbisi & Ben-Porath, 1999; Brulot, Strauss, & Spellacy, 1997; Edwards et al., 2003; Glassmire et al., 2003). Additionally the hypothesis that neurological symptoms of a TBI will reduce the validity of a MMPI-2 assessment has many critics (Glassmire et al., 2003; La Chapelle & Alfano, 2005). Despite these criticisms the Gass (1991) correction procedure is included in some contemporary clinical assessment guides (Butcher, 2006). These issues are discussed in greater detail in Chapter Two.

Whether the MMPI-2 in its uncorrected format can be used to diagnose psychopathology in patients' with a TBI is a question that falls under what Widaman and Reise (1997) call the 'rubric of invariance testing'. Measurement invariance is discussed in Chapter Three. Measurement invariance evaluates the precise quantative generalisability of construct validity inferences and is essential in determining whether a psychological test is appropriate for use amongst different groups (Brown, 2006). Therefore, by completing tests of measurement invariance we can assess the appropriateness of the uncorrected MMPI-2 when diagnosing psychopathology in persons suffering a TBI. By extension, measurement invariance permits a test of the Gass (1991) correction procedure using modern statistical techniques. Chapter Three will also provide an overview of the theory underpinning measurement invariance tests undertaken to assess the Gass (1991) correction procedure.

It is important to understand the various conceptual explanations that are provided to understand the relationship between TBI and psychopathology. Understanding the nuances between a neurological and a psychosocial explanation for the increased prevalence of psychopathology after a TBI, helps better identify the strengths and weaknesses of theory underpinning the Gass (1991) correction procedure. Additionally recognising the various forms of psychopathology associated with TBI, along with the enormous economic and personal costs from injury make it clear that determining the appropriateness of Gass (1991) correction procedure is essential. Even more so when one considers that accurate diagnosis is required for timely therapy to treat psychopathology. Timely therapy can minimise the impact from a TBI on both the individual and the community. The Stress-Appraisal-Coping model is provided as a valuable clinical guide to understand psychopathology in persons suffering a TBI. These issues are covered in Chapter One. Details about the relationship between neurological damage and psychopathology in persons with a TBI are discussed. The tension between the research into severity of TBI and psychopathology with the theory underpinning the Gass (1991) correction procedure is highlighted in Chapter Two.

Chapter One – Traumatic Brain Injury

This section will discuss research into the relationship between TBI and psychopathology. The proposed aetiology of psychopathology subsequent to TBI is covered, which includes a suggestion for the role of both neurologic and psychosocial features. The risk factors for developing psychopathology after a TBI, and the risk factors for having a TBI will be examined. The interdependent relationship between psychopathology and TBI is highlighted. The surprising inverse relationship between TBI severity and levels of psychopathology is reviewed. As will be seen in Chapter Two this relationship is also found in MMPI-2 profiles of TBI patients. Finally the difficulty of assessing psychopathology with patients who have suffered a TBI is discussed.

1.1 Prevalence of TBI

It is important to highlight the impact of TBI on the community. As previously stated WHO (2006) define TBI as the leading cause of disability in people under the age of 40. Additionally WHO predict that due to factors such as the increasing numbers of vehicles on the roads, motor vehicle accidents by 2020 will sit third on the world ranking of burden of disease. With motor vehicle accidents being implicated in almost two-thirds of TBI's in Australia (Brain Injury Association of Queensland, 2006) the WHO predictions suggest an expected increase in the number of TBI's and cost to the community.

In Australia there were over 16,000 hospitalisations during 2008, the last year for which detailed results are available, that involved TBI's (Access Economics Pty Ltd, 2009). It is estimated that in Australia the total cost of TBI is $8.6 billion with $3.7 billion attributable to moderate TBI and $4.8 billion attributable to severe TBI, with costs apportioned to individuals (64.9%), State (19.1%) and Federal Governments (11.2%) (Access Economics Pty Ltd, 2009). These figures exclude costs associated with mild TBI's which account for 70-80% of cases (Rapoport, McCauley, Levin, Song, & Feinstein, 2002). As mild TBI's are associated with higher rates of psychopathology including this group would lead to substantially increased costs (Crowe, 2008; Lillie et al., 2010; Rosenthal, Christensen, & Ross, 1998).

In addition to the challenges faced by someone who suffers a TBI during the acute injury phase, there are substantial long term difficulties. Cameron, Purdie, Kliewer and McClure (2008)

compared a non-injured cohort with a TBI group and found after excluding the first 60 days the TBI group, regardless of injury severity, had a higher mortality rate even after adjusting for pre-existing health and demographic characteristics. Cameron et al. also found the TBI group had more post-injury hospitalizations and greater cumulative lengths of stay. The protracted nature of the consequences from a TBI highlights the importance of rehabilitation programmes, especially as two thirds of the people who suffer a TBI are under the age of 30 (Access Economics Pty Ltd, 2009). TBI is the leading cause of death and disability in people under the age of 45 (Crowe, 2008; Dombrowski, Petrick, & Strauss, 2000).

In summary the statistics show that living with the effects of TBI is an ongoing burden to large numbers of people. The financial costs to the Australian community are at a minimum just below $10 billion and likely much higher. With many of the injured being young at the time of the incident they are required to endure the life-altering consequences from a TBI for many years. Ongoing treatment for medical and psychological issues along with diminished capacity to independently complete the activities required for daily living including community involvement, education and work, are some of the more debilitating consequences.

1.2 Psychopathology and TBI

In this section the literature covering the comorbidity of TBI and psychopathology is reviewed. Overwhelmingly the research finds higher rates of psychological disorders and psychological distress in groups of individuals who suffered a TBI compared with community levels. These higher prevalence rates continue to persist decades after injury. The papers reviewed find higher rates of both Axis I and Axis II disorders in TBI populations. The increased rates indicate the need to assess and treat each person with a TBI carefully and being mindful of the diagnostic risk associated with interpreting symptomology as purely neurological in aetiology. As Palav, Ortega and McCaffrey (2001) suggested, psychological functioning should be a standard element of assessment after TBI to inform treatment. Note that while psychopathology and distress are interrelated constructs the research finds them to be distinct entities (Coyne & Schwenk, 1997; Fechner-Bates, Coyne, & Schwenk, 1994). In this dissertation ‘distress’ is used as generic shorthand for many different types of psychopathology.

Overview

Research shows that persons with a TBI are at increased risk for a variety of psychopathologies and this risk continues long after injury. Koponen et al. (2002) assessed 60 patients with a TBI who on average were injured 30 years earlier. They found 48.3% of people with a TBI met criteria for an Axis I disorder that began after the injury and 61.7% for an Axis II disorder during their lifetime. The most common Axis I disorders were major depression, substance use or dependence, panic disorder, specific phobia and psychotic disorders. Axis II personality disorders were more commonly avoidant, paranoid and schizoid. This study exemplifies both the variety of psychopathology associated with TBI and the enduring nature of these pathologies. Other studies have also found the prospect of developing psychopathology remains elevated many years, even decades after the injury, as seen by a study of World War 2 veterans which found TBI produced an increased risk for developing depression decades after the injury (Holsinger et al., 2002).

The research literature on Axis I disorders comorbid with a TBI is largely supportive of the Koponen et al. (2002) results, and will be discussed in more detail later when reviewing specific psychopathologies. However, research into Axis II personality disorders is less clear. Hibbard et al. (2000) assessed Axis II disorders using Diagnostic and Statistical Manual (DSM) Fourth Edition criteria. In a sample of 100 TBI patients they found one-quarter met criteria for a personality disorder prior to injury, and two-thirds of the sample met criteria for at least one personality disorder after injury. Pre-injury the most common personality disorders were antisocial and obsessive-compulsive, whilst after injury the most common were borderline, narcissistic, avoidant, paranoid and obsessive-compulsive. Koponen et al. (2002) in conjunction with Hibbard et al. (2000) find almost all personality disorders are implicated as sequelae to a TBI. These findings likely occur because there is no homogeneous effect of TBI on personality style. Rather, persons suffering a TBI are exposed to an environment with continual challenges and individuals can apply a variety of compensatory strategies and, when maladaptive, these strategies lead to various personality disorders (Hibbard et al., 2000).

It may be that people who are prone to TBI are also more likely to suffer from psychopathology. Interestingly we will see later that psychological disorders are indeed a risk factor for suffering a TBI, especially depression and substance use disorders (Hibbard, Uysal, Kepler, Bogdany, & Silver, 1998). However, Cameron et al. (2008) found that even after controlling for pre-injury

psychopathology in the 10 years after their injury a TBI sample had 2.82 times the number of mental health hospitalizations than a sample matched by aboriginal status, age, gender and geographical location. The strength of this paper in minimising potential confounds, provides strong evidence for the increased risk of psychopathology after a TBI, regardless of pre-injury factors. The results are supported by Silver et al. (2001) who interviewed 5034 participants of which 361 recorded a history of severe brain trauma. After controlling for sociodemographic factors and alcohol use, the risk for major depression, anxiety disorders and drug abuse/dependence was still increased after a TBI. Gualtiera and Johnson (1999) identify that TBI's have the tendency to worsen any pre-existing negative traits. Therefore, pre-existing psychopathology does not preclude the injury leading to further reduced psychological functioning.

Many people who suffer a TBI are burdened daily by irritability, cognitive fatigue and memory dysfunction with the complication of emotional difficulties reducing the quality of life (Rapoport et al., 2002; Whitnall, McMillan, Murray, & Teasdale, 2006). TBI survivors with comorbid psychopathology report more severe symptomology, perceive their injury and cognitive impairment as more severe, are more functionally disabled, and make comparatively poorer recoveries than TBI survivors without psychopathology (Fann, Katon, Uomoto, & Esselman, 1995; Mooney & Speed, 2001). Of particular concern is the risk of suicidal behaviour. Studies have found that persons with psychopathology are at increased risk for attempted suicide when their illness is comorbid with a TBI (Alderfer, Arciniegas, & Silver, 2005; Silver et al., 2001). TBI comorbid with psychopathology appears to produce an additive cumulative impact on the sufferer when compared with persons suffering psychopathology or TBI alone.

Research on specific psychological disorders and TBI

Anxiety disorders are common sequelae of a TBI. Mooney and Speed (2001) found 24% of their participants with a mild TBI were classified as having developed an acquired anxiety disorder. Similarly whilst reviewing 1199 patients from 1942-1990 Epstein and Ursano (1994) discovered approximately 29% of individuals with a TBI were diagnosed with clinical anxiety subsequent to the injury. Disorders such as Obsessive Compulsive Disorder, Generalised Anxiety Disorders, Panic Disorder, and Post Traumatic Stress Disorder have all been found to have higher prevalence rates in the TBI population compared with the general population (Moore, Terryberry-Spohr, & Hope, 2006).

Hiott and Labbate (2002) reviewed the literature surrounding TBI and anxiety disorders and reported that a meta-analysis of Generalised Anxiety Disorder found prevalence rates were double those found in community samples. Similar levels of Generalised Anxiety Disorder in TBI samples have been found in other studies (Bachna, Sieggreen, Cermak, Penk, & O'Connor, 1998). Hibbard, Uysal, Kepler, Bogdany, & Silver (1998) compared the prevalence of psychopathology after TBI across two groups, those with and without a history of psychopathology. They found a prevalence rate for Obsessive Compulsive Disorder of 14% in both groups. This prevalence rate is higher than community rates and similar to the rates found with other studies using a TBI sample (Hiott & Labbate, 2002). The research consistently finds Generalised Anxiety Disorder and Obsessive Compulsive Disorder at higher rates in TBI samples than is found in the community.

Understanding the relationship between TBI and Post Traumatic Stress Disorder is complicated. Not least by diagnostic DSM-IV-TR criterion A2 that requires response to the traumatic event involves intense fear, horror, or helplessness (American Psychiatric Association, 2000). Criterion A2 can be problematic for a TBI patient who suffers amnesia. Recent studies are finding a large number of persons, both with and without a TBI, who develop Post Traumatic Stress Disorder symptoms fail to endorse criterion A2 (O'Donnell, Creamer, McFarlane, Silove, & Bryant, 2010). The authors note this finding supports Post Traumatic Stress Disorder being a valid psychopathology in a traumatic brain injured person regardless of loss of consciousness. The removal of criterion A2 in DSM-5 may lead to a clearer understanding of the relationship between TBI and Post Traumatic Stress Disorder.

Furthermore, the presence of trauma in TBI makes comparison with community prevalence rates difficult to interpret. The most common incident leading to a TBI is a motor vehicle accident (Brain Injury Association of Queensland, 2006). Ursano et al. (1999) found victims of serious motor vehicle accidents were more likely to report Post Traumatic Stress Disorder symptoms than those involved in minor accidents, which supports the notion that the severity of the accident is related to the psychopathology. However, Ursano and colleagues failed to control for the presence or absence of a TBI in their estimates of Post Traumatic Stress Disorder symptoms. Therefore it is difficult to determine whether a TBI is specifically implicated in the higher rate of Post Traumatic Stress Disorder or the severity of the motor vehicle accident is the primary determinant of Post Traumatic Stress Disorder. Bryant and Harvey (1999) found comparable

rates of Post Traumatic Stress Disorder in TBI and non-TBI samples of motor vehicle accident victims, suggesting Post Traumatic Stress Disorder may be more related to the incident causing the TBI rather than the injury itself.

Individuals with comorbid depressive and anxiety disorders are regularly found in mental health settings. It is therefore not surprising that this scenario is replicated amongst survivors of TBI. Jorge et al. (2004) found 76.7% of TBI sufferers who presented with Major Depression also met criteria for an anxiety disorder. Furthermore, Jorge et al. observed that one-third of the TBI patients suffered Major Depression inside one year after injury. Continuing on this topic, studies that investigate the relationship between depression and TBI will now be reviewed. Depressive disorders are the most common form of psychopathology comorbid with a TBI. In an attempt to understand the role of TBI severity and depression Busch and Alpern (1998) critically reviewed a dozen research papers from 1984 through 1996. They found consistent evidence for high rates of depressive disorders in TBI samples. The studies reviewed revealed prevalence rates from 35% up to 87%. Additionally Busch and Alpern found the disorders occurred many years after the injury and not just as an acute response to the trauma.

These findings are consistently replicated in the literature with Kreutzer, Seel and Gourley (2001) finding 42% of 722 individuals referred to a regional trauma centre met criteria for a DSM-IV diagnosis of Major Depressive Disorder. Sell et al. (2003) found 27% of patients from a demographically diverse sample of 666 outpatients who were evaluated from one to ten years after injury also met DSM-IV criteria for a Major Depressive Disorder. It is clear from the research that an individual who suffers a TBI is at increased risk of developing major depression in both the short and long term subsequent to the injury.

With TBI's often resulting from a motor vehicle accident, and younger males over-represented in these accidents, it is worth considering whether age at time of injury is a specific risk factor for psychological distress after a TBI (Brain Injury Association of Queensland, 2006). Rapoport, Kiss and Feinstein (2006) found that in a sample of 77 patients over 50 years of age who attended a TBI clinic, major depression occurred at incidence levels higher than community prevalence rates. This finding suggests the injury rather than the age of the injured person is related to the experience of psychological distress. Furthermore TBI patients with comorbid depression, when compared with those patients without, reported elevated degrees of

psychological distress and psychosocial dysfunction, and were rated as having inferior performance in activities of daily living (Rapoport et al., 2006). Therefore the challenge of living with a TBI is related to psychological distress regardless of a person's age.

There is much less research into the development of psychotic disorders, such as schizophrenia, after a TBI. This despite the proposition by Kraeplin (1919 in Corcoran, McAllister, & Malaspina, 2005) that childhood TBI is a risk factor for developing psychosis. Until recently the literature was predominantly case studies. In a literature review, Fujii (2005) found between 6.5% and 43.5% of patients with schizophrenia had suffered a TBI. He concluded that TBI may contribute to a high proportion of cases with schizophrenia but it is unclear from the studies whether the psychopathology predates injury.

However there is some evidence to suggest that individuals who suffer a TBI are at increased risk for developing psychosis. Davison and Bagley (1969) reviewed the research into psychosis after a TBI and discovered prevalence rates ranging between 0.07% and 9.8%, with the psychosis often developing many years after injury. The higher end of the range is consistent with TBI being a risk factor for psychosis whereas the lower end of the range suggests no additional risk.

More recent research is unable to provide further clarity. Nielsen, Mortensen, O'Callaghan, Mors and Ewald (2002) analysed head injury rates and severity in 5179 male and 3109 female patients with schizophrenia. The findings were compared with a gender and age-matched control sample. They found higher rates of head injury in males, but not females. Nielsen and colleagues conservatively concluded that the findings do not exclude the possibility that for some males, head injury may contribute to risk of schizophrenia. The authors suggest that as females recorded higher levels of trauma, this may explain why no relationship between TBI and schizophrenia was observed.

It has been proposed that family attributes may increase the risk of TBI which in turn increases the risk of schizophrenia in persons with a genetic vulnerability (Malaspina et al., 2001). This proposal may explain the variability in the research findings, with a failure to control for genetic vulnerability confounding the results. A limitation of some previous research is the failure of patient assessments to investigate a history of TBI. Furthermore when an injury is noted there

appears to be confusion about whether to diagnose psychosis due to a TBI, or a schizophrenia spectrum disorder (Fujii, 2005; Kim et al., 2007).

Therefore the literature tentatively indicates that suffering a TBI increases the risk that someone will experience a psychotic disorder. Genetic vulnerability may play a specific additional risk factor with males at higher risk. To consolidate future research efforts clear diagnostic guidelines which cover TBI and the spectrum of psychotic disorders are required (Kim et al., 2007). However there is one clear implication of research to date: psychosis is a potentially serious and debilitating consequence of TBI, making accurate diagnoses to facilitate appropriate treatment crucial (Arciniegas et al., 2003).

Summary

The prevalence of various Axis I and Axis II disorders in groups of TBI survivors is higher than for the community. In particular mood disorders are implicated as complicating factors for recovery in people with a TBI. The role of a TBI in the development of psychotic illness requires more research. In addition, consistent use of well-established, explicitly defined measures of the constructs being investigated is necessary to compare and interpret the research into psychosis after a TBI.

It is possible that Axis I and Axis II prevalence rates are higher partly due to the personality influences that increase the risk of having a TBI. This will be discussed in more detail below. Regardless, the challenge of living with a TBI clearly creates additional stress that can lead to a psychological disorder. The presence of this comorbidity further diminishes one's quality of life, with the resultant higher risk of suicide being particularly concerning. Finally those patients suffering TBI comorbid with psychopathology have a worse prognosis than those with TBI alone.

1.3 Aetiology and risk factors for the development of psychopathology after TBI

In this section the various explanations of why a TBI increases the risk of developing psychopathology will be discussed. The most common explanations are 1) neurological, which includes both biological and anatomical alterations, 2) psychosocial, which considers difficulty

adapting to the changes in personal capacity and 3) a combination whereby the deficits associated with the neurological damage result in psychosocial dysfunction.

Neurological model of psychopathology

Although some researchers suggest the psychopathology subsequent to a TBI is due to a biochemical imbalance or dysfunctional neurotransmitters, most of the literature suggesting a neurological hypothesis implicates diminished functioning of specific neuroanatomical regions (Gualtieri & Johnson, 1999; Jorge & Starkstein, 2005). People who have suffered a TBI can have severe frontotemporal effects along with severe multi-focal and diffuse effects (Iverson & Lange, 2011). Patients with damage to their frontal lobes are noted to present with symptoms of anergia, anhedonia and lack of initiative which are also symptoms of depression (Gualtieri & Johnson, 1999). Additionally damage to this region is typified by disinhibited and socially inappropriate behaviour, which may be especially pertinent to personality disorders diagnoses (Miller, 1996). Similarly damage to the temporal lobe can result in depressive symptoms such as irritability or fatigue resulting from difficulty in processing volumes of information simultaneously (Folzer, 2001). The common theme from these studies is the symptoms of psychopathology are purported to result from neurological damage. However, none of the studies have effectively excluded the presence of the very psychopathology the symptoms indicate.

Busch and Alpern (1998) reviewed research into depression comorbid with mild TBI. Some important conclusions were that pathophysiological changes generated by a TBI may result in a depressive episode in a vulnerable subset of the population. The authors propose that the aetiology for depression after TBI and non-traumatic depression may differ. Proposing differences in aetiology implies an exclusively neurological basis for depressive symptoms that develop after a TBI. If the damage is altering the neurological functioning then it must differ from those undamaged who also present with depression. In a seeming contradiction to the central thesis of their paper, Busch and Alpern acknowledge the role of psychosocial factors in depression subsequent to a TBI and the need for further research.

In the previous section, Depressive Disorders were identified as the most common psychopathology comorbid with TBI. It is not a surprise that most neurologic theories focus on this disorder. However, a neurological basis for anxiety disorders has also been proposed. Gray and McNaughton (1996) propose that damage to specific neurological regions generates

hypersensitivity to stimuli resulting in chronic anxiety. It is proposes that anxiety symptoms are associated with damage to the behavioural inhibition system which includes regions such as the septo-hippocampal system, the anterior thalamus, Papez circuit', cingulate cortex, pre-frontal cortex and ascending noradrenergic fibres of the locus coeruleus.

Jorge et al. (2005) suggest that while mood disorder following a TBI likely occurs from an interaction between biological and psychosocial factors, in specific circumstances neurologic factors are implicated. The authors propose alcohol abuse prior to a TBI has neurotoxic consequences that in combination with the injury produce interference with the neural circuits that modulate reward and mood. Consequently individuals with prior alcohol abuse are at increased risk of developing a mood disorder comorbid with their TBI. Therefore only in particular circumstances is psychopathology in a person with a TBI specifically related to neurologic impairment. However, substance abuse is observed to occur often in conjunction with a mood disorder in a variety of populations (Ostacher, 2007). The comorbidity of substance abuse and mood disorders independent of a TBI weakens the neurological argument.

Crowe (2008) concludes that research fails to clearly implicate neurological damage as causing psychopathology after a TBI. Crowe argues that if psychopathology is due to neuroanatomical damage from injury then a specific pattern of impairment and location of lesion would correlate with types of psychopathology. Additionally increased injury severity of TBI would increase the risk of psychopathology. Neither proposition is supported by the literature. Similarly the recovery process is frequently disproportionately more difficult than would be expected purely based on the objective facets of the injury (Al-Adawi et al., 2007). Consequently a neurologic explanation for psychopathology is inadequate.

Psychosocial model of psychopathology

A psychosocial explanation suggests psychopathology develops when the challenges in adapting to changes in life functioning after a TBI are overwhelming. In both acute and post-acute phases of recovery change in autonomy, interpersonal relationships and self-image present extensive adaptational problems for the individual with a TBI (Alderfer et al., 2005). The immediate challenges after a TBI include hospitalization or acute care, diminished psychosocial functioning, diminished cognitive functioning, along with an unclear prognosis. Development of psychopathology when confronted with these challenges is also likely related to pre-injury

personality characteristics and access to a social support network. During the acute phase of recovery a patient with TBI may view these difficulties as temporary, when the difficulties are prolonged, the demoralizing effect makes adjusting to the new self increasingly difficult.

Additionally there are often new and complicating challenges for a TBI survivor such as loss of job, chronic pain, social isolation, memory loss, financial difficulties and litigation (Moore et al., 2006). The combination of additional challenges with duration after TBI, along with the demoralizing effects of a prolonged recovery period might be expected to increase the risks for psychopathology. Supporting this expectation, studies have found prevalence rates of psychopathology in TBI samples do increase with duration after injury. In a longitudinal study by MacNiven and Finlayson (1993) the paper found greater levels of psychopathology in a TBI sample at 24 months after injury, than at 12 months after injury.

The challenges in adapting to functional changes in a person who suffered a TBI are also implicated in the development of a psychotic illness. Fujii (2005) discusses the role of adapting to the consequences from a TBI as a stressor that may lead to schizophrenia in people with a genetic vulnerability. Persons suffering a TBI who have a genetic vulnerability for psychosis are more likely to develop associated psychopathology than those without the genetic vulnerability. Additionally persons with a TBI and genetic vulnerability are at increased risk of psychosis if they utilise maladaptive coping strategies. This proposal identifies adaptational difficulties as a fundamental component in developing psychotic illness after TBI, supporting a psychosocial explanation.

The above research has identified specific difficulties prevalent during the acute and ongoing phases of recovery that relate to psychopathology. These psychosocial difficulties are suggested to increase the risk for psychopathology in a person with a TBI.

Neuropsychosocial model of psychopathology

Some research proposes the combination of neurological and psychosocial factors are operating concurrently in the expression of psychopathology (Rosenthal et al., 1998). A neuropsychosocial paradigm considers that a variety of neurological and psychosocial features of TBI combine to increase the risks of psychopathology. The interaction of these features can vary dramatically across individuals with the same pathology (Alderfer et al., 2005). Individual who suffer a TBI

have specific deficits due to the neurological damage such as cognitive dysfunction and behavioural dyscontrol which along with physical limitations reduces their capacity to adapt to their environment (Fujii, 2005). Adjustment difficulties can result in psychological distress, along with further influence from factors such as being unemployed and lacking social support networks (Folzer, 2001; Prigatano, Altman, & O'Brien, 1990).

Folzer (2001) describes how interpersonal relationships suffer after a TBI. She provides a scenario that exemplifies the processes in the neuropsychosocial model. Folzer explains that communication deficits may be interpreted by others as disinterest. Furthermore, as neurological deficits are associated disfluency, interpersonal communication may suffer which could lead to reduced social interaction. In both cases the person suffering a TBI is at increased risk of social isolation, thus removing the social support network as a protective mechanism against psychopathology. Additionally reduced information processing capabilities can make social environments overwhelming which may lead to feelings of irritability and frustration that might be projected onto others. As social support networks are important protective mechanisms against developing psychopathology the neurological deficits that impact these networks are specifically implicated in the experience of psychological distress.

A neuropsychosocial model of psychopathology integrates psychosocial impacts from specific neuropsychological symptoms which can combine to increase the risk of psychopathology. Ongoing lack of awareness as to the impacts on a person's changed capabilities after injury can increase the risk for developing psychopathology. Below the Stress-Appraisal-Coping model will be described and used to show how reduced awareness can be protective in the short term. However in the long term the same reduced awareness can become problematic.

Another explanation for the MacNiven and Finlayson (1993) findings is that neurological recovery can increase the risk for psychopathology (Hibbard et al., 2000). This can be explained as a function of awareness. Self-insight or awareness of a person's reduced functional capabilities after a TBI can increase the risk for psychopathology. Awareness of adaptational challenges is an important component of a psychosocial understanding of psychopathology in persons with a TBI. The Stress-Appraisal-Coping model of psychopathology (Lazarus & Folkman, 1984) will be discussed in more detail below but importantly recognition of a stressor is fundamental to the experience of psychological distress. If someone is unaware of their

deficits they may not perceive a situation as potentially difficult, or stressful, and consequently they experience less psychological distress. With neurological recovery comes increased awareness of functional limitations and prognosis, which can explain the increase in rates of psychopathology with duration after injury found by MacNiven and Finlayson. Hibbard and colleagues propose a neurological influence on the expression of psychopathology is mediated by insight into psychosocial challenges.

There is a substantial body of research that identifies diminished insight as a common symptom after TBI. Fleming and Strong (1995) propose reduced insight can present in three different ways. First is reduced insight into the injury related deficits. Second is failing to identify the implications of these injury related deficits. Third is lacking the necessary insight to determine realistic goals or recognise prognosis. Similarly Kortte, Wegner, & Chwalisz (2003) suggest reduced insight can present as a lack of awareness of the injury, underestimating impairments or failing to recognise the severity of neuropsychological symptoms.

McAvinue, O'Keefe, McMackin and Robertson (2005) compared error awareness in 18 TBI patients with 16 matched control participants. They found evidence for reduced insight in patients who had suffered a TBI. These findings were replicated in a second study from the same paper with a different sample of 19 TBI patients and 20 controls. The paper also found a significant correlation between increase in TBI severity and reduced insight.
Diaz et al. (2012) completed psychopathological assessment on 33 patients with severe TBI, 18 months after hospitalization. One of their findings was that TBI patients with or without personality change after injury were not significantly different on measures of mental health domains. Diaz et al. propose that a lack of insight into the limitations caused by a TBI may explain this finding. Importantly studies have observed that the presentation of reduced insight after TBI appears to result from disruption in the person's conscious experience, and not as a defence mechanism (Prigatano, 1999).

Bogod, Mateer and McDonald (2003) found the Self-Awareness of Deficits Interview was able to predict the level of injury severity in a TBI sample. This finding shows that the level of insight into deficits after a TBI varies with the severity of injury. The study found that an increase in TBI severity was associated with a decrease in awareness of injury related deficits. This finding is important as the relationship between TBI severity and the risk for developing

psychopathology is proposed to be mediated by insight. More information about this topic is covered in sections 1.5 and 1.6 below.

For a person with TBI, awareness can become problematic when insight reveals challenges that are considered to require a coping strategy. A copy strategy is adaptive when the outcome is a reduction in distress or maladaptive when the strategy fails to reduce distress. Maladaptive coping strategies employed to manage the effects of a TBI are implicated in psychological disorders. Epstein and Ursano (1994) propose patients develop compulsions as coping mechanisms to deal with the loss of control in their life, which may lead to Obsessive Compulsive Disorder. Maladaptive coping strategies, such as avoidance, are a common focus in therapy for anxiety disorders. This same non-organic explanation for a behavioural basis of the development of an anxiety disorder can apply to Post Traumatic Stress Disorder, Generalised Anxiety Disorder and PD (Hiott & Labbate, 2002; Moore et al., 2006). Therefore awareness of deficits can be implicated in the development of psychopathology.

Folzer (2001) found relatives of TBI patients report a variety of neuropsychological symptoms which can increase the risk for psychopathology. Symptoms include overestimating one's capacity to emotionally handle arguments and unexpected change, being unaware of one's own emotional state, and difficulty controlling anger. These behaviours are associated with problems in social and work relations which explain the high unemployment rates of TBI sufferers (Prigatano, Altman & O'Brien, 1990). The concern is that the mistaken beliefs of capabilities on the part of a TBI patient may alter their employment status and social support network. With social support networks and employment act as protective mechanisms against psychopathology, neuropsychological deficits that negatively impact these mechanisms increase the risk for psychopathology. Furthermore psychological therapy which can reduce these risks by developing adaptive coping strategies is dependent upon patient insight (Godfrey, Knight, & Partridge, 1996). Therefore prolonged diminished awareness can be detrimental for patient rehabilitation.

Research that suggests duration after injury can vary the influence of neurological or psychosocial factors supports a neuropsychosocial model of psychopathology. Jorge, Robinson, Arndt et al. (1993) found a relation between neuroanatomical damage and acute onset depression but not later developing depression in a TBI sample. The findings imply that acute depression is

more strongly influenced by neurological factors while delayed onset depression may be more influenced by psychosocial factors (Alderfer et al., 2005). The difficulty for clinicians working with TBI patients is that no clear distinction exists between the acute and post-acute phases of rehabilitation. Furthermore the rate of neurological recovery in persons with a TBI varies considerably (Crowe, 2008). The strength of a neuropsychosocial model is that multiple factors are considered independently and interactively by a clinician to explain symptoms presentation in persons with a TBI.

Summary

The suggested risk factors for developing psychopathology are diverse. Some research provides support for a neurological basis and this research appears strongest when psychopathology develops in the acute phase. In contrast support for the role of adaptational difficulties in the development of psychopathology is found especially when a prolonged recovery period ensues. In reality the picture is quite complex, with neurological and psychosocial facets both being implicated in the literature. The implication of both facets may reflect the shortage of clear research into the topic, the complexity of the situation and the nuances of all factors being sensitive to the individual with a TBI.

The strength of the neuropsychosocial model is that it incorporates all the benefits of the previous models discussed while minimising any weaknesses. Neurological trauma is related to specific cognitive and functional deficits in a person suffering a TBI. These deficits create particular challenges whilst also making coping with these challenges more difficult. Therefore whilst in some cases psychopathology may purely result from neurological or psychosocial factors in most cases psychopathology is a consequence of both. Consequently the requirement for sensitive assessment of each person with a TBI is paramount.

1.4 Risk factors for a TBI

Any person may suffer a TBI. However, existing psychopathology, a person's age and gender increase the risk. An example of existing psychopathology as a risk factor is when brain injury occurs from self-injurious behaviour associated with depression (Moldover, Goldberg, & Prout, 2004). Research that finds antisocial personality disorder and substance use disorders occur pre-injury at higher levels in TBI samples compared with community samples infers these

pathologies as risk factors (Hibbard et al., 2000; Hibbard et al., 1998). These findings may be a consequence of more interpersonal violence associated with antisocial personality disorder and the increase risk in motor vehicle accidents associated with substance use disorders.

Motor vehicle accidents are the leading cause of TBI. Consequently, any psychopathology that increases the risk for a motor vehicle accident also increases the risk for a TBI. There are two scenarios which best exemplify the influence of psychopathology in motor vehicle accidents: the accident which is a suicide attempt and the accident which is the consequence of driving under the influence of alcohol or drugs (Moldover et al., 2004). Accurate data measuring these scenarios is difficult to obtain. However, according to a survey in the US of drug use and driving behaviour around 12% of the population aged 12 or older drove while under the influence of alcohol, and just over 4% of the same population drove while under the influence of illicit drugs (National Institute on Drug Abuse, 2010). Amongst young adults aged from 18 to 25, driving while under the influence of illicit drugs occurred in 12.8%. Males were more likely than females to drive under the influence of either alcohol or illicit drugs. Suicide and suicidal ideation is associated with depression, bipolar disorder and anxiety disorders indicating an increased risk of suicidal behaviour for individuals suffering a mood disorder (Boden, Fergusson, & Horwood, 2007; Dumais et al., 2005; Ohberg, Penttila, & Lonnqvist, 1997). Thus a motor vehicle accident which is an unsuccessful suicide attempt in a person with psychopathology can lead to comorbid TBI. Unfortunately data illustrating the relationship between suicidal behaviour and motor vehicle accidents is scarce. This is predominantly because the road safety data systems do not record this information, resulting in possible underestimation of the prevalence rates (Routley, Staines, Brennan, Hawoth, & Ozanne-Smith, 2003).

Investigation of substance abuse and mood disorders identifies the interactive nature of these factors in increasing the risk for a motor vehicle accident. Grant and colleagues (2006) found a DSM-IV mood disorder among 20% of respondents with a substance use disorder compared with 8% amongst those without a substance use disorder. A substance use disorder may develop from the use of licit or illicit drugs to deal with psychological distress. This position is reflected in the frequent comorbidity between substance use disorders and mood disorders (Moore et al., 2006). Substance use is thought to explain the increased susceptibility of people with a psychological disorder to a TBI (Ostacher, 2007). In support of this position Hibbard, Uysal, Kepler, Bogdany and Silver (1998) studied 100 patients aged from 18 – 65 who had a TBI, finding substance use

disorders commonly predated the patient's injury.

It is likely that the existence of a mood disorder moderates the relationship between substance use disorders and TBI. Some individuals' self-medicate through the use of drugs to deal with their psychopathology and this behaviour increases their risk of having a motor vehicle accident (Ostacher, 2007). While for others the symptoms of the mood disorder increase the risk of motor vehicle accident (Rihmer, 2007). With motor vehicle accident being the leading cause of a TBI (Brain Injury Association of Queensland, 2006) and a mood disorder increasing the risk for having a motor vehicle accident, it can be seen that an individual's chances of developing a TBI is increased by having a mood disorder.

Along with these psychopathologies, gender is implicated as a risk factor for suffering a TBI. Males are twice as likely to suffer a TBI compared with females. In 2008 almost 70% of TBI's in Australia was suffered by males with this proportion consistent across severity ranges (Access Economics Pty Ltd, 2009). Additionally the highest incidence rate for TBI is amongst males aged between 15 and 24 (Access Economics Pty Ltd). These findings indicate young males with current mental health problems are at specific risk for suffering a TBI.

Summary

The interaction between mental health and TBI is bidirectional. Research has found that having a psychological disorder increases the risk for suffering a TBI and suffering a TBI increases the risk for developing psychopathology. Additionally a person suffering a TBI is more likely to be a young male who faces an extended period of coping with the functional limitations resulting from their injuries.

1.5 The relationship between TBI severity and psychopathology

Intuitively under the neuropsychosocial model one might expect the more severe the TBI the greater the risk for psychopathology. That is, with more severe deficits from greater neurological damage a person will be exposed to more situations that require a coping strategy. Additionally the person with a more severe TBI has a decreased capacity to cope. As a result the risk for psychopathology increases.

Supporting the intuition that the risk for psychopathology increases with the extent of neurologic damage are the studies that find that the rates of depression and psychosis is greater in those with a severe TBI (Arciniegas et al., 2003; Fujii & Ahmed, 2002; Holsinger et al., 2002). In contrast some studies have observed no clear relationship between severity of TBI and the subsequent occurrence of depression or psychosis (Dikmen, Bombardier, Machamer, Fann, & Temkin, 2004; Malaspina et al., 2001). There are two key limitations to research suggesting greater severity of injury increases the risk for psychosis. The first is the shortage of replication studies, and the second is differences across studies in the operational criteria (Kim et al., 2007). The Holdinger et al. (2002) study, which found higher rates of depression in the severe TBI group, is limited in generalisability due to use of a sample of veterans from World War II with retrospective diagnoses of TBI.

In contrast there is a large body of research that suggests as TBI severity decreases the risk for psychopathology increases. This research will now be reviewed. Diminished insight is a neuropsychological symptom with increased TBI severity associated with lower levels of insight (McAvinue et al., 2005). In the neuropsychosocial model reduced insight was identified as protective against psychopathology. Therefore it would be expected that the risk of psychopathology decreases as the severity of a TBI increases. The research into the relationship between injury severity and psychopathology finds compelling support for this explanation (Cooper-Evans, Alderman, Knight, & Oddy, 2008; Crowe, 2008; Lillie et al., 2010; Miller & Donders, 2001; Morton & Wehman, 1995; Rosenthal et al., 1998).

In a study evaluating 150 patients with TBI using the MMPI-2, individuals with mild injuries recorded higher levels of symptomology than patients with moderate-to-severe injuries (Lillie et al., 2010). Rosenthal, Christensen and Ross (1998) also found a greater number of depressive symptoms were reported in individuals with mild TBI compared to those with more severe TBI. Furthermore, studies have observed that people with mild brain injuries are more aware of their functional limitations than those with severe injuries (Morton & Wehman, 1995). Therefore, studies are findings that the mild TBI patients have both increased insight, and increased risk for psychopathology. These findings may reflect the role of insight in development of psychopathology.

Dikmen et al. (2004) found no difference in emotional functioning across severity ranges; however they did observe greater problems with hyperarousal in the more severely injured group. This finding may suggest that anxiety disorders would be at increased risk with greater severity of TBI. However like the research into depressive disorders the prevailing view is that persons suffering mild TBI report higher levels of anxiety symptoms than those with a moderate or severe TBI (Crowe, 2008). Therefore lower severity of TBI increases the risk of suffering a mood and anxiety disorder. This pattern of findings is consistent with understanding that deficit awareness, or insight, is associated with psychological distress.

The role of insight in development of psychopathology is advocated by Godfrey, Partridge, Knight and Bishara (1993) who found the return of insight was associated with increased risk for emotional distress. Accordingly even in a sample of severely injured TBI patients' higher levels of awareness were related to lower levels of self-esteem; and self-esteem was related to levels of psychological distress (Cooper-Evans et al., 2008). Therefore the role of insight in increasing the risk for psychopathology is not specific to any one end of the injury severity continuum. The comparative risk from differing levels of insight is observed across the complete injury spectrum.

Summary

The emerging theme from research is an inverse relationship between injury severity and psychopathology, also known as the injury severity paradox. That is, individuals who suffer a mild TBI are at increased risk of psychological distress compared with those suffering moderate or severe injuries. This role of insight into deficits, in the development of psychopathology, is proposed to explain the injury severity paradox.

1.6 Stress-Appraisal-Coping Model and TBI

This section will outline key elements of the Stress-Appraisal-Coping model. The model will be shown to provide an explanation for the injury severity paradox. Additionally the model can assist a clinician to understanding the experience of psychological distress in TBI patients and inform treatment. Finally the model can provide valuable information that can be used to guide therapy.

The Stress-Appraisal-Coping model proposes that psychological distress is not specifically the by-product of particular characteristics of an event, but an outcome from an individual's evaluation of their ability to cope with the consequences of the event. Lazarus and Folkman (1984) proposed the Stress-Appraisal-Coping model explains why people differ in developing psychopathology despite similar life events. In other words, the development of psychopathology after a TBI is the outcome of a person's belief in their capacity to cope with identified adaptational challenges associated with their injury.

According to Lazarus and Folkman (1984), psychological distress occurs when an individual appraises their environment as requiring coping skills they do not possess. There are two stages in this appraisal process. First is the primary cognitive appraisal in which the demands of the environment are deemed to be either irrelevant, benign-positive or stressful. If deemed stressful a secondary appraisal occurs. In this stage a person determines what can be done to meet the environmental needs, and evaluates the likelihood any given strategy will achieve its goals. Importantly, coping is defined as the capacity to manage stressful demands and no one strategy is deemed better than another. The essential issue is the emotional benefit of a strategy. The Stress-Appraisal-Coping process is iterative. For each new encounter the success or failure of a coping strategy applied in a previous encounter is important information in determining one's capacity to cope with the demands of the current situation.

The Stress-Appraisal-Coping process demonstrates that two steps are involved in generating psychological distress. The first is evaluating whether the situation warrants the application of a coping strategy, and secondly whether the person has the capabilities to successfully employ a coping strategy. Therefore not perceiving a situation as stressful can have a similar emotional benefit as applying a successful coping strategy.

This process provides an explanation for the role of insight into the injury severity paradox highlighted earlier. Studies were reviewed which found that mild TBI was associated with higher levels of psychopathology than that associated with moderate to severe TBI (Crowe, 2008; Rosenthal et al., 1998). Additionally research indicated that individuals with mild TBI are more aware of their problems than those with severe TBI (Morton & Wehman, 1995). This reduced awareness may result in a person with severe TBI being less inclined to determine that their environment requires a coping strategy than a person with a mild TBI. In stage one of the Stress-

Appraisal-Coping model a person is required to determine the environment it potentially stressful. If someone does not perceive the environment as potentially stressful then they are at reduced risk for psychopathology under the model. Therefore a person with severe TBI may be less likely than a person with mild TBI to experience psychological distress because they fail to determine their environment may require a coping strategy. The decision that the environment does not require a coping strategy may result from their reduced insight into the demands of the environment.

The neuropsychosocial framework for understanding psychopathology (see Section 1.3) is compatible with the Stress-Appraisal-Coping model for psychopathology. It may appear that including neurologic influences contradicts the Stress-Appraisal-Coping model. However neurologic deficits were shown to lead to behaviours that resulted in various psychosocial challenges (Folzer, 2001). These challenges are examples of potential stressors in the Stress-Appraisal-Coping model. Whilst neurological damage may generate the stressor, the appraisal and coping process is the mechanism that leads to psychological distress. Additionally for a person suffering a TBI, both the deficits from neurological damage and their psychosocial capacities are considered when determining coping options.

The cognitive processes of appraisal and coping embody the key characteristics of Cognitive Behavioural Therapy (CBT). That is, the automatic 'thoughts' a person has in a situation result in the associated emotions and behaviours. It is not the 'situation' itself that leads to the distress. Lazarus and Folkman (1984) recognise the synergy between CBT and the Stress-Appraisal-Coping model. In their conclusion they write "the view that how we act and feel depends on the way we think, specifically the way we appraise the significance of encounters for our well-being, is a major premise of our formulation of stress and coping….(which) is the assumption underlying the cognitive-behavioural therapies" (p350). The authors note that CBT as congruent with their cognitive theory of stress and coping.

Research supporting the Stress-Appraisal-Coping model

Folkman and Lazarus (1986) found a difference in the appraisal and coping process in subjects with high levels of depressive symptoms compared to those with low levels. They found that members of the high depressive group rated the importance of a situation higher during 'primary appraisal'. A finding which supports understanding that appraising a situation as stressful, and

consequently requiring a coping strategy, is related to psychological distress. Folkman and Lazarus also found differences between the groups on which types of coping strategies were utilised. Those with higher depressive symptoms used more self-control, confrontive coping, and escape-avoidance strategies. Additionally they accepted more responsibility for a situation and responded with more worry or fear and anger or disgust. As a result, Folkman and Lazarus found psychological distress was related to both cognitive appraisals and coping strategies.

The importance of specific coping strategies in the experience of psychological distress is also found in studies where the participants' were the victims of trauma. Pollard and Kennedy (2007) found coping strategies explain emotional adjustment in a sample of individuals who suffered a spinal cord injury. While Walsh, Fortier and DiLillo (2010) found coping strategies explained much of the variance in long term emotional functioning of people with a history of childhood sexual abuse. Importantly these studies find coping strategies relate to both the acute and long term functioning in victims of trauma.

The applicability of the Stress-Appraisal-Coping model is found in a variety of life events. The coping approach employed by parents of children with Autism Spectrum Disorder was found to mediate parental psychological distress (Pottie & Ingram, 2008). Even the experience of migraines is associated with the appraisal and coping processes. Huber & Henrich (2003) found individuals who suffer migraines typically used avoidance, preoccupation and social isolation as coping strategies, while considering themselves less able to relax and more irritable than people who do not suffer migraines.

These studies are just a sample of those that found support for the Stress-Appraisal-Coping model. Importantly they show the role of both the primary and secondary appraisal processes, across a variety of circumstances, is related to the experience of psychological distress. From a clinical perspective the information garnered around the coping strategies that fail to provide emotional benefits is important. That is because these maladaptive coping strategies can be targeted in therapy.

Research into the Stress-Appraisal-Coping model and TBI

Research supports the Stress-Appraisal-Coping model as an explanation of psychological distress in persons who have suffered a TBI. Crowe (2008) identifies that appraisal of events was

a better predictor of psychological distress than the number of stressful events experienced. Strom and Kosciulek (2007) found perceived stress in patients with TBI predicted self-reported depression, which in turn predicted lower levels of hope and life satisfaction. Additionally differences in coping with stressors associated with mild traumatic brain injury can account for the variations in levels of symptom complaint and disability (Machulda, Berquist, Ito, & Chew, 1998).

Whilst these studies support the Stress-Appraisal-Coping model by highlighting the importance of appraisal and coping mechanisms Godfrey, Knight and Partridge (1996) specifically used a Stress-Appraisal-Coping framework to study the emotional adjustment in persons suffering a TBI. They conclude emotional adjustment of the individual with TBI is seen as a response to neuropsychological symptoms and associated losses. However the response is mediated by the individuals coping skills, social support, and insight. The mediating variables are described as altering the severity of psychological distress in TBI patients by modifying the impact of stressors. Mediators include TBI patients' appraisal of the difficulty of a challenge, the capacity of their coping skills to meet the challenge and the degree to which they effectively employ social support networks for assistance.

The role of insight in the appraisal stage can explain both the injury severity paradox and increased rates of psychopathology with longer duration after injury. Godfrey et al. (1993) found TBI patients under-report the impact of behavioural problems within the first 6 months after injury when compared with reports by individuals injured between 12 and 36 months earlier. However, those in the latter group also presented with higher levels of emotional dysfunction. The Stress-Appraisal-Coping model explains this finding as an outcome of the awareness of stressors that require coping strategies. This awareness allows some individuals to successfully employ a coping strategy. Yet others will employ a maladaptive strategy, such as avoidance, or experience psychological distress from appraising the needs of the environment as exceeding their capacity. In the early stages of recovery the diminished insight is protective because the environmental needs are not appraised as requiring a coping strategy. Thus removing the potential to utilise maladaptive approaches or deem the needs as exceeding the individual's capacity. With the return of insight the needs of the environment become more often appraised as requiring a coping strategy, the first step towards psychological distress.

The capacity of the Stress-Appraisal-Coping model to explain the injury severity paradox follows the same logic as the reasons provided for increasing levels of psychological distress with time since injury. Insight into the needs of the environment is diminished with increased severity of injury (Crowe, 2008). Previously discussed was the research which found lower levels of psychopathology with increased TBI severity. Therefore, again, diminished awareness can be seen as protective. It is likely the finding of increased psychopathology with duration after injury is occurring because the severity of the injury is reducing over time with natural recovery of the neurological structures damaged in the incident. Both the injury severity paradox, and the differing levels of psychopathology found over duration, reflect the specific role of insight into deficits outlined in the Stress-Appraisal-Coping model.

1.7 Treatment for persons who have suffered a TBI

Overview

As psychopathology has been shown inhibit recovery from a TBI, treatment for psychological distress is crucial for rehabilitation (Davis, Reeves, Hastie, Graff-Radford, & Naliboff, 2000; Fann et al., 1995). Fortunately the psychological distress commonly experienced by individuals suffering a TBI is ameliorable with treatment. In studies of pharmacological and psychological therapies, benefits such as improved cognitive ability and reduced experiences of pain for patients with a TBI reflect both physical and psychological gains from treatment (Anson & Ponsford, 2006c; Fann, Uomoto, & Katon, 2001). These findings are reflected in the recommendation that therapy for psychopathology is a fundamental aspect of TBI rehabilitation programs (Folzer, 2001; Hiott & Labbate, 2002; Koponen et al., 2002; Kurtz, Shealy, & Putnam, 2007; Rapoport et al., 2006; Rosenthal et al., 1998).

The importance for treating psychopathology in TBI rehabilitation programs is supported by findings that the outcomes from injury are better predicted by psychological symptoms rather than physical symptoms (Rosenthal et al., 1998). The implication of psychosocial factors in the emergence of psychopathology, even long after injury, supports utilising therapies such as CBT that can focus on these psychosocial influences (Alderfer et al., 2005). Research reviewed previously established high rates of comorbid TBI and psychopathology (see Section 1.2). Therefore without appropriate therapy for psychopathology a large proportion of individuals suffering a TBI are hampered in their rehabilitation. Furthermore as the archetypal TBI sufferer

is under the age of 25, an untreated psychological illness can persistently diminish a person's quality of life for decades.

The APA Task Force on Promotion and Dissemination of Psychological Procedures (Chambless et al.,1996) notes the importance of employing cross-validated assessment instruments, and the application of a test such as the MMPI-2 can guide the approach to treatment (Gori, Lauro-Grott, Giannini, & Schulderberg (2010). Studies have found the MMPI-2 scales can predict outcomes from psychotherapy programs (Chisolm, Crowther, & Ben-Porath, 1997; Forbes et al., 2002). The incremental validity of the MMPI-2 has been questioned when it is utilised with other diagnostic instruments (Lima et al, 2005). However it is considered that the MMPI-2 can assist in complicated diagnostic decisions while also informing a clinician as to the client's typical behaviours and thinking styles (Butcher & Rouse, 1996). This conclusion points to the value of the MMPI-2 as both a personality and a clinical instrument

Psychotherapy

Without diminishing the potential role for pharmacotherapy in the rehabilitation of some TBI sufferers there is little doubt that withholding psychotherapy from a patient with TBI risks delaying and limiting their recovery (Folzer, 2001; Pollack, 2005). Koponen et al. (2002) found patients with a TBI who suffer major depression in the first few months after injury have inferior outcomes. This finding highlights the importance for prompt treatment of psychopathology. Addressing the issues of pain, lowered self-esteem, coping with adjustments, peer support and the change in family dynamics are all aspects which require careful consideration during treatment and may require either or both cognitive and behavioural approaches (Hibbard et al., 2002; Hiott & Labbate, 2002; Machulda et al., 1998; Seel et al., 2003). Therefore the use of CBT techniques is well suited for individuals suffering a TBI.

The proposal that psychopathology after a TBI develops due to factors identified in the Stress-Appraisal-Coping model is theoretically congruent with the CBT framework. Under the Stress-Appraisal-Coping model the coping process includes conscious or unconscious applications of cognitive and behavioural strategies to ameliorate the experience of stress (Vosvick, Martin, Smith, & Jenkins, 2010). Maladaptive cognitive and behavioural strategies can be targeted in therapy. From a therapeutic perspective, a key advantage of the Stress-Appraisal-Coping model is that it can help understand the process of adjusting to the consequences of a TBI (Strom &

Kosciulek, 2007). Strom and Kosciulek found the Stress-Appraisal-Coping model could be used to identify when the appraisal and coping process is stressful, or overwhelming, which is central to the experience of emotional distress. These findings further illustrate the role of awareness, or insight, in psychological distress for sufferers of a TBI.

CBT provides practitioners with a variety of techniques, which include using behavioural therapy, problem solving skills, assertiveness training, and peer support along with cognitive therapy techniques. Behavioural therapy has been recommended for TBI patients with Obsessive Compulsive Disorder, whilst peer support is shown to benefit both the individual with a TBI and their family (Hibbard et al., 2002; Hiott & Labbate, 2002). Peer support programs may help generate an understanding of the reasons some problems occur, and to develop necessary skills that can assist in adapting to life with a TBI (Lazarus & Folkman, 1984). The strength of CBT is found the repertoire of techniques available in therapy which can be refined to meet the specific needs of a patient with a TBI. Some specific adaptations to therapy are recommended when working with TBI patients such as using memory techniques, attentional training, relaxation techniques to assist with over arousal and using a more concrete approach to cognitive work (Folzer, 2001).

Anson & Ponsford (2006c) completed a study of group based CBT over 10 sessions with 33 TBI sufferers. They reported that better outcomes following intervention were associated with greater understanding of injury-related deficits. Additionally the same authors found that CBT can modify the coping strategies applied by people with a TBI (Anson & Ponsford, 2006b). However, a difficult aspect of therapy is determining the appropriate time to discuss injury-related deficits. Without insight into deficits psychotherapy cannot assist in the implementation of an adaptive coping strategy. However as previously discussed, insight into deficits is also related to levels of psychological distress. The impact of psychoeducation on symptomology is seen in the findings of Smith and Godfrey (1995), where family education improved symptoms awareness in patients with a TBI, but the increased awareness was associated with higher levels of depression.

The findings by Smith and Godfrey (1995) may be interpreted as suggesting psychotherapy for some people with TBI is inappropriate. That is, if reduced insight protects against psychopathology then psychoeducation is unnecessary and potentially harmful. However,

Godfrey and colleagues (Godfrey et al., 1996; Godfrey et al., 1993) suggest that because patients with a TBI and diminished insight have no improved long term psychosocial outcomes, despite the immediate reduction of emotional distress, timely assistance to generate patient insight is appropriate. The authors propose cultivating insight when symptoms can be adequately monitored along with designing appropriate support structures to assist the patient during this difficult phase. An implication from the recommendations by Godfrey and colleagues is that individuals with a TBI will at some stage become aware of their injury related deficits. It is inappropriate to suggest this prognosis for all TBI sufferers, but the lack of insight or psychopathology at one point in time does not negate the possible development of insight or psychopathology in the future. MacNiven and Finlayson (1993) found greater levels of psychopathology in a TBI sample at 24 months after injury, than for the same sample at 12 months after injury reflecting this understanding.

As such rehabilitation after a TBI would benefit from periodic reassessment to screen for developing psychopathology. Clinical instruments such as the MMPI-2 which provide scores for psychopathology on a continuum are especially useful in this context. If a patient generates higher scores at reassessment then even if scores remain below diagnostic levels they may be reflecting a return of awareness. Therefore the time may be right to protect against worsening psychological distress by following the approach recommended by Godfrey and colleagues while utilising CBT to assist in the adjustment process (Godfrey et al., 1996; Godfrey et al., 1993). Additionally periodic reassessment may uncover psychopathology that has developed since the previous evaluation. Failure to diagnose psychopathology at any stage of recovery leads to denial of treatment, leading to worse symptoms and compounded consequences (Handel, Ovitt, Spiro, & Vani Rao, 2007).

1.8 General Diagnostic Challenges with TBI patients

Handel and colleagues (2007) identify the assessment process as crucial for TBI sufferers, with eligibility for treatment and the subsequent likelihood of a successful outcome being aided by accurate diagnosis. McAllister and Flashman (1999) caution that without careful initial assessment there is an increased risk of following a treatment approach that is unhelpful, inappropriate and potentially harmful for patients who suffer a neurological injury. McAllister and Flashman advise that even minor functional changes in individuals can result in emotional

distress as the injured person attempts to meet the demands of their environment. In this section the specific diagnostic challenges that confront clinicians working with patients suffering a TBI is discussed.

Symptomology

Determining whether symptoms presented during an assessment are reflective of physical injury, neurological damage or psychopathology is a key challenge. For example if a patient reports apathy, fatigue or confusion, is this due to neurological dysfunction or does these symptoms reflect the presence of a depressive disorder, or a combination of both. If neurological, then treatment may consider waiting to see if the natural recovery process alleviates the discomfort. If psychological then CBT addressing the difficulty adapting to the altered functional capabilities may provide improved psychological functioning. Efficient use of limited therapeutic resources is necessary to contain the already high community costs from TBI. To achieve this goal accurate diagnosis is vital as it will assist in determining which individuals suffering a TBI would likely best benefit from psychotherapy.

Cripe (1999) identifies a key clinical challenge is trying to understand the role of a variety of factors in the presentation of a patient with TBI. These influences include neurological disorder, psychological disorder, and the relative contribution of both. Additionally the presentation of symptoms is affected by psychological reactions to adjustment difficulties, premorbid psychological factors along with the role of poor motivation and malingering. The potential for neurological damage to influence self report is an added challenge to the diagnostic process (Caplan & Shechter, 1995; First, 2005). Some of these factors can be concatenated into the single issue of whether the person suffering a TBI is further burdened by psychological distress. Only if the symptoms presented are purely neurological or malingering does the issue of psychopathology become redundant.

The key issue is how to interpret symptoms with potentially either a neurological or a psychological foundation. A common concern is that symptoms generated by neurological damage are often misconstrued as psychological (Alfano et al., 1993; Caplan & Shechter, 1995; Gass & Russell, 1991; Rosenthal et al., 1998). It is likely the underlying neurological damage associated with TBI can present as symptoms of psychopathology. However, the research reviewed has shown that numerous patients with a TBI are suffering the very psychopathology

those symptoms indicate. A clinician must be wary of dismissing the presence of such indications of psychopathology as a predictable reaction to TBI (Al-Adawi et al., 2007; Gualtieri & Johnson, 1999).

Diagnostic instruments

Along with the difficulty surrounding how to interpret symptoms of psychopathology is the related quandary of what diagnostic instrument to employ. The difficulty is how to interpret measures that are based on psychiatric or mental health samples when utilized on patients who suffer a TBI (Kreutzer et al., 2001). Critics of the DSM, and its various editions, generally point out the overlap in symptomology amongst disorders and that the manual was not specifically designed for assessment of TBI patients (Jorge, Robinson, & Arndt, 1993; Kendler, Gardner, & Prescott, 2002). The MMPI-2 and the Hospital Anxiety and Depression Scale are two other frequently used diagnostic instruments whose appropriateness for patients with a TBI is also controversial (Alfano et al., 1993; Gass, 2009; Hiott & Labbate, 2002).

The primary controversy is that misinterpretation of neurological symptoms in individuals with TBI may generate spuriously inflated scores on diagnostic measures (Nelson et al., 1989). When reviewing the challenges for clinicians who work with TBI patients Cripe (1999) cautions that assessment must consider the limitations of a test when interpreting the results. The implication is that a clinician should consider the various mechanisms that may lead to measured symptomatology and reflect on the tests limitations when interpreting results. Knapp and VandeCreek (2006) recommend using specific tests that are appropriate for the patient being diagnosed or implementing empirically supported alterations to the diagnostic process for the individual patient.

These issues are central to this thesis. The key issues that will be covered in more detail in Chapter Two are as follows. The MMPI-2 is regularly employed during assessment of persons suffering a TBI. The perceived limitations of the MMPI-2 have resulted in the creation of a non-standard procedure for assessment of psychopathology in persons suffering a TBI. Included amongst these non-standard procedures is the Gass (1991) correction procedure. This procedure has received conflicting reviews from researchers (Arbisi & Ben-Porath, 1999; Brulot et al., 1997; Edwards et al., 2003; Gass & Wald, 1997; Glassmire et al., 2003; Rayls et al., 1997). Importantly there is a widely recommended procedure to evaluate the construct validity of a

diagnostic instrument across different populations. This procedure is measurement invariance analysis and has not been previously completed on the MMPI-2 with a TBI sample. More details on measurement invariance and how this approach can assess the validity of the Gass (1991) correction procedure are provided in Chapter Three.

Summary

This chapter highlights the immediate and enduring impacts from TBI on sufferers. Specifically the literature reviewed emphasizes the high risk to individuals of experiencing psychological distress and developing various psychological disorders. A variety of psychopathology was found at higher prevalence rates in samples of TBI patients compared to community rates with mood disorders specifically implicated. Noteworthy was that the severity of TBI was correlated with the development of psychopathology but surprisingly injuries at the milder end of the spectrum produced higher levels of psychopathology.

A range of conceptualisations that consider the neurological and psychosocial factors involved in the development of psychopathology were discussed. A neuropsychosocial framework which recognises the importance of neurological symptoms in creating the environment in which the psychosocial influences are specifically implicated in the presentation of psychopathology was preferred. Importantly the suggestion that psychopathology is an expression primarily from neurologic damage is inconsistent with the greater levels of psychopathology found in the mild TBI samples compared with the severely injured. The diminished insight into the impacts of a TBI on a sufferer when the injury is severe, compared with a mild injury, was provided to explain the injury severity paradox.

It is suggested that the Stress-Appraisal-Coping model appropriately identifies the mediating cognitive processes that can be targeted for psychotherapy. Additionally the Stress-Appraisal-Coping model recognises the importance of insight in commencing a personal evaluation of environment needs and individual capacity which may lead to psychological distress. CBT is an appropriate therapeutic technique that can be applied concurrently with an understanding of psychopathology derived from the Stress-Appraisal-Coping model.

Psychotherapy is necessary for better rehabilitation outcomes when psychopathology is present comorbid with a TBI. However, the specific challenge of diagnosing psychopathology accurately

is an impediment to the appropriate allocation of limited resources available for TBI rehabilitation. More importantly misdiagnosis can deny a person suffering a TBI appropriate treatment. This can lead to prolonged experiences of psychological distress which can further hamper rehabilitation and in the worst case scenario lead to suicide.

Chapter Two – MMPI-2

2.1 Overview of the MMPI-2

The MMPI-2 is one of the most widely used tests of personality and psychopathology in both clinical and research settings (Archer & Newsom, 2000; Butcher, Rouse & Steven, 1996; Smith, Gorske, Wiggins & Little, 2010). First published in 1942 and revised in 1989, the MMPI-2 is an empirically developed questionnaire containing 567 items that require True or False responses which are scored on 10 clinical scales and numerous supplementary scales to assess various forms of psychopathology. The approach used to construct the MMPI clinical scales is known as the criterion-keying method whereby items are selected based exclusively on their ability to predict the criterion (Strauss & Smith, 2009). When developing the original MMPI, the authors analysed over 1000 items assembled from existing personality inventories, psychiatric texts and clinical experience. A sample of questions was generated from this list and responses from psychopathological groups were compared with responses from a normative sample (Greene, 2000). Items were included in the MMPI when responses by the group with psychopathology (criterion group) were divergent from responses by the normative group. The various clinical scales reflect the different criterion groups and the understanding of psychopathology during the 1940's. Scale elevations were interpreted based on reference values obtained from a normative sample (Butcher, Lim, & Nezami, 1998). Importantly there was no requirement that items represented any theoretical relationship to measured psychopathology.

Published in 1989, the MMPI-2 is an update to the original test. Scores were restandardised to appropriately represent ethnic minorities and item content was updated where required, but the same clinical scales and assessment approach was maintained from the original MMPI (Greene, 2000). Along with removal of items, 82 questions were rewritten in the MMPI-2 to remove awkward phrasing, refine content and modernize the language (Butcher et al., 1998; Helmes & Reddon, 1993). Content scales and supplementary scales were developed under the original MMPI and are available to provide more information for clinicians when interpreting the MMPI-2.

When using the MMPI-2 for the first time it becomes clear the scale terminology no longer reflects current psychopathological nosology. For the novice user this immediately raises

concerns about the utility of the MMPI-2. Additionally clients often generate elevated scores on multiple scales. At first glance, interpretation of profiles with multiple elevations on scales reflecting unfamiliar constructs appears confusing. However, the strength of the MMPI-2 has been the employ of code types to inform diagnosis, and the substantial body of research surrounding the use of these code types. A code type is a two or three character key that reflects the structure of elevated clinical scales. These code types are the source of a wealth of clinical information (Caldwell, 2006).

The clinical scales of the MMPI-2 are heavily loaded with a single factor representing non-specific psychological distress. This single factor is suggested to be an outcome of the criterion-keying methodology of comparing clinical groups with a 'normal' group (Tellegen et al., 2003). Another likely influence underlying a single overarching factor is that many of the MMPI-2 items form part of the scoring criteria on multiple clinical scales. The single factor has been labelled demoralisation and reduces the discriminant validity of the MMPI-2 clinical scales (Frank, 1974; Tellegen et al, 2003).

The demoralisation factor is considered a weakness of the MMPI-2 and in an attempt to rectify this weakness Tellegen and colleagues (2003) embarked on a project to restructure the clinical scales. The key aims of the restructured clinical scales (RC) were to remove item overlap and create restructured scales that reflect the core components of the existing clinical scales. This procedure was expected to generate new restructured clinical scales with greater divergent validity enhancing the clinical utility of the MMPI-2. The project created a new clinical scale named 'Demoralisation' which aimed to represent the overarching first-factor found. Additionally eight new clinical scales were developed which aimed to represent the core construct of each existing MMPI-2 clinical scale after removing the demoralisation component. MMPI-2 scales 5 and 0 were excluded from the project. Finally when developing these nine new clinical scales, items were allocated to only a single scale, thus removing item overlap. Reviews of the RC project largely support the success in achieving these goals, however the debate over the enhanced clinical utility of the RC continues (Avdeyeva, Tellegen, & Ben-Porath, 2012; Bolinskey & Nichols, 2011; Calabrese, Rudick, Simms, & Clark, 2012; Finn & Kamphuis, 2006; Greene, Rouse, Butcher, Nichols, & Williams, 2009; Nichols, 2006; Rogers, Sewell, Harrison, & Jordan, 2006; Rouse, Greene, Butcher, Nichols, & Williams, 2008; Scholte, Tiemens, Verheul,

Meerman, & Egger, 2012; Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006; Tellegen, Ben-Porath, & Sellbom, 2009; Tellegen et al., 2006).

Reitan and Wolfson (1997) provide an overview of the previous research into the interaction between neuropsychological deficits and emotional distress which includes a section specifically discussing these issues in relation to using the MMPI. The key points from their review are that concurrent neuropsychological deficits and emotional distress are not uncommon occurrences. Additionally, regardless of severity, people who have suffered a head injury often have many physical and emotional complaints. The authors note that research finds brain-damaged groups repeatedly produced elevated clinical scale scores, especially those related to anxiety and mood disorders. However no differences in scores are observed based on the location of the injury or damage. Furthermore scores are not significantly different to groups of neurotic subjects. The review notes that head-injured groups are characterised by primary elevations on Hy1, D2 and Hy3 and a secondary elevation especially on Sc8. The authors suggest MMPI profiles likely reflect stress associated with impairment of the person's adaptive capability rather than any intrinsic aspect of the neurological damage.

The criterion-keying method is central to a common criticism of the MMPI-2, one that remains despite the improvements noted under the RC project. Namely the atheoretical nature of the item selection process diminished the utility of the MMPI-2 in specific populations. Due to the atheoretical nature of item selection critics suggest that responses to specific items can reflect physical or neurological symptoms of an illness or injury, yet endorsement of the items is interpreted as reflecting psychopathology (Nelson et al., 1989). This ambiguous item endorsement can generate elevated profiles which are interpreted as reflecting psychological distress when the elevation may be the results of physical or neurological distress. Concern regarding the different explanations for item endorsement has led to the validity of the MMPI-2 being questioned for patients suffering from TBI, chronic fatigue syndrome, a stroke, amnesia and epilepsy (Alfano, Neilson, Paniak, & Finlayson, 1992; Bachna et al., 1998; Gass, 1991; Gass & Lawhorn, 1991; Johnson, DeLuca, & Natelson, 1996; Nelson, Elder, Groot, Tehrai, & Grant, 2004). More coverage of the research into assessment of individuals suffering a TBI is provided later.

In addition to the clinical scales the MMPI-2 also provides multiple validity scales that assess the

test-takers attitude at the time of assessment. Validity scales measure the propensity to answer true or false, inconsistent responses and whether the responses reflect a desire to present oneself with more or less psychological distress. These scales are widely regarded as a particular strength of the MMPI-2 and have become important component in both clinical and medicolegal settings (Arbisi & Butcher, 2004; Greve, Bianchini, Love, Brennan, & Heinly, 2006).

The MMPI-2 is an example of using self-report measures to glean diagnostic information. The benefits of this approach include reducing misinterpretation of patients subjective reports of symptomology; self-report provides the opportunity for a patient to disclose information they may omit during a face-to-face interview; and self-report is found to be an effective and reliable approach to obtaining important clinical information (Gass, 2006; Helmes & Reddon, 1993). These benefits in conjunction with validity measures of the test-takers attitude, and the extensive coverage of psychopathology by the MMPI-2, demonstrate the clinical gain from using a diagnostic instrument. Additionally the extensive research into the components of the MMPI-2 provides professional confidence when employing the instrument in clinical settings.

As noted above (see section 1.7 Treatment for persons who have suffered a TBI) there is evidence that the MMPI-2 is a valuable instrument for guiding therapy. While the research was indefinite, the capacity for the MMPI-2 to provide both clinical information and personality details was identified as a particular advantage of using the MMPI-2 prior to engaging in therapy. The identification by the APA Task Force on Promotion and Dissemination of Psychological Procedures (1995) of the need for cross-validation in all populations of interest for a diagnostic instrument highlights the importance of validating the MMPI-2 for use with persons who have suffered a TBI.

2.2 MMPI-2 Endorsement patterns in people with a TBI

There is no prototypic MMPI-2 profile for individuals suffering a TBI with studies finding significant variability in the structure of elevated clinical scales (Berry et al., 1995). However, a general trend appears with clinical scales Hy1, D2, Hs3, Pt7 and Sc8 producing higher, although not necessarily clinically elevated scores (Berry et al., 1995; Gass & Wald, 1997; Reitan, & Wolfson, 1997; Youngjohn, Davis, & Wolf, 1997). Alfano, Neilson, Paniak and Finlayson (1992) found that with the exception of scale Si0, all the clinical scales in a TBI sample were

elevated compared with non-injured samples. The finding by Alfano and colleagues may have tapped into the generalised factor of non-specific distress which loads onto all clinical scales. The authors identified a trend for elevated scales Hs1, Hy3 and Sc8 in TBI samples and suggested the scale scores were potentially increased due to the neurologic content of some items.

However, the higher score on the clinical scales Hy1, Hs3 and Sc8 may reflect the psychological distress associated with individuals suffering a TBI. Arbisi, Ben-Porath and McNulty (2003) reviewed the MMPI-2 profiles of a male psychiatric inpatient sample (n=1213) and found code types 68/86, 78/87, 13/31 were among the five most frequently occurring. These findings suggest amongst a male psychiatric sample elevated scale of Hy1, Hs3 and Sc8 are common. With males being the most prominent gender represented in TBI samples, and the accepted higher rates of psychopathology amongst TBI samples, elevation on MMPI-2 clinical scales Hy1, Hs3 and Sc8 is not surprising. Greene (2000) provides a guide as to the clinical meaning of MMPI-2 code types. A review of this information found repeated examples of code type interpretations using scales Hy1, Hs3 and Sc8 that are appropriate for the increased prevalence of anxiety and depressive disorders in the TBI population. For example code type 1-2 characterised by anxiety and dysphoria; code type 1-3 characterised by mild emotional distress, worries and dysphoria; code type 2-3 worries, dysphoria and anhedonia; code type 6-8 is characterised by severe emotional distress, anhedonia, agitation, worry and dysphoria. Importantly Sc8 is a measure of unusual or bizarre ideations which are symptoms of a psychotic disorder. Previously discussed was the virtually non-existent research into psychotic symptoms after a TBI when Alfano et al. (1992) published their study. The studies since this date suggest psychotic symptoms after a TBI may be more prevalent than previously believed (Fujii, 2005; Malaspina et al., 2001). These findings might explain the elevated rates on Sc8 identified as unexpected by Alfano and colleagues (1992). Therefore one interpretation is that the elevation in scales Hy1, Hs3 and Sc8 may accurately represent the types of psychological distress noted to occur at increased levels in people who suffered a TBI.

Injury severity paradox and the MMPI-2

Previously discussed were the findings that individuals with a mild TBI are at a greater risk of developing psychopathology than those who suffer more severe injuries. This injury severity paradox is also found on MMPI-2 profiles where the more severely injured score lower in

clinical scales, especially on those scales (Hs1, D2, Hy3, Pt7 and Sc8) identified as more commonly elevated by sufferers of TBI (Cripe, 1999; Kurtz et al., 2007; Youngjohn et al., 1997). Kurtz et al. (2007) also find a similar paradoxical effect on the Personality Assessment Inventory for patients with TBI, suggesting the MMPI-2 findings reflect real psychological distress in TBI groups rather than any artefactual bias. The strength of the MMPI-2 is the additional information available to a clinician exemplified by the various validity scales. The validity scales are especially valuable to investigate whether poor motivation, or malingering, is related to the injury severity paradox.

In this dissertation patient insight has been proposed as a mechanism to explain the injury severity paradox using the Stress-Appraisal-Coping model. To reiterate, the reduced awareness of deficits after injury in a severely injured person minimises their experience of psychological distress and this is reflected in MMPI-2 profiles. Others have suggested a more parsimonious explanation is that MMPI-2 profiles reveal the personalities and psychopathology of patients with a mild TBI rather than any influence from insight (Youngjohn et al., 1997). These two positions are not mutually exclusive under the Stress-Appraisal-Coping model. That is, regardless of injury severity, reduced insight generates a different appraisal of the objective deficits encountered after injury which leads to the personalities and psychopathology reflected in MMPI-2 profiles.

Kurtz and colleagues (2007) propose that insight is unrelated to the severity paradox. The authors suggest that as validity scales from the Personality Assessment Inventory and MMPI-2 reflect the ability of individuals with a TBI to provide accurate self-report then diminished awareness is unrelated to the severity paradox. The finding that patients with TBI and their relatives provided concordant rating on personality assessment instruments is interpreted to support their argument (Kurtz & Putnam, 2006). However, this conclusion may misconstrue the central thesis of the role of insight under the Stress-Appraisal-Coping model in the presentation of psychopathology. Under Stress-Appraisal-Coping it is not that responses are invalid but that a lack of insight into deficits actually reduces psychological distress and therefore lowers MMPI-2 scale scores. A lack of insight, if it reduces psychological distress can still be the explanatory variable for the paradox because it reflects the true experience of psychological distress in a patient. Therefore a consistent rating between patients and relatives does not invalidate this proposition.

The role of litigation has also been proposed as explaining the injury severity paradox. There are some individuals who when seeking compensation choose to exaggerate or fabricate psychological symptoms and disability (Bianchini, Greve, & Glynn, 2005; Greve, Bianchini, Love, Brennan, & Heinly, 2006). It has been suggested that malingering for the purpose of seeking compensation may be more likely in those with a mild TBI compared with those suffering more severe injuries (Kurtz et al., 2007). The regular use of MMPI-2 profiles as evidence of a patient's psychological state in medicolegal proceedings further highlights the importance of this proposition. Research with the MMPI-2 has found that assessing whether individuals with a TBI are feigning symptoms is best completed using the validity scales (Dearth et al., 2005). Therefore if malingering is the reason behind the injury severity paradox, and validity scales can detect malingering, then validity scales should be able to identify the level of injury severity in a TBI sample. Tests of the new validity scales designed under the Restructured Clinical Scales project found these scales were not able to predict injury severity in a TBI sample, whilst other scales continued to generate scores reflecting the injury severity paradox (Youngjohn, Wershba, Stevenson, Sturgeon, & Thomas, 2011). This finding fails to support the proposition that the likelihood of feigning symptoms for financial gain differs across the TBI severity spectrum. Therefore the injury severity paradox is not explained by compensation seeking behaviour. This conclusion is strengthened by research that finds the restructured validity scales as well suited to assess symptom validity in neuropsychological evaluations (Hoelzle, Nelson, & Arbisi, 2012; Tarescavage, Wygant, Gervais, & Ben-Porath, 2013).

Studies using the MMPI-2 scales also find compensation seeking behaviour cannot explain the severity paradox. Youngjohn et al. (1997) profiled 30 patients with mild TBI and 30 with moderate/severe TBI using the MMPI-2. All 30 of the mild group and 18 of the moderate/severe group were involved in litigation. One could speculate that if malingering behaviour is equivalent across the severity spectrum then finding more litigation in the mild TBI group provides some implicit support for the role of litigation in the severity paradox. Additionally differences on clinical scales were found between the litigating and non-litigating severe TBI groups, supporting the role of litigation in item endorsement. However, the study found the mild TBI group had significant elevations on scales Hy1, D2, Hs3 and Pt7 over both the litigating and non-litigating severe TBI groups. The authors concluded that with clinical scale scores in the litigating mild TBI being greater than scores in the litigating moderate/severe TBI group, the presence of litigation cannot fully account for the injury severity paradox. The authors suggest

only those who perceive themselves as significantly damaged or disabled are likely to pursue litigation, or that the litigation process is inherently stressful and will exacerbate symptoms.

Similarly Hoffman, Scott, Emick, and Adams (1999) used the MMPI-2 to profile 55 mild TBI and 57 moderate/severe TBI individuals. Like Youngjohn et al. (1997) they also found an interaction between injury severity and litigation status. Importantly they observed that elevations on MMPI-2 scales for individuals involved in litigation were independent of TBI severity. Hoffman et al. concluded that the stress of the litigation process may generate scale elevation on MMPI-2 profiles, which further supports the Youngjohn et al. (1997) conclusion.

For example, Thomas and Youngjohn (2009) found a lack of effort during cognitive testing was correlated with higher levels of symptom reports on MMPI-2 clinical and validity scales in a sample of TBI litigants. Measurement of effort was used as a measure of malingering, and the authors reported a non-significant relationship between TBI severity status and a measure of potential malingering. This would initially appear to remove malingering as a potential explanation for the injury severity paradox. However, Thomas and Youngjohn (2009) also noted that clinical scale scores were not significantly predicted by TBI severity after controlling for malingering status. Together this pattern of findings provides mixed evidence regarding malingering as a potential mechanism underpinning the injury severity paradox, as acknowledged by the authors. This paper highlights the uncertainty around explaining the injury severity paradox. In the dissertation the focus will be given to how the Stress-Appraisal-Coping mechanism can explain the injury severity paradox. This should not be interpreted as suggesting that malingering is without potential value as an explanatory variable. In addition it should also be noted that MMPI-2 cases were only included in the dissertation if the patients did not show invalid profiles, including any indication of malingering.

Described above were the concerns that the neurologic content in some MMPI-2 items may lead to these items being endorsed for reasons other than the psychopathology they are interpreted to identify. However, there is an inherent incongruity between the injury severity paradox, and the proposition that MMPI-2 scales are inflated in TBI profiles due to the neurologic content of items. If the neurologic items were the source of spuriously inflated clinical scales, then it would be expected that the more severely injured generate higher scores on these scales. That is, the

greater the severity of the TBI, the more severe the neurologic damage. The increase in damage would increase the likelihood that an item reflecting neurologic content would be endorsed. As a consequence, the MMPI-2 scores would be higher. Research which finds evidence for the injury severity paradox, fails to support this expectation. Therefore a different explanation for the distribution of clinical scale scores is required and notwithstanding the issue of malingering, the Stress-Appraisal-Coping model better explains the distribution of MMPI-2 scores in non-malingering TBI populations.

While psychological distress has been presented as a key variable in understanding the injury severity paradox, the relationship between individual MMPI-2 scales and psychological distress varies considerably. Almagor and Koren (2001) found that while D2, Hy3, Pd4, Pa6 and Sc8 appear to tap into distress, Ma5 did not. The authors also note that some of the content scales, namely DEP, WRK, ANX and A were highly correlated with a measure of distress. In a study of compensation seeking veterans elevated scores on the F scale (a validity scale) were identified as being related to extreme distress without intentional over-reporting of symptoms (Franklin, Repasky, Thomson, Shelton, & Uddo, 2002). Alternatively the F(p) validity scale is found to be a valuable addition to the F scale alone in identifying individuals who are experiencing psychological distress (Arbisi & Ben-Porath, 1997). The correlation between the MMPI-2 clinical scales and the 'Demoralization' scale in the RC project has been observed to range from .35 to .92 (Simms, Casillas, Clark, Watson, & Doebbeling, 2005). 'Demoralization' was defined as representation of the overarching single factor of distress from the original MMPI which supports the finding by Simms et al. as illustrating the variance in the relationships between MMPI-2 clinical scales and distress.

Summary

The injury severity paradox for persons suffering a TBI is replicated in research with the MMPI-2. The increased risk of psychopathology in persons who have suffered a TBI is consistently reflected in elevation of MMPI-2 clinical scales Hs1, D2, Hy3, Pt7 and Sc8. The additional information available on MMPI-2 profiles allows investigation of the mechanisms that may underlie the presentation of psychopathology in samples of TBI sufferers. Compensation seeking may increase scale scores but this could be due to the stresses associated with the litigation process. Whilst malingering is a concern, the validity scales of the MMPI-2 serve as an appropriate device to determine the motivations of persons completing the TBI. The proposition

that neurologic content in some MMPI-2 items generates inflated scores on the clinical scales is contrary to the injury severity paradox. The specific role of insight under the Stress-Appraisal-Coping model provides an understanding of the injury severity paradox reflected in MMPI-2 profiles. The complex and speculative nature of the issues described above merit analysis with rigorous methods, such as measurement invariance testing.

2.3 Specific MMPI-2 diagnostic challenges with TBI patients

The atheroetical nature of the MMPI-2 design process was previously highlighted as resulting in items with potential neurologic content being included (Alfano, Finlayson, Stearns, & Neilson, 1990; Gass, 1991; Gass, 1992; Nelson et al., 1989). The inclusion of these items has resulted in uncertainty regarding the applicability of the diagnostic instrument when used with person's suffering a TBI. The confusion and contradictory advice as to how the MMPI-2 should be utilised with patients suffering a TBI is illustrated in the Textbook of Traumatic Brain Injury (Silver, McAllister, & Yudofsky, 2005). On the one hand, the chapter discussing personality disorders strongly advises against interpretations from a MMPI-2 profile (O'Shanick & O'Shanick, 2005). On the other hand, the chapter discussing chronic pain describes the MMPI-2 as a valuable source in understanding patient's emotional distress and coping styles (Zasler, Martelli, & Nicholson, 2005). A common reminder from critics is that while the MMPI-2 is well validated and easy to administer, it was never developed to measure psychopathology in people with neurologic injury or illness (Nelson et al., 2004).

The primary rationale for raising concerns with the atheoretical nature of the MMPI-2 is that profiles for TBI patients are an artefact of the test items content reflecting neurological symptoms commonly occurring after injury (Gualtieri & Johnson, 1999). This occurs because individuals who have suffered a TBI may acknowledge the neurological symptoms of their injury when completing the MMPI-2 regardless of any relation between these symptoms and psychopathology (Gass & Wald, 1997). That is the neurological symptoms associated with a TBI are reflected in the content of some items and as such accurate endorsement of these items by brain injured patients will generate falsely inflated scale scores.

This line of argument was proposed as early as 1970 by Taylor who asserted that physical disability moderates item endorsement probability which potentially reduces the MMPI validity.

However, as discussed earlier people suffering a TBI are at increased risk of developing psychopathology. As such, the general recommendation is to exercise caution when interpreting MMPI-2 profiles due to the possibility of inflated scale scores, whilst recognising the injury suffered significantly increases the likelihood that the measured psychopathology may exist (Gass, 2009).

The finding that clinical scales Hs1, D2, Hy3, Pt7 and Sc8 are often inflated in MMPI-2 profiles of TBI patients, and that these scales have items that refer to neurological symptoms, is provided as evidence to support concerns with the validity of MMPI-2 in TBI patients (Bachna et al., 1998; Gass, 1991). Gass (2009) reviewed the MMPI-2 and identified the content of 27 items as representing typical neurological symptomology (see Table 2.1). These items were represented across scales Hs1, D2, Hy3, Pt7 and Sc8 with no items from the remaining scales identified as problematic. Gass suggested that the inflated scores for individuals with a TBI on scales Hs1, D2, Hy3, Pt7 and Sc8 are related to the content of the items identified.

Notwithstanding the potential influence of item content on profiles of TBI patients it is important to remember that the symptoms described are not incompatible with psychopathology (Bachna et al., 1998). Fishbain et al. (2006) completed a meta-analysis on the influence of pain on a variety of measures, including the MMPI-2. They found treatment of pain resulted in improved MMPI-2 scores. In relation to diagnosis of psychopathology using the MMPI-2, many of the studies included were using clinical scales and changes in these scores may be interpreted as reduced psychopathology with effective treatment. These findings support the proposition that physical and psychological symptoms exist concurrently and the treatment of one may reduce the experience of the other.

Gass (2000) recognises that some clinicians propose psychological distress, rather than neurological injury, is the reason for endorsement of MMPI-2 items identified in Table 2.1. He criticizes this interpretation for assuming patients only endorse an item due emotional distress, not neurologic factors. Gass points out that valid response to MMPI-2 items require acknowledging symptoms regardless of the existence or absence of psychological distress. The implication from his criticism is that an item is a valid measure only when psychological distress is the sole reason for item endorsement, and there is no overlap between neurological and psychological symptoms.

It is reasonable to consider that multiple causes may influence whether or not a person endorses an item. This is seen in the fact that some MMPI-2 items load onto multiple scales making the endorsement of an item relevant to the constructs being measured on multiple scales, including validity scales. Gass (2000) identifies three factors after injury that can result in personality changes and altered emotional functioning 1) psychological reaction to change in functioning 2) damage to specific neuroanatomical regions related to mood and 3) difficulties in long term adjustment to the individuals post injury environment. These factors at a minimum would be expected to influence item endorsement.

The items in the MMPI-2 are designed to assist diagnosing psychopathology, and at a minimum, the three factors identified by Gass (2000) influence psychological functioning. All these mechanisms identified by Gass as altering the emotional functioning of a person are potentially valid components of MMPI-2 responses which include, but are not limited to, neurological damage. It is important to remember the strength of the MMPI-2 is not the endorsement of a single item, but the aggregation of multiple items to produce clinically informative scale scores.

Correction procedures are discussed in more detail below. Briefly, a correction procedure removes specific items from the MMPI-2 that have been identified as potentially introducing bias in profile scores. In their review of the Stress-Appraisal-Coping model and TBI patients Godfrey, Knight and Partridge (1996) note a common response is distress about neuropsychological symptoms. If distress about neurological symptoms is a typical response, and stress increases the risk of developing psychopathology, then endorsing the presence of neurological symptoms by TBI patients may accurately reflect the increased risk for psychopathology. However a correction procedure assumes endorsement of items reflecting neurological symptoms will falsely inflate scores of psychopathology. This assumption is incongruent with the research by Godfrey and colleagues.

Rayls et al. (1997) propose that item endorsement would vary based on duration post injury. That is, endorsement of some of the items identified in Table 2.1 reflects acute neurologic symptoms that are likely to resolve following mild head injury. Whereas endorsement of the same items after the acute phase, is most likely related to psychological factors. This raises the difficulty for a clinician to differentiate between an acute and chronic presentation of symptomology. Incorrectly differentiating leads to the risk of missing psychopathology in persons suffering a

TBI. This can lead to denying psychotherapy and the associated benefits for TBI patients in their process of rehabilitation post injury.

Notably, a limitation in the research literature is the shortage of validation studies completed on the MMPI-2 when used with persons suffering a TBI (Arbisi & Ben-Porath, 1999). The situation has not been rectified in the decade since this original statement. Interestingly, Gass (2000) identifies this limitation yet he concludes this is why clinicians should be wary of using the standard MMPI-2 with persons suffering a TBI. The implication is that a non-standard version of the MMPI-2 that corrects for the neurologic bias is more appropriate. If validation studies are required before using the uncorrected MMPI-2, then the same criterion should be applied to use of non-standard or corrected version. As will be shown the research has yet to produce the necessary findings to support using a non-standard MMPI-2 with persons suffering a TBI. In fact much of the research suggests a non-standard approach diminishes the MMPI-2 validity when used with TBI patients (Brulot et al., 1997; Dunn & Lees-Haley, 1995; Edwards et al., 2003).

2.4 Correction Procedures

The aim of a correction procedure is to remove any items that diminish the clinical validity of the MMPI-2 and restore the test to, or as close as possible to its original effectiveness. Nelson and colleagues (2004) describe two approaches to follow when generating a correction procedure. The first, named a ‘rational procedure’, utilises expert knowledge to guide selection of items to be removed according to established criteria, such as symptoms representing the disease or medication side-effects. The second is called a ‘statistical procedure’ where items are selected without regard to content provided those items can differentiate subjects with the specific condition from normal controls. From this set of items a review is completed to determine which items likely reflect the specific medical condition (to be removed), and which likely reflect true psychopathology (to remain). A limitation for both procedures is that at some point they rely on clinical intuition. Neither approach can be regarded as a purely empirical exercise which leaves both open to subjective bias.

Generating culturally appropriate norms is considered a fundamental step when translating the MMPI-2 for different languages (Butcher et al., 1998). Similarly, generating TBI population specific norms is necessary for the use of any MMPI-2 TBI correction procedure (Arbisi & Ben-

Porath, 1999). One of the enduring gaps in knowledge regarding the correction approach to the MMPI-2 has been the dearth of analysis to ascertain whether the norms for the MMPI-2 can continue to be applied for the correction procedure or if these alterations reduce the validity of the *T*-scores for the various scales. If the goal of correction procedure is to improve the validity of a test then it is vital to remember "any departure from standard administrative practice changes the meaning of scores (Cronbach, 1960, p.185)".

Clinicians are faced with a dilemma when interpreting correction procedure *T*-scores for the MMPI-2. The dilemma is whether the risk from abandoning the standard administration procedures is sufficiently offset by enhanced clinical validity (Caplan & Shechter, 1995). This risk is exacerbated by the failure to recalibrate *T*-scores to account for the reduced number of items within a scale. The dilemma would be best served by clinical research rather than intuition. However, the MMPI-2 TBI correction procedure research provides conflicting results (Arbisi & Ben-Porath, 1999; Brulot et al., 1997; Edwards et al., 2003; Gass & Wald, 1997; Glassmire et al., 2003; Rayls et al., 1997).

Caplan and Shechter (1995) suggest assessment using correction procedures requires a blend of solid technical skills and 'judicious creativity'. By 'judicious creativity' the authors are endorsing a balanced use of nonstandard methods to complete evaluations. This position assumes the nonstandard methods will provide test scores that can better inform treatment than standard practice. The further one moves away from the standard application of a test instrument the further they move away from evidence-based practice.

Looking at guides on how to use the MMPI-2 it is apparent that a clinician should consider the individual patient's particular circumstances when interpreting profiles. The clinician's role is more challenging that just summing numbers to produce a profile. While this is true it is imperative a clinician respects the substantial research completed into understanding MMPI-2 scale scores and code-types. This understanding should be at the forefront of a clinician's mind when they consider using a non-standard interpretation of a MMPI-2 profile. The removal of items from scales has the potential to reduce scale validity as readily as the potential confound of neurologic symptoms. An assessment of the impact from a TBI on the psychometric properties of MMPI-2 items can provide valuable information. Such an assessment would help define what, if any, impact exists from including items with neurologic content.

Table 2.1

MMPI-2 items proposed by Gass (2009) to reflect potential neurologic content

		MMPI-2 Scales				
Item	Content	Hs1	D2	Hy3	Pt7	Sc8
10	work capacity			X		
31	distractibility		X	X		X
45	general health	X				
47	pain			X		
53	paresthesias	X				
106	speech changes					X
142	convulsions		X			
147	reading difficulty		X		X	X
148	general health			X		
152	tiredness & fatigue	X		X		
164	dizzy spells	X		X		
165	memory difficulty		X		X	X
172	Tremor			X		
173	tiredness & fatigue			X		
175	weakness	X	X	X	X	
177	paralysis or weakness					X
179	difficulty walking	X	X			X
224	Pain	X		X		
229	blank spells					X
247	numbness	X				X
249	Vision			X		
255	Tinnitus					X
295	paralysis or weakness					X
298	anosmia					X
299	distractibility					X
308	forgetfulness				X	
325	concentration difficulty				X	X

There is a risk of eliminating important items that measure a common condition comorbid with a TBI because the item also refers to neurologic content (Arbisi & Ben-Porath, 1999). The concern for the MMPI-2 TBI correction procedures is the removed item also measures the types of psychopathology known to be regularly comorbid with a TBI. A correction item may be endorsed due to a variety of reasons such as head injury, an emotional response to the injury, premorbid psychopathology or a combination of these factors (Glassmire et al., 2003). Removing MMPI-2 items may reduce the capacity for a clinician to understand the role of psychological factors in item endorsement. As recommended by Gass (2009) a prudent approach to using the MMPI-2 with a patient suffering a TBI requires methodical consideration of the possible impact on profiles from endorsement of neurologically sensitive items. A comparative analysis of the MMPI-2 psychometric properties in a TBI group and a non-injured group would provide important information lacking in the literature to assist in this methodical consideration. Examination of measurement invariance can provide this information as it assesses the MMPI-2's validity when applied to patients suffering a TBI.

Gass (1991) Correction Procedure

The most commonly used MMPI-2 correction procedure for TBI patients is the one proposed by Gass (1991). This section reviews the method followed to select the items for removal under this procedure. Furthermore the research into the Gass (1991) correction procedure is critiqued.
The Gass (1991) correction procedure was designed using the MMPI responses from 75 TBI patients (70 male and 5 female) compared with the 1,138 normal adult men in the normative sample on which the MMPI-2 is based. In what appears to be an attempt to control for confounding psychopathology, the TBI sample was screened to ensure there was no evidence of a pre-injury psychological disorder or substance abuse. Additionally patients were excluded from the study if they had other known neurological disorders, medical diagnoses or invalid MMPI profiles. The TBI sample was noted to be skewed towards recent injury with 52% of participants within 1 year post injury.

The Gass correction procedure was developed using what Nelson et al. (2004) termed a 'statistical procedure'. In the first stage item endorsement frequencies were compared across samples. The percentage of subjects in the normative sample who responded in the scored direction was multiplied by 75, the sample size of TBI patients. This allowed a comparison of the endorsement frequencies across the two samples.

In the second stage, principal components analysis (PCA) was applied to determine which items differentiated the normative sample from the TBI sample. Items were selected based on two criteria: (a) the degree of discriminative power ($p < .001$), and (b) endorsement frequency in the TBI sample (endorsement frequency $\geq 25\%$). Applying these two criteria resulted in 23 items being extracted for further analysis. PCA with an oblique rotation generated a 2 factor solution with each of the 23 items loading > .30 on the designated factor. Factor-1 has 14 items which explained 24.8% of variance and Factor-2 has 5 items that explain 3.7% variance. Factor-1 was labelled Neurologic Complaints items and their corresponding MMPI-2 items are 31, 101, 106, 147, 149, 165, 170, 172, 175, 179, 180, 247, 295 and 325. Factor-2 was labelled Psychiatric complaints and the MMPI-2 items are 10, 32, 229, 241 and 307. Interestingly items 149, 170 and 180 from Factor-1 were not included in the items identified by Gass (2009) as involving neurologic content (see Table 2.1) but were included in the correction procedure. Whereas item 229 in Factor-2 and not included in the correction procedure was included in Table 2.1.

The correction procedure recommends scoring the MMPI-2 profile of a person with TBI twice if necessary (Gass, 1991, 2009). On the first occasion the profile is scored according to the standard procedure and reviewed for signs of psychopathology. If the profile indicates psychological distress then the profile is rescored with the 14 items from the Neurologic Complaints factor removed. It is expected that a person suffering a TBI who is experiencing psychological distress would continue to generate an elevated profile despite the removal of items from a clinical scale. As previously noted one oversight with this approach is the failure to update the *T*-score calculation procedure to allow for the reduced number of items in the clinical scales (Arbisi & Ben-Porath, 1999).

A review of the Factor-1 and Factor-2 items identified by Gass (1991) shows that the items allocated to the Neurologic Complaints factor refer to content that may reflect neurologic symptoms post a TBI. However some of items in the Psychiatric Complaints factor are not as clearly differentiated from the Neurologic Complaints factor. Item 229 refers to blank spells which interrupt activities and cause confusion, while item 32 refers to strange and peculiar experiences. The content of these two items could be equally representative of neurologic symptoms as psychiatric symptoms. The content of these items alone does not invalidate the correction procedure as these items were allocated to the factor which explained less variance but it does highlight the challenge in demarcating a neurologic from a psychiatric symptom.

Interestingly Gass (2006) wrote that correction items are somewhat sensitive to various types of psychopathology and should be interpreted cautiously in individuals who have premorbid or comorbid emotional disturbances or substance abuse. He recommends in these cases if the correction procedure results in no, or a minor change, in scale scores, then the clinician can be confident in the validity of elevated scores on scales Hs1, D2, Hy3, and Sc8. This recommendation is somewhat confusing. If the content of an item reflects neurologic damage after a TBI then it would be expected to be endorsed. Additionally, if the item measures psychopathology in the person with a TBI it would again be expected to be endorsed. However, Gass suggests a clinician interpret a profile using the complete set of MMPI-2 items only when items with neurologic content are not endorsed. The result would be no change in scale score when applying the correction procedure, providing confidence in the complete MMPI-2 scores. This seems an unlikely scenario as the item now reflects both valid neurological and psychological symptoms.

Research supporting Gass (1991) correction procedure

The Gass (1991) correction procedure is supported by some replication studies that have attempted to control for some of the limitations in the original study. These studies will be discussed below.

The high proportion of males and the process of recruiting the sample from a veteran affairs medical centre are limitations of the procedure used to generate the Gass (1991) correction procedure. Gass and Wald (1997) further queried whether the TBI sample was not representative of the acute cases more commonly seen in clinical practice, by having individuals suffering moderate to severe injuries with an average post-injury time of 4.1 years. The severity of injury is worth controlling for in a study because, as was noted in Chapter One, mild TBI is associated with greater levels of psychopathology. As 52% of the sample was assessed within 1 year after the TBI this reduces concerns about duration since injury. However, concerns with the gender imbalance are warranted.

Gass and Wald (1997) cross-validated the Gass (1991) correction procedure by analysing the MMPI-2 profiles of 54 TBI patients (28 male, 26 female) who were referred to an outpatient service for neuropsychological assessment. The paper describes the ‘vast majority’ of the TBI sample as mildly injured with the average time post injury being 24.2 weeks. Thus the study has

selected a more representative sample which accounts for some of the limitations in the original paper. Gass and Wald describe the purpose of the paper to establish: 1) whether items with neurologic content represent a major source of variance in MMPI-2 profiles from the TBI sample; 2) the reliability of the items from the Gass (1991) correction procedure in differentiating this TBI sample from a normative sample; and 3) the clinical importance of the Gass (1991) correction procedure items as defined by their frequency of endorsement in the pathological direction. To facilitate these goals the endorsement frequency of items from the TBI sample was compared with that of the full MMPI-2 normative sample (n=2600).

The analysis found 44 items that were considered effective discriminators between the two samples ($p<.001$). Ten of the top 15 items that discriminated the two samples were also part of the 14 item pool previously identified by Gass (1991). For both genders 13 of the 14 Gass (1991) correction procedure items were found to be effective discriminators between samples. Item 172, which refers to frequent hand shaking, was the item which failed to differentiate between samples. Items 39, 40, 229, 308, 299 comprise the additional five items in the top 15 that were not included in the Gass (1991) correction procedure. Item 229 was part of Factor-2, labelled Physical Complaints, in the original study. When reviewing the design method for the correction procedure, this item was identified as one which also likely refers to neurologic symptoms.

Gass and Wald (1997) suggest the content of the 15 items found as the top discriminators between samples in their study reflect neurologic rather than psychiatric symptoms. Notwithstanding conjecture about the assumed mutual exclusivity of these categories, the items do appear to reflect neurologic content. One apparent exception is item 39 which refer to fitful and disturbed sleep. This item reflects content that explicitly relates to criteria for diagnosing Major Depressive Disorder, Generalised Anxiety Disorder and Post Traumatic Stress Disorder. Across the two papers at least 13 of the 14 items from the Gass (1991) correction procedure have been deemed to act as effective discriminators between TBI and control samples (Gass, 1991; Gass & Wald, 1997). The TBI samples have covered both genders, a range of severity, inpatients and outpatients, and varied durations of time since injury. Additionally premorbid psychopathology was controlled for as a potential confound when creating the Gass (1991) correction procedure. This still leaves uncontrolled the impact of current psychopathology on the selection of correction procedure items.

To investigate this issue Gass and Wald (1997) completed a second study to investigated what were the top 15 items to differentiate the normative sample from a sample of 524 psychiatric inpatients. It was concluded that as none of the 15 items from this second study were also part of the Gass (1991) correction procedure, this indicates that the TBI sample and psychiatric samples differ in item endorsement. The implication of this finding is that psychopathology does not reflect the endorsement patterns found by the TBI sample and therefore the correction procedure reflects items selected only because of their neurologic content.

The problem with this conclusion is that the item endorsements between the psychiatric and TBI sample may vary due to differences in psychopathology prevalence rates. Gass and Wald (1997) noted that in the psychiatric sample the breakdown of psychopathology were depressive disorders (26%), schizophrenia and other psychotic disorders (36%), bipolar disorder (9%), adjustment disorders (10%) and other disorders (19%). Clearly psychotic disorders are overrepresented with the highest prevalence rates for developing psychosis post TBI at 9.8% (Davison & Bagley, 1969). The representation of depressive disorders in the sample is at the lower end of the studies reviewed earlier, while the absence of any noted anxiety disorders is unrepresentative of TBI samples. Therefore whilst the desire to investigate the role of comorbid psychopathology with TBI on the endorsement patterns is admirable, the psychiatric sample appears inadequate to control for potential confounds.

Gass (2009) refers to a study by Edwards, Holmquist, Wanlass, Wicks and Davis (1998) as supporting the specificity of his TBI correction procedure. Edwards and colleagues examined whether endorsement by individuals suffering a TBI of the items in the Gass (1991) correction procedure reflects either premorbid psychopathology or an effect of the injury. To answer this question each study patient and a "significant other" informant were interviewed for their opinion on whether item endorsement reflected an effects of the patient's TBI. The study found 109/119 endorsements were attributed to the head injury "supporting the specificity of the correction when applied to TBI patients". However, again the study failed to control for post injury psychopathology. Thus whilst this study may conclude that the endorsement of the correction items is most likely due to the effects of the TBI, this does not preclude one of the effects being newly developed psychopathology.

Gass (2006) cites a conference paper in which an incidental finding was that TBI injury severity,

as represented by the Glasgow Coma Scale (GCS) scores, were positively related to endorsement frequency of the items comprising the Gass (1991) correction procedure (Gass, Luis, Rayls, & Mittenberg, 1999). This finding was interpreted to support the correction procedure. This conclusion is presumably because the increasing endorsement frequency, along with increased severity logically reflects the expectations that the more severely injured would present with higher frequencies of neurological symptoms. This finding is interesting because when items are measuring psychopathology, the injury severity paradox would expect lower frequencies of endorsement as TBI severity increases. Thus the neurological explanation for item endorsement appears to be disentangled from the psychopathological explanation in this paper. Gass et al. (1999) appears to provide some empirical support for the Gass (1991) correction procedure, although being unpublished it is not available for closer inspection of the research methodology. Rayls, Mittenberg, William and Theroux (1997) studied the longitudinal effects of mild TBI on endorsement of Gass (1991) correction procedure items. They found immediately following injury (range 1- 15 days) TBI patients endorsed more correction items than the MMPI-2 normative sample. When the TBI patients were contacted on average 8 months after injury, none of the items were endorsed at frequencies higher than the normative sample. Additionally patients endorsed significantly fewer items than immediately after their injury. The authors concluded the correction items reflect acute neurological symptoms likely to dissipate following mild TBI, whilst chronic endorsement of the items likely reflects psychological factors. This is an interesting proposition which the data appears to support particularly as research has found higher levels of measured psychopathology using the MMPI-2 with increased duration post TBI (MacNiven & Finlayson, 1993). These findings are reflected in the Gass (2006) rule that a clinician does not need to apply the correction procedure with uncomplicated mild head injury patients. A difficulty with this rule is that the MMPI-2 may be required to determine whether any comorbid psychopathology is complicating a mild head injury.

Research not supportive of Gass (1991) Correction procedure

Along with the studies described above that support the Gass (1991) correction procedure there is also a body of research that suggests using the correction procedure is inappropriate. This section reviews some of these papers.

In a forensic sample of 191 adults, Brulot, Strauss and Spellacy (1997) assessed the validity of the Gass (1991) correction procedure in three sets of analyses. In the first set they found no

significant association between the correction factor and measures of injury severity (loss of consciousness or post traumatic amnesia). This finding suggests that increased neurologic damage is not associated with increased endorsement of the 14 correction procedure items. The failure to find an association between injury severity and item endorsement is in contrast with the unpublished data described above that was provided to support the correction procedure (Gass et al., 1999). Secondly, no significant associations were found between the correction factors and performance on neuropsychological tests, which fails to support the hypothesis that neurological damage is specifically reflected in endorsement of correction items. Finally, Brulot and colleagues found significant correlations with correction procedure items and the MMPI-2 depression content scale. The depression content scale was used in the analysis as it contains none of the Gass (1991) correction procedure items. Brulot et al. acknowledge skewness in their study towards mild TBI is a limitation. However the natural distribution of injury severity is skewed towards the mild end of the spectrum, so this may not be such a significant qualification on the quality of the study (Rapoport et al., 2002). Brulot et al. concluded that the Gass (1991) correction procedure may be more related to depression than to TBI. Therefore they argue against using the Gass (1991) correction procedure.

Edwards et al (2003) scored the MMPI-2 twice, once with all items and a second time applying the Gass (1991) correction procedure. These two scoring procedures were completed on both a TBI sample (screened to ensure no normal level profiles) and a psychopathological sample. When applying the Gass (1991) correction procedure the number of profiles that moved to normal levels was recorded as was the number of 2-point code type changes. The frequency of code type changes was compared across groups with the expectation that if the Gass (1991) correction procedure is specific to neurologic damage, then significantly more changes should be observed in the TBI sample compared with the psychopathological sample. The analysis found no significant difference between the samples. Additionally across both samples only one participant ended up with no clinical scale elevation after applying the Gass (1991) correction procedure. The authors conclude that the Gass (1991) correction procedure is unnecessary as there was no clinical impact from removing the correction items from the MMPI-2. They explain that despite observing a number of code-type changes in the TBI sample when completing the Gass (1991) correction procedure, as these changes were also found in the clinical sample, the profile alterations from the procedure are not specific to TBI. The authors re-iterate that validity of the MMPI-2 is compromised once items are removed because existing norms and validity

studies of *T*-scores and code types are no longer relevant.

An important finding in the study by Edwards et al. (2003) was the high proportion of participants in the TBI sample (43%) who had different two-point codes for their MMPI-2 profile after the correction procedure. It was noted above that one of the strengths of the MMPI-2 is the wealth of well researched clinical information provided by these two-point codes (Caldwell, 2006). Therefore the findings by Edwards and colleagues suggest that by applying the Gass (1991) correction procedure, while it may not alter an individual scale score from a diagnostic to a non-diagnostic range, correction may change the two code profile. Such a change may alter the interpretation of the MMPI-2 for a particular patient.

Dunn and Lees (1995) assessed utility of the Gass (1991) correction procedure in a forensic setting. All participants were undergoing forensic neuropsychological evaluations with 59 suffering a TBI and 102 non-TBI participants. Endorsement rate of the 14 Gass correction items was compared across groups, with differential response frequencies found on only five of the items (147, 165, 179, 247 and 295). The paper reviewed the mean *T*-score differences on the clinical scales Hs1, D2, Hy3 and Sc8, comparing results using either the Gass (1991) correction procedure, or the five items found by Dunn and Lees. Using these five items identified by Dunn and Lees produced no significant profile changes. Using the Gass (1991) correction procedure *T*-score changes ranged from 4.7 on scale Sc8 to 3.0 on scale Hs1. Unfortunately no significance tests were detailed. The paper concludes that the Gass (1991) correction procedure is not recommended in a forensic setting, an important consideration in view of the widespread use of the MMPI-2 in medico-legal cases.

Glassmire et al. (2003) reviewed MMPI-2 responses for three groups: a TBI sample (n=218), a psychopathological sample (n=656), and, the MMPI-2 normative sample (n=2600). The endorsement frequency of Gass (1991) correction procedure items was calculated for each participant. Receiver operating characteristic analysis was utilised to evaluate the sensitivity and specificity of the Gass (1991) correction procedure in identifying persons as being either from the TBI sample, the psychopathological sample, or the normative sample. The Gass (1991) correction procedure produced strong sensitivity, discriminating TBI participants from the normative participants, but poor specificity for discriminating TBI participants from psychopathological participants. These results suggest the correction items are effective at

selecting participants with a TBI from those without a TBI. But the items are poor at differentiating participants with a TBI from participants with psychopathology. This occurs despite the fact the TBI sample was screened for premorbid psychological disorders. Therefore the endorsement of correction items by the TBI sample may be reflecting true psychopathology. These results lead the authors to argue against use of the Gass (1991) correction procedure.

Furthermore the theoretical proposition that the physical and neurological symptoms associated with a TBI will reduce the validity of the MMPI-2 item endorsement has many critics (Glassmire et al., 2003; La Chapelle & Alfano, 2005). Glassmire et al. explain that an assumption underlying the correction procedure is that item endorsement is solely due to the effects of a TBI. Yet endorsement may reflect other factors such as premorbid psychopathology, an emotional response to the injury, generalised distress, or a combination of these factors. Cripe (1999) adds malingering as another potential influence on item endorsement. Importantly, due to the atheoretical nature of the MMPI-2, as long as items can help identify psychopathology in a person with TBI, the content of the items is unimportant.

Alfano and colleagues Correction procedure (1991; 1993)

It would be remiss not to mention that there is another MMPI-2 correction procedure proposed for use with persons suffering a TBI. Alfano, Finlayson, Stearns and MacLennan (1991) applied the 'rational' process to generate a correction procedure that consists of 44 items. The MMPI was reviewed by 18 specialists in clinical neuroscience to identify items they believed represented valid neurologic damage or dysfunction. They compared MMPI profiles of 115 patients suffering neurologic disease with 109 non patients and concluded the endorsement rates of the identified items suggest a clinician should to be wary of the impact of neurologic content items when interpreting MMPI profiles. A limitation of the study is that only 43% of the neurologic disease sample had suffered a TBI.

Alfano and colleagues (1993) rectify this limitation by reviewing their 44 item correction procedure with a sample of 102 patients suffering a TBI. To refine the model they completed a two-step process. Initially they identified 24 of their correction items that were endorsed in the pathological direction by at least 30% of the sample. Next they completed Principal Components Analysis and found a 2-Factor solution accounted for 33.2% of total variance with Factor-1 comprising 13 MMPI-2 items that accounted for 25% of total variance. Factor-1 was labelled

'Neurobehavioural' and included items 10, 12, 31, 38, 106, 147, 177, 179, 180, 247, 295, 299 and 325. Similar to Gass (1991) they suggest a correction procedure should be applied to a second scoring of the MMPI-2 with the removal of the 13 Neurobehavioural items identified. Despite the similar number of items included in both correction procedures the specific items used in each procedure differs considerable. Only eight items are included in both the 13 item Alfano and colleague's procedure and the 14 item Gass (1991) correction procedure.

Research into the Alfano and colleagues 13 item model is limited but on the whole not supportive. Brulot, Strauss and Spellacy (1997) found no association between injury severity and the Alfano and colleagues (1993) correction procedure items and conclude the model should be avoided when diagnosing psychopathology in persons suffering a TBI. Glassmire et al. (2003) found the Gass (1991) correction procedure performed better than the proposed model by Alfano and colleagues. Nevertheless the authors conclude neither model is recommended as they cannot discriminate between item endorsement from samples of individuals suffering a TBI and one of individuals suffering psychopathology. Edwards et al. (2003) only assessed the full 44 item neurocorrection procedure and again found no significant difference between profile changes in the TBI and psychopathological groups.

Summary

Conflicting research findings into the Gass (1991) correction procedure leave clinicians working in TBI populations unsure as to the appropriateness of the procedure. There are two key limitations with the papers that support the Gass (1991) correction procedure. The first is controlling for the comorbidity of psychopathology and TBI. Some papers have controlled for premorbid psychopathology but have failed to consider the development of psychological disorders and distress post injury. Secondly no papers have been able to ascertain the impacts on scale validity as a result of removing items. Gass, Luis, Rayls and Mittenberg (1999) found a positive correlation between injury severity and endorsement of correction item which supports the proposition that neurologic symptoms are primarily responsible for item endorsement. However the lack of detail about sample selection prevents a more thorough review of the findings. Additionally the replicability of the finding is questionable as Brulot, Strauss and Spellacy (1997) found no relationship between correction procedure item endorsement and injury severity.

Alfano and colleagues (1993) have also proposed a correction procedure. Interestingly despite being derived from the same theoretical foundation, the items included in this correction procedure were seen to differ markedly from the items in the Gass (1991) correction procedure. The research into the applicability of the Alfano and colleagues (1993) correction procedure is predominantly unsupportive (Brulot et al., 1997; Edwards et al., 2003; Glassmire et al., 2003).

Clinical application of the MMPI-2 profiles necessarily requires the blend of solid technical skills and consideration of the unique characteristics of a patient when interpreting profiles. However fundamental in this approach is informative research. The conflicting conclusions from the reviewed research suggest further analysis of the Gass (1991) correction procedure is required. Noteworthy is that none of the studies assessed impacts on the psychometric characteristics of the MMPI-2 when applied to individuals suffering a TBI. Testing the measurement invariance of a model which facilitates such an assessment is now considered fundamental when using diagnostic instrument in different populations (Muniz & Bartram, 2007).

Chapter Three – Measurement Invariance

3.1 What is measurement invariance?

Measurement invariance is the condition whereby the psychometric properties of latent constructs being measured by a test are equivalent across different populations. For the purpose of this paper the test is the MMPI-2, and the populations are persons with and without a TBI. Examination of measurement invariance allows a direct test of the hypothesis that in different populations the same latent constructs underlie test scores and the same metric relationship exists between these observed and latent variables (Bowden et al., 2008). For the MMPI-2 this means psychopathology is measured equivalently for a person regardless of whether or not they have suffered a TBI. Establishing measurement invariance does not establish the same causal factors underlie the responses, but it does establish that the responses equivalently measure the same psychological constructs.

A full latent variable model comprises both a structural model and a measurement model (Byrne, 1998). Measurement invariance testing is primarily concerned with the measurement model which represents the relationship between the latent constructs and the observed measures. Measurement invariance testing follows a sequential approach where model parameters are allowed to be freely estimated across groups in the first stage, and parameters are constrained to equality in subsequent stages of analysis. When model fit is equivalent despite the additional imposition of parameter equality across populations, then invariance is established. When the observed measured are dichotomous items, such as with the MMPI-2, there are three parameters in the measurement model that are required to be equivalent to establish strict measurement invariance; thresholds, loadings and residuals (Millsap & Yun-Tein, 2004). Additionally, with dichotomous data, the results of the measurement invariance testing with theta parameterisation are equivalent to a test of differential item functioning under item response theory (Glöckner-Rist & Hoijtink, 2003).

When all measurement parameters meet the requirements for invariance this is known as strict measurement invariance or strict invariance. Technically a violation of invariance occurs when one or more parameters in the factor model differ across the populations of interest (Millsap, 2005). Consequently the reliability or validity of a test differs across groups of people when

violations of measurement invariance are found. If some measurement model parameters fail and some parameters meet the requirements for invariance this condition may satisfy what is known as partial measurement invariance or partial invariance (Brown, 2006; Meredith, 1993; Widaman & Reise, 1997).

Measurement invariance is relevant to any diagnostic instrument that is utilised in multiple populations (Millsap, 1997). Meredith (1993) has shown that strict measurement invariance is important to ascertain equality in classification decisions across different groups. Therefore measurement invariance is a test of whether the MMPI-2 scores can be interpreted to diagnose psychopathology with equal accuracy for both persons with and without a TBI. Failure to find measurement invariance on a test implies construct bias across the populations, which suggests psychopathology is measured differently for persons in different conditions (Kline, 2010). For this study, measurement invariance tests are completed to ascertain whether the MMPI-2 measures psychopathology equivalently for persons with and without a TBI. Importantly if failure to find strict invariance is observed, then analysis can identify the specific items that are the source of variance. Marsh and colleagues (Marsh et al., 2013; Marsh, Nagengast, & Morin, 2010) have applied measurement invariance techniques to assess the big-five factor structure over lifespan and across genders. However, a review of the literature was unable to find further research that employed measured invariance to assess a multi-trait measure of personality or psychopathology such as the MMPI-2. This lack of research may reflect the lack of awareness of measurement invariance techniques amongst applied personality researchers, together with the fact that the methods of measurement invariance are cumbersome and challenging to implement for test with many items, especially categorical items (Marsh et al., 2010).

How does measurement invariance answer the research question?

If the Gass (1991) correction procedure is required for persons suffering a TBI then the MMPI-2 would fail a test of measurement invariance and specifically the 14 items identified by Gass would be the sources of variance. Furthermore completing measurement invariance analysis allows a more thorough examination of the core issue. That is whether the MMPI-2 is appropriate in its current format for diagnosing psychopathology in a person suffering a TBI. Along with assessing the Gass (1991) correction procedure, measurement invariance analysis examines whether any items on the scales analysed generate a bias in the MMPI-2 profiles for persons with a TBI.

Measurement invariance tests of the MMPI-2 in a TBI population are necessary regardless of the confusion around the use of the Gass (1991) correction procedure. José Muniz and David Bartram are members of the International Test Commission and the European Federation of Psychologists' Associations Standing Committee on Tests and Testing respectively. Together they reviewed research and projects of the two bodies in regards to best practices for psychological assessment (Muniz & Bartram, 2007). Findings from their review include that: invariance of diagnostic instruments should be established before completing comparisons across different populations; and, test developers should use appropriate statistical techniques to establish equivalence of the test instruments for all intended populations. Furthermore APA guidelines recommend some validity evidence should be part of test development and usage (American Education Research Association, 1999). Therefore even if the validation studies unanimously rejected the Gass (1991) correction procedure, measurement invariance analysis with a TBI sample is still required to support the MMPI-2 for this population.

Establishing measurement invariance for the MMPI-2 in a TBI population is necessary because the scores on clinical scales are used to inform diagnosis and whether treatment for psychological distress is required. Borsboom (2006) writes a commentary on the clinical importance of measurement invariance tests. He explains that invariance is required to assert that differences in observed scores for different populations reflect real variation in the theoretical attributes being measures. Without a test of measurement invariance a clinician cannot infer that differences in MMPI-2 item endorsement frequencies are due to differences in psychopathology. There is clear evidence that psychopathology is more frequent in survivors of a TBI compared with the general population. Yet without measurement invariance analysis of the MMPI-2 one cannot conclude whether different item endorsement frequencies reflect these different prevalence rates or some other phenomena, such as neurological injury symptoms after injury.

The reason for failure of measurement invariance is often difficult to ascertain as it may be due to the accumulation of minor inconsequential parameter differences. This is a distinct possibility in tests with as many parameters as the MMPI-2 clinical scales (Byrne & van de Vijver, 2010). However, violations of measurement invariance have important practical consequences when these violations contribute to inaccurate selections of which persons may benefit from psychological therapy (Millsap, 2005). This statement includes a subtle alteration to the previous discussion of invariance. Thus far invariance has been referred to as either existing or failing to

exist, with failure implying the test is inappropriate for a particular condition. Millsap redefines the focus of invariance testing to consider the objective of a test. In practical terms Millsap sets the issue of clinical importance as whether failure of invariance results in reduced diagnostic accuracy. Therefore, according to Millsap's argument a test that fails to satisfy the requirements of measurement invariance across different populations, may still be used provided there is none, or minimal, practical impact when failure of invariance occurs.

Borsboom (2006) suggests that technically measurement invariance will be violated in all conditions. That is, because technically any difference in the psychometric properties of a measurement parameter between groups of interest, represents failure of invariance. Borsboom notes the precise equivalence of measurement model parameters across two samples is unlikely. Therefore every diagnostic tool is biased to some degree, and empirically the situation is exacerbated with larger sample sizes. When diagnostic instruments are tested for measurement invariance in clinical settings the more common finding is one of partial invariance (Millsap & Meredith, 2007). Partial invariance is the situation where some but not all of the model parameters meet the requirements of invariance. The challenge for researchers is that if one accepts the condition of partial invariance what does this mean? Are the numbers of parameters that fail invariance too small to be a concern? In essence is the violation of invariance meaningful? We can define a meaningful violation of invariance as one that interferes with the purpose of the measure (Millsap & Meredith, 2007).

3.2 Assessing the impact of partial invariance?

Schmitt and Kuljanin (2008) completed a review of measurement invariance papers published since 2000. In this review they observed that half of the 75 papers that studied measurement invariance concluded with a finding of partial invariance. The repeated finding that partial invariance exists for tests across conditions demonstrates the difficulty in establishing strict invariance as described by Borsboom (2006). However an assumption underlying the retention of non-invariant items discovered when determining partial invariance is that these items do not excessively affect cross-group comparisons (Cheung & Rensvold, 1999). Millsap and Kwok (2004) outline a procedure which investigates the practical impact on a diagnostic instrument when the instrument failed the test of strict invariance. Sensitivity and specificity values can be calculated and reviewed to determine the practical impact on diagnostic accuracy when non-

invariance is found. Specificity is the probability that a test of psychopathology will correctly diagnose a patient as healthy when no psychological disorder exists (Glassmire et al., 2003). Sensitivity is the probability the test will correctly diagnose a patient as having a psychological disorder when psychopathology does exist (Glassmire et al., 2003).

This approach is recommended in the review by Schmitt and Kuljanin (2008) who also note the lack of invariance research that has implemented the Millsap and Kwok (2004) procedure. A search of PsycINFO of the papers published since 2008 fails to reveal any invariance studies using the Millsap and Kwok practical analysis to understand the clinical impact of a partial invariance model.

3.3 Research Questions

Factor analysis will be completed on the MMPI-2 clinical scales investigated for measurement invariance. There is a paucity of factor analysis studies published for the individual MMPI-2 clinical scales. The two factor model found for D2, the Depression scale is an exception (Chang, 1996). Quereshi and Kleman (1996) completed factor analysis on the three validity and ten clinical scales, but these were combined into a single analysis with no results provided that describe the factor structure of the individual clinical scales. Furthermore, other factor analytic studies of the MMPI-2 were completed using item parcels rather than individual items in the analysis. Little, Cunningham, Shahar and Widaman (2002) identify that parcels may better reflect the constructs being measured compared to individual items. However, Little et al. explain a disadvantage of parcelling is that the process may obscure misspecification that would be found if analyses were performed on item-level data. Importantly the authors note the pitfalls of parcelling are most evident when attempting to identify the exact psychometric relationship amongst the individual items.

As will be discussed below a clearly defined factor model for the MMPI-2 clinical scales is necessary to commence measurement invariance testing. Therefore it is necessary to define a factor model for each of the clinical scales being assessed for measurement invariance. Additionally as measurement invariance is concerned with the performance of individual items, factor analysis will be completed using individual items rather parcelling items together.

Measurement invariance analysis will assess the appropriateness of the MMPI-2 in its current format when used with persons suffering a TBI. If strict invariance is found then it can be concluded that MMPI-2 items do not bias profiles for persons suffering a TBI. Importantly strict invariance would establish that items with content referring to neurologic symptoms associated with a TBI are valid inclusions in the MMPI-2. If strict invariance is found it can be concluded the Gass (1991) correction procedure for TBI patients is unnecessary.

If failure of invariance is found then the analysis will attempt to define a partial invariance model. The partial invariance model procedure will free the minimum number of parameters necessary to meet the invariance requirements. The backwards elimination procedure described by Cheung and Rensvold (1999) will be employed during this stage. This procedure facilitates determining whether any items in the Gass (1991) correction procedure are a specific source of variance. Specifically, this is concluded if measurement parameters corresponding to the items in the Gass (1991) correction procedure are not equivalent across groups when defining the partial invariance model. If no items from the Gass (1991) correction procedure are defined as a source of variance this is further evidence against the correction procedure.

If a partial invariance model is found, the Millsap and Kwok (2004) sensitivity and specificity analysis will follow. This procedure allows a practical impact analysis of the failure to find strict invariance. Reviewing the sensitivity and specificity values can answer the question as to whether the violation of strict invariance has clinical impact (Millsap & Meredith, 2007). If the impact is negligible, then failure of invariance is concluded to reflect the view that most diagnostic instruments contain some bias across populations, but this alone should not deem the instrument unsuitable (Borsboom, 2006). If items in the Gass (1991) correction procedure are a source of variance, yet this variance is found result in no practical impact, this would provide evidence against applying the correction procedure.

The primary goal of the measurement invariance analysis is to assess the Gass (1991) correction procedure. However, the conclusions will provide important information regarding the underlying theory that neurological content in items falsely inflated the MMPI-2 profile in people with e TBI. An understanding of the neurological content hypothesis can inform as to the mechanisms that lead to the injury severity paradox. In particular the analysis of the items in Gass (1991) correction procedure will provide information about the role of insight in the injury

severity paradox. Support for retaining the 14 items from the correction procedure would suggest the injury severity paradox found in MMPI-2 profiles is not confounded by neurological item content.

By removing the potential confusion that profiles are inflated by neurological damage, rather the reflecting psychopathologies then the potential mechanisms underpinning the injury severity paradox become clearer. Indirectly this would permit the contention that the injury severity paradox reflects the role of insight as defined in the Stress-Appraisal-Coping model, although there are no direct tests of this hypothesis. Finding indirect support for the Stress-Appraisal-Coping model in understanding psychopathology amongst persons with TBI is valuable as the Stress-Appraisal-Coping model provides important information which can assist the therapeutic process.

Because of the computational and fitting challenges of modelling item level data, particularly with large numbers of items, this study will model items from individual clinical scales of the MMPI-2 and test for invariance between a TBI sample and the MMPI-2 normative data (Little et al., 2002). The Hs1, Hy3 and Sc8 scales were selected for measurement invariance analysis. These three scales were selected for a variety of reasons. Most importantly using Hs1, Hy3 and Sc8 scales provide complete coverage of the items in the Gass (1991) correction procedure. Additionally these three scales were identified as being excessively inflated in TBI samples (Alfano et al., 1992). Therefore Hs1, Hy3 and Sc8 scales are an excellent review of the neurologic content hypothesis which is integral to the generation of correction procedures. Finally these scales are three of the five scales commonly referenced in the injury severity paradox (Cripe, 1999; Kurtz et al., 2007; Youngjohn et al., 1997).

The key research questions for each scale are: (1) What item-level factor structure best represent the respective clinical scale?; (2) Does the scale meet the requirements for strict invariance across a TBI and the MMPI-2 normative sample?; (3) If partial invariance is established which items fail the test of strict invariance?; and (4) What is the practical impact from any failure to establish strict invariance?

Chapter Four – Method Section

4.1 Materials

To handle the overlap of items across clinical scales Hs1, Hy3 and Sc8, two options were considered. Firstly, to model all items in each clinical scale with 32 items on Hs1, 60 items on Hy3 and 78 items on Sc8. The second option is to assess each scale sequentially and remove items from the subsequent scales that were included as part of a preceding scale's analysis. That is to remove duplicate items which had already been assessed. The decision taken was to remove duplicate items on the subsequent scales. This is because measurement invariance analysis will have previously assessed whether any duplicated item is invariant. With dichotomous data, the results of the measurement invariance testing with theta parameterisation are equivalent to a test of differential item functioning under item response theory (Glöckner-Rist & Hoijtink, 2003). This equivalence as a test of differential item functioning further supports the decision to remove items on subsequent scale. Furthermore, the considerable computational and model fitting challenges of modelling item level data, especially with scales containing large numbers of items, further supports removing duplicate items from the longer scales (Little et al., 2002). The result was an analysis on the full 32 items from Hs1, 40 items from the Hy3 scale and 68 items on the Sc8 scale.

In preparation for practical impact analysis following the Millsap and Kwok (2004) approach, item were rescored to account for polarity. The MMPI-2 scores item endorsement of psychopathology with a combination of True and False response. Therefore it was necessary to recode the responses with 1 indicating a response that endorses the psychopathology and 0 a response not reflecting psychopathology. For example item 2, True response was scored as 0 while False was scored as 1, whilst item 18 had True scored as 1 and False scored as 0.

Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and measurement invariance testing was completed using Mplus Version 6.11. The weighted least squares means and variance adjusted (WLSMV) estimator is employed as the estimation method for analysis. Beaducel and Herzberg (2006) compared the WLSMV and maximum likelihood (ML) estimators in a simulation study and found WLSMV to be more appropriate with categorical data. Importantly for this study they found when data had two or three categories the model

rejection rates using WLSMV were closer to the alpha level expected than those using ML with similar sample sizes.

The geomin rotation is the Mplus default for categorical data using WLSMV. The geomin solution is an oblique rotation. Analysis using oblique rotations is recommended when assessing a questionnaire, like the MMPI-2, with interrelated dimensions of broader constructs (Brown, 2006) and avoiding arbitrary decisions a priori about factor correlations (Floyd & Widaman, 1995). The item to factor relationship was determined by finding the factor in the EFA analysis upon which the item loaded with the highest absolute value (e.g. -0.521 is a higher loading than 0.520). All other aspects of CFA model development followed contemporary recommendations (Brown, 2006; Hensen & Roberts, 2006; Kline, 2010).

Item levels scores from the MMPI-2 normative sample (n=2600, female n = 1462, male n = 1138; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) were used under research license agreement between the University of Minnesota Press and the University of Melbourne. While the practical impact analysis was completed using IBM SPSS v.20 (IBM Corp, 2011).

4.2 Participants

The MMPI-2 normative data comprises a representative community sample from the USA of 2600 adult men (n=1138) and women (n=1462). In the analysis of each scale (Hs1, Hy3 and Sc8) after removing cases with missing data five samples were generated from the MMPI-2 normative data. The Female Norm sample included only female participants and the Male Norm sample included only male participants. Norm A and Norm B samples were randomly generated halves of the normative sample including mixtures of males and females. Finally measurement invariance testing was completed with a Community sample of the MMPI-2 normative data that reflects the gender proportions observed in the full TBI sample (62.5% male and 37.5% female) described below. The Community sample consists of the all participants in the Male Norm sample and a random subset of the participants from the Female Norm sample.

The study also included a sample of TBI patients assessed at a private practice specialising in forensic TBI evaluations who were tested using the MMPI-2 from 1995 – 2005. This sample comprises 162 males and 97 females collected consecutively and from whom MMPI-2 data were

available. Data on TBI severity was available for 212 participants, derived from the examining neuropsychologists report and stratified according to the criteria of Williamson, Scott & Adams, (1996); 117 (55.2%) of participants had experienced mild injuries, 46 (21.7%) moderate, 32 (15.1%) moderately-severe and 17 (8%) severe injuries.

Data on time between injury and MMPI-2 assessment was available for 205 cases from the same source, and was 2.81 (s.d. = 3.9) years on average with a minimum interval of .1 year and a maximum of 40 years. The median interval was 2.00 years with an inter-quartile range of 1.65 years. In the whole sample, 24 cases experienced workplace-related injuries, 6 assaults, 4 sports or domestic injuries and the remainder were motor-vehicle or pedestrian injuries, reflecting the referral pattern in the personal injury practice from which the MMPI-2 data were obtained.

This study was granted approval by the Human Research and Ethics Committee of St Vincent's Hospital, Melbourne. Table 4.1 provides information for each sample employed with scale Hs1, Hy3 and Sc8. Apart from the TBI sample, the age demographic information was only calculated for the Female Norm and Male Norm samples as the remaining normative samples were derived from these participants in each scale.

Table 4.1

Number of participants, mean age and standard deviation for each sample employed in the analysis on scales Hs1, Hy3 and Sc8

	Hs1		Hy3		Sc8	
Sample	N	Age *(s.d.)*	N	Age *(s.d.)*	N	Age *(s.d)*
TBI	242	35.7 *(12.9)*	233	36.0 *(13.1)*	229	35.8 *(12.8)*
Female Norm	1432	40.4 *(15.1)*	1393	40.2 *(15.0)*	1376	40.1 *(15.0)*
Male Norm	1116	41.6 *(15.2)*	1104	41.4 *(15.2)*	1079	41.5 *(15.2)*
Norm A	1275		1249		1228	
Norm B	1273		1248		1227	
Community	1786		1766		1726	

4.3 Procedure to define a Baseline Model

Measurement invariance testing requires a defined baseline model which must meet the criteria for configural invariance (Byrne, 1998; Widaman & Reise, 1997). Configural invariance is established when items load onto the same latent variables across samples, indicating a similar factor structure in each sample (Widaman & Reise, 1997). Configural invariance supports the inference that similar constructs, although not necessarily identical, are found in the separate groups. In view of the paucity of research which defines factorial models on the individual clinical scales of the MMPI-2, a procedure using EFA and CFA was employed to define a baseline model. The procedure employed has been termed the 'EFA into CFA' method, to convert the best fitting, admissible EFA model into a simple structure CFA for replication across samples (Brown, 2006; Muthén & Muthén, 1998-2010).

The first step was to define a candidate baseline model using EFA and CFA in the Norm A, Norm B and TBI samples (from now on called a candidate model). Floyd and Widaman (1995) recommend that when samples are large enough, participants should be assigned to multiple groups which allow one sample to define a factor solution and the remaining samples to confirm the solution. The MMPI-2 normative data sample is large enough to allow creation of multiple samples; however the TBI sample is too small for this approach.

The next step was a replication process using CFA to analyse the performance of the candidate model in all five samples. If there is no obviously preferred model evident from examination of model fit criteria then item R-squared, factor articulation and review of item content of items for model interpretability will also be considered (Kline, 2010).

Admissibility and model fit criteria

Confirmatory factor models derived from EFA were assessed against admissibility criteria before evaluation of fit (Brown, 2006). Firstly a minimum of three indicators is required per factor for identification. Secondly, standardised item loadings cannot be greater than 1. Thirdly, to ensure model factors are clearly articulated the standardised factor covariance must be more than two standard errors less than 1. Fourthly, all items should preferably load significantly on respective factors.

Admissible models were subjected to χ^2 comparison tests for nested models. Additionally, Root-Mean-Square-Error-Approximation (RMSEA), Tucker-Lewis Index of fit (TLI), and Comparative Fit Index (CFI) were compared for both nested and non-nested models. Multiple fit indices were evaluated when assessing factor models as recommended by Vandenberg and Lance (2000) and Brown (2006). The RMSEA was required to be less than .05 which indicates the data fits the model well (MacCallum & Browne, 1993). When an acceptable RMSEA is found and if the TLI and CFI levels are close to or greater than .95 then this is indicates acceptable model fit, if the levels fall below .90 then the model will be rejected (Brown, 2006; Hu & Bentler, 1999; Kline, 2010).

Procedure for finding a candidate model in Norm A, Norm B and TBI samples

Following recommendation from Henson and Roberts (2006), and Floyd and Widaman (1995), EFA was conducted in each sample, to find the model with the least number of factors that was statistically non-significant using a χ^2 test ($p>.05$). Because EFA has a tendency to overestimate the number of factors, the model found by EFA was selected as the upper limit of the number of factors possible in the candidate model (Brown, 2006). The second step was to run a simple structure CFA using the EFA results (Muthen & Muthen 2007; Brown, 2006; Kline, 2010). The item to factor relationship was determined by finding the factor in the EFA upon which the item loaded with the highest absolute value (e.g. -0.521 is a higher loading than 0.520). The marker variable for a factor should be an item with good psychometric characteristics, and therefore the item with the highest EFA loading was selected (Bontempo & Hofer, 2007; Kline, 2010).

Candidate model respecification procedure

Once the preferred factor structure for a candidate model is determined, it is important to consider whether any model respecification is necessary. An important aspect of model refinement is to consider whether correlation between item residuals is worth specifying in the factor model. There are two types of unique variances are represented by measurement errors: random error (score unreliability) and all sources of systematic variance not due to the factors (e.g. method effects). A method effect refers to an additional correlation found between items due to the measurement approach, such as item referring to similar content. If method effects are present then defining correlations between pairs of residuals is appropriate (Kline, 2010). In CFA this method effect can be specified in the model (Brown, 2006).

The modification indices in the output will be reviewed to find possible pairs of correlated residuals. These correlated residuals will be specified in a model to be tested, provided the content of the items is similar. Similar item content is important as this supports the notion that method effects are present for these items. A respecified model will be assessed using a χ^2 difference test for significant fit improvement after inclusion of the correlated residual. If the χ^2 difference test is non-significant, this means the additional specified parameters (the residual correlations) result in no model improvement. Therefore under the preference for parsimony the factor model with the additional parameter specifications is rejected. If the χ^2 difference test is significant the respecified model is preferred as it improves the model performance.

Modification indices may also suggest items be moved from the factor on which they are allocated to a different factor. Should this occur the model will be respecified to reflect the modification indices output and the respecified model will be assessed for significant improvement in fit following the same procedure outlined above. After completion of this respecification phase the final structure of the candidate model will be defined. That is the number of factors, the allocation of items to a factor, and whether any item residuals should be correlated will be defined.

An important issue to remember when reviewing the performance of the candidate models is that frequently the model definition procedure is intended to define best possible factor model from the variables available. This process assumes the ability to exclude a variable that is detrimental to the model performance, or if necessary to reject all possible models generated with the variables provided (Kline, 2010). However, for the purpose of this study the measurement invariance testing requires the best available measurement model that utilizes all the unique items on the MMPI-2 scale. Therefore all items included for analysis on a scale will be included in the selected candidate model.

Replication procedure to compare candidate models

Once a best fitting candidate model is defined for each of the samples described above, the next step will be a replication process using CFA to analyse the performance of each candidate model in multiple samples. During this replication procedure two additional samples from the normative data were included. One is all the male participants (Male Norm) and a second is all the female participants (Female Norm). It was considered worthwhile to include these samples to

ascertain the appropriateness of a baseline model for both genders. Furthermore the additional samples will increase confidence in the selected factor model. Performance of a factor model in a variety of samples is fundamentally important (Thomson, 2004). A well replicated model is shown to fit the data well despite individual sample variance. Furthermore the admissibility of a candidate model is assessed in all samples. If a candidate model is inadmissible in any sample it is removed from consideration as the baseline model.

If there is no obviously preferred model evident from the replication procedure then further investigation is required. To assist in the identification of a preferred model the item R-squared, factor rho coefficient and review of item content for model interpretability will also be considered (Kline, 2010). Item R-squared is a measure of item reliability. Factor rho coefficient is a measure of the factor reliability. Reviewing item content is a theoretical assessment of the logical coherence of the defined item to factor structure. This review permits a comparative assessment of the theoretical strengths and weaknesses of the three candidate models.

Summary of baseline model selection procedure

EFA and CFA are employed to define three candidate models (Norm A, Norm B and TBI). EFA is used to define the upper limit on number of factors and the allocation of items to a specific factor. CFA is used to assess the admissibility of a factor model, along with the comparative goodness of fit of admissible factor models. Once the preferred number of factors and item to factor loadings is determined, the model will be evaluated for possible respecification to primarily account for method effects. Additionally the reallocation of any items to alternative factors is considered. At all times the goal is to define the best performing factor model in each sample without sacrificing parsimony.

Each candidate model is then subjected to a replication procedure in five samples. The three samples used in the model definition stage were augmented by a normative female and a normative male sample. If a candidate model fails the admissibility criteria during the replication stage then it is removed from further analysis and no longer available for selection as the baseline model. Admissible candidate models are compared for goodness of fit using RMSEA, CFI and TLI. If no candidate model is found as preferred after this procedure, then candidate models were compared on measures of item reliability, factor articulation and a theoretical review of the candidate model structure.

4.4 Procedure to complete measurement invariance testing

As mentioned above measurement invariance testing sequentially holds the parameters of the measurement model to equality across the groups of interest. With the MMPI-2, the parameters of interest are; thresholds, loadings and residuals (Millsap & Yun-Tein, 2004). Provided model fit is equivalent with each additional constraint, then invariance is established.

Measurement invariance requires clear assessment criteria to determine that model fit is equivalent despite the additional constraints of equality imposed on model parameters in the sequential tests. Three options were considered for determining requirements to meet the standard of invariance. First, if RMSEA remains below .05, a minor change to CFI or TLI, such as .01, is acceptable within the definition of an invariant measurement model (Cheung & Rensvold, 2002). Second, invariance is established if there was no significant difference in RMSEA, and the CFI decreased by less than .005 (Chen, 2007). The third option was to apply a more stringent application of the CFI decrease by .002 or less, as recommended by Meade, Johnson and Braddy (2008). The third, and more challenging, option was the approach followed in this study.

Multi-group CFA will be used for measurement invariance testing as it is the preferred methodology in multifactorial frameworks (Meade & Lautenschlager, 2004). Previous analysis has found multiple factors in Hs1 and D2 suggesting a multifactorial approach is required (Alkemade, 2007; Chang, 1996). Furthermore CFA has the ability to separately model measurement error (item residuals) enhancing appraisal of measurement reliability (Ullman, 2006). Additionally when the threshold parameter is investigated in CFA this is equivalent to investigating the difficulty parameter in Item Response Theory, that is, the test is equivalent to examination of differential item functioning (Kim & Yoon, 2011).

There will be two samples used in the measurement invariance analysis for each scale. The first group is a Community sample and the second is the TBI sample. The Community sample is created from the MMPI-2 normative data with gender proportions equivalent to those observed in the overall TBI sample. Thus reducing gender as a potential confound. This decision strengthens the interpretation of any difference across groups on tests of measurement invariance

as being due to factors specifically related to the existence of a TBI.

There are two different procedures recommended when completing measurement invariance testing using dichotomous data. Millsap and Yun-Tein (2004) recommend an approach whereby thresholds are constrained to equality for identification purposes in Test 1. In Test 2 item loadings are constrained to equality across groups, which facilitate a test of scalar invariance. In Test 3 residuals are constrained across groups to assess strict invariance. An alternative approach is recommended by Muthen and Muthen (1998-2010) where residuals are constrained in Test 1 with loadings and thresholds freely estimated. In Test 2 loadings and thresholds are constrained to equality with residuals freely estimated. This approach reflects the view that invariance of residual parameters is unnecessary to permit comparisons of mean scores across groups in a test instrument (Raykov, 2004; Widaman & Reise, 1997).

Concerns about the Millsap and Yun-Tein (2004) approach have been raised because when applied to models involving dichotomous items there is no separate testing of threshold invariance (Bontempo & Hofer, 2007). Investigating the invariance of threshold parameters is valuable because this test is equivalent to a test of the difficulty parameter in Item Response Theory (Kim & Yoon, 2011). While the approach advocated by Muthen and Muthen (1998-2007) does not include a test of strict invariance. A test of strict invariance is required to implement the practical impact analysis proposed for partial invariance models.

Therefore a hybrid approach will be employed which follows the first stage of the Muthen and Muthen (1998-2007) approach with the final stage of the Millsap and Yun-Tein approach (2004). In Test 1 residuals are constrained to equality across groups with loadings and thresholds allowed to vary (except for loadings on marker variables). In the Test 2 all residuals, loadings and threshold were constrained to equality across groups. This latter test provides a test of strict invariance. If the second test fails to meet the requirement for measurement invariance, then item parameters (loadings and thresholds) will be freed sequentially to find a model that meets the requirements of partial invariance. Test 3 is a test of partial invariance. That is, a factor model defined with the minimum number of parameters freed (or the maximum level of constraint), that meets the invariance criteria discussed above. The advantage of this approach is that by constraining all parameters in Test 2 if the model fails the test of strict invariance then assessment of the practical impact of the non-invariance could be completed following the

approach suggested by Millsap and Kwok (2004). Another advantage of this approach is that if a partial invariance model is found, the analysis can identify which items have the threshold parameters differ across group and which remain equivalent. Additionally, Vandenberg and Lance (2000) identify an advantage of employing a two-step process in invariance testing to reduce the alpha error rate associated with multiple steps in invariance analyses.

If failure to find strict invariance is found then the backwards elimination approach will be applied to determine which parameters violate the hypothesis of measurement invariance. Cheung and Rensvold (1999) raised the problem of having to set the loadings of a variable to invariance to meet identification needs in multigroup CFA. This can become problematic if the item is a source of model variance. The backwards elimination method frees parameters sequentially until the model is accepted as statistically meeting the requirements of measurement invariance. The backwards elimination approach minimizes the risk of identification problems when defining a partial invariance model (Millsap & Meredith, 2007; Yoon & Millsap, 2007). This is because the variable constrained to equality for identification can be altered should the modification indices recommend freeing the item. As only one variable loading per factor is required to be constrained for identification, this only becomes problematic should all respective item loadings requiring freeing to meet the partial invariance criteria for a factor.

Review of the modification indices will show which item parameters are to be freed (Brown, 2006; Kline, 2010). A modification indices value is provided in Mplus output for the loadings with each sample, and a single value for thresholds. As such there are potentially three modification indices provided for each item. These three values will be summed together and the item with the largest total modification indices value will have its parameters freed.

Once the parameter constraints are removed, the model is assessed using a χ^2 difference test to ascertain whether the model with the unconstrained parameters is a significant improvement on the model with the parameters constrained. If this test is passed then the updated model is re-evaluated to determine if the model meets the requirements of partial invariance. If the model fails to meet the criteria for partial invariance then additional item parameters are freed, and the process is repeated until a model with appropriate fit is retained as invariant.

4.5 Procedure to complete practical impact analysis

The following section is a synopsis of the detail described by Millsap and Kwok (2004) to assess the practical impact of a non-invariant model when partial invariance is established. The information is presented as it pertains to using the MMPI-2. Briefly to complete practical impact analysis, sensitivity and specificity values are required for the TBI sample in the strict and partial invariance condition for any factor that failed the test of strict invariance. Observed scores and factor scores are respectively used as proxy measures of diagnosis with or without psychopathology, and the presence or absence of psychopathology. These measures are necessary to calculate the sensitivity and specificity values. The following procedure is required for any factor that fails the test of strict invariance but fulfils the criteria for partial invariance.

For each factor a participant's observed score and a factor score are calculated. The factor score is calculated by Mplus, while the observed score is calculated by summing the raw item scores within each factor. The observed score represents the assessment of psychopathology on the respective scale, assuming strict invariance were to hold. Factor scores represent the estimate of psychopathology, and can be used to estimate the presence or absence of psychopathology. The Millsap and Kwok (2004) procedure requires selection of a cut-off score to differentiate those with and without psychopathology. An observed score and factor score that reflects the 93.32 percentile rank will be used to define cut-off points. This percentile rank is 1.5 standard deviations above the mean which corresponds to a *T*-score of 65. A MMPI-2 *T*-score of 65 or greater is often used to reflect a score of clinical interest. A reference sample is required to define the cut-point scores, which is the Community sample in this study (Millsap & Kwok). It should be noted that the cut-point selected is pragmatic and was chosen to both best represent a typical clinical decision process and to find cases that provide non-zero frequencies each of the four categories for analysis shown in Figure 4.1

It is expected that the distribution of observed scores will find the percentile rank of 93.32 occurring between the distribution points, for example hypothetically between observed score three and observed score four. Should this occur then an observed score of four will indicate the presence of psychopathology and score three the absence of psychopathology. The cut-off point for the factor score will be set at the equivalent point for the percentile rank on observed score

three. That is if observed score three has a percentile rank of 90.0 then the factor score at percentile rank of 90.0 is set as the cut-off point.

Therefore each person in a sample will have an observed score and factor score. Each factor will have a cut-off point in scores to reflect the presence or absence of psychopathology. While each observed scores will have a cut-off point to represent diagnosis. In traditional notation for assessing the sensitivity and specificity of a diagnostic instrument, the factor scores represent the presence or absence of the latent estimate of psychopathology, while the observed scores represent the diagnosis of psychopathology (present or absent). As such, each person is located within one of the following four categories (1) the diagnostic zone for both the observed score and the factor score (True Positive) (2) the non-diagnostic zone for both the observed score and the factor score (True Negative) (3) the non-diagnostic zone for the observed score and the diagnostic zone for the factor score (False Negative) and (4), the diagnostic zone for the observed score and the non-diagnostic zone for the factor score (False Positive). See Figure 4.1 which represents a hypothetical distribution of factor scores and observed scores into each of the four categories described with a cut-point at the 90^{th} percentile.

A count of the number of participants that are located within each category, or quadrant in Figure 4.1, is calculated. These tabulated data will be used to calculate sensitivity and specificity values for each factor in each MMPI-2 scale that fails the test of strict invariance. This procedure is repeated for the partial invariance condition and the strict invariance condition. Difference in values across the invariance conditions may occur as changes in the underlying measurement model potentially alter the calculated factor scores. Changes in factor scores can potentially alter the proportions in each category and alter the sensitivity and specificity values. The practical impact of non-invariant items is then evaluated by comparing the sensitivity and specificity values in the partial invariance condition (the clinical application of the MMPI-2) with the strict invariance condition (the desired condition of measurement model equivalence across groups). See Figure 4.2 which represents a hypothetical distribution of factor scores and observed scores into each of the four categories described with a cut-point at the 90^{th} percentile, in both a partial invariance and strict invariance condition.

The TBI population is the focus for this study therefore only the sensitivity and specificity values for the corresponding sample are required for review. An important practical impact is suggested

if, for the TBI sample the value of sensitivity or specificity in the partial invariance condition is below the lower bound of the 95% confidence interval (CI) in the strict invariance condition. The exception to this rule is when the absolute value for either sensitivity or specificity is acceptable in the partial invariance condition.

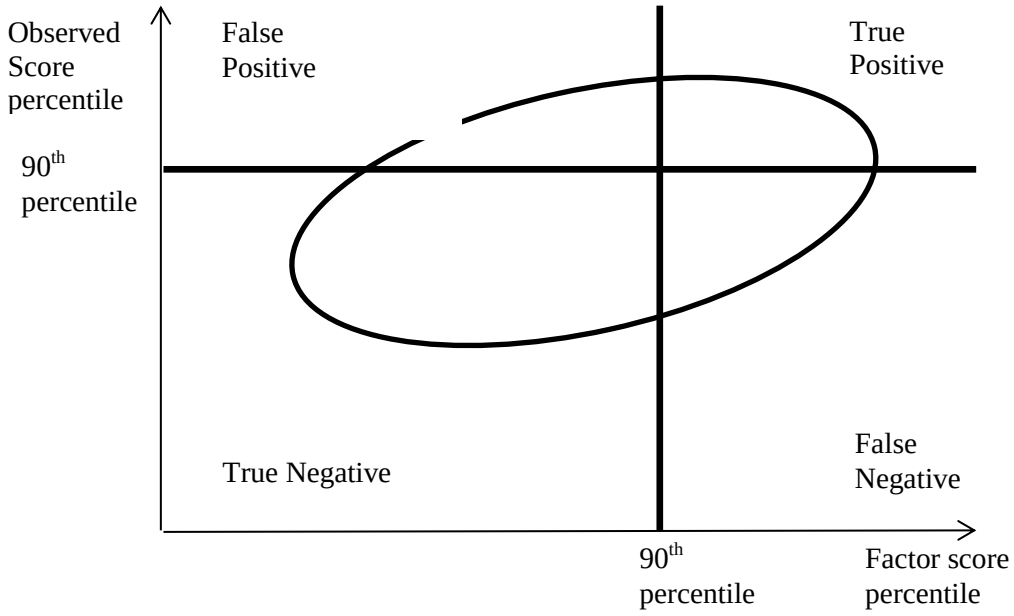

Figure 4.1

Distribution of Factor scores and Observed Scores into quadrants assuming cut-point at 90th percentile

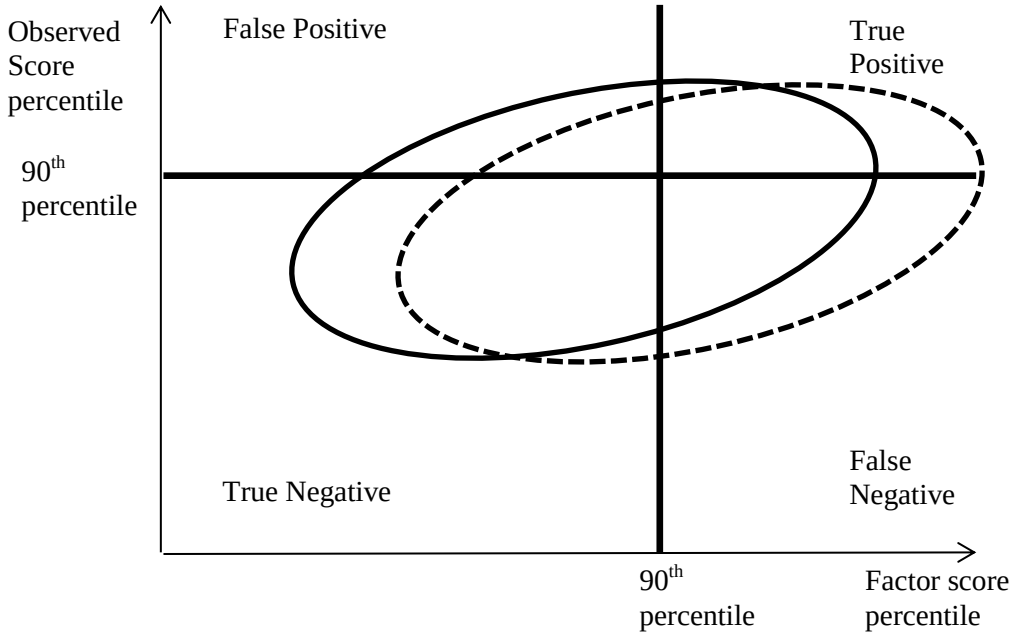

Figure 4.2

Distribution of Factor scores and Observed Scores into quadrants assuming cut-point at 90th percentile for the Partial Invariance (hard line) and Strict Invariance (broken line) conditions.

Chapter Five – Results Hs1

5.1 Defining a Baseline model

Candidate Model definition - Norm A Sample

Using the 32 items from the MMPI-2 Hs1 scale, EFA showed that a 7-Factor model fitted the data [WLSMV χ^2 (293, N = 1275) = 327.252, p = .08]. Therefore seven factors were set as the upper limit for the candidate model in the Norm A sample. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed the 7-Factor model and the 6-Factor model were inadmissible as they had factors with less than three items. All remaining simple structure CFA models were admissible. The 4-Factor model was not nested under the 5-Factor model precluding a χ^2 difference test. Table 5.1 shows the fit indices for the 4-Factor model were equal to or better than the 5-Factor model, which supported selecting the more parsimonious 4-Factor model. Table 1 indicates significant χ^2 differences were produced for the 3-Factor compared to the 4-Factor model and each subsequent lower factor model which supported selecting the 4-Factor model as the candidate model. Therefore, the 4-Factor model was chosen as the candidate model. The RMSEA, TLI and CFI for the 4-Factor model indicated reasonable fit under the model fit criteria.

Table 5.1

Confirmatory factor analysis of potential Hs1 candidate models in the Norm A sample (n=1275)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5-Factor	772.718	454	<.0001	.023	.944	.938	
4-Factor	766.932	458	<.0001	.023	.945	.941	4 v 5 (not nested)
3-Factor	804.795	461	<.0001	.024	.939	.935	3 v 4 (p < .0001)
2-Factor	900.506	463	<.0001	.027	.923	.917	2 v 3 (p < .0001)
1-Factor	1163.736	464	<.0001	.034	.876	.868	1 v 2 (p < .0001)

Respecification Norm A 4-Factor model

To generate modification indices for item residual correlations in Mplus it is necessary to specify at least one item residual correlation parameter in the factor model. Items 111 and 28 loaded onto the same factor in all candidate models, and both have content related to stomach discomfort. Therefore these items were specified as having correlated residuals in the factor model.

Norm A 4-Factor post hoc model 1

This model followed the same structure as Norm A 4-Factor model with the addition of a residual correlation between items 28 and 111. The resulting 4-Factor post hoc model 1 was significantly better fitting compared with the 4-Factor model [WLSMV χ^2 (1, N = 1275) = 5.266 p = .0217].

Compared with the 4-Factor model, the CFI increased by .003, TLI increased by .003 and RMSEA was reduced by .001. The results supported the Norm A 4-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm A 4-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (457, N = 1275) = 762.290 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 45 and 141 may improve model fit. The content of item 45 (comparative physical health) and item 141 (recent physical health) are alike which may result in method effects. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 4-Factor post hoc model 2

This model followed the same structure as Norm A 4-Factor post hoc model 1 with the addition of a residual correlation between items 45 and 141. The resulting 4-Factor post hoc model 2 was significantly better fitting compared with the 4-Factor post hoc model 1 [WLSMV χ^2 (1, N = 1275) = 22.776 p < .0001].

Compared with the 4-Factor post hoc model 1, the CFI increased by .003, TLI increased by .003 and RMSEA was reduced by .001. The results supported the Norm A 4-Factor post hoc model 2 as preferred even though the overall test of fit showed that the Norm A 4-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (456, N = 1275) = 745.264 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 173 and 249 may improve model fit. The content of item 173 (eye fatigue whilst reading) and item 249 (eyesight) is alike which may result in method effects. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 4-Factor post hoc model 3

This model followed the same structure as Norm A 4-Factor post hoc model 2 with the addition of a residual correlation between items 173 and 249. The resulting 4-Factor post hoc model 3 was significantly better fitting compared with the 4-Factor post hoc model 2 [WLSMV χ^2 (1, N = 1275) = 27.104 p < .0001].

Compared with the 4-Factor post hoc model 2, the CFI increased by .004, TLI increased by .004 and RMSEA was unchanged. The results supported the Norm A 4-Factor post hoc model 3 as preferred even though the overall test of fit showed that the Norm A 4-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (455, N = 1275) = 723.645 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 28 and 59 may improve model fit. The content of item 28 (stomach complaints) and item 59 (stomach discomfort) is alike which may result in method effects. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing model fit.

Norm A 4-Factor post hoc model 4

This model followed the same structure as Norm A 4-Factor post hoc model 3 with the addition of a residual correlation between items 28 and 59. The resulting 4-Factor post hoc model 4 was significantly better fitting compared with the 4-Factor post hoc model 3 [WLSMV χ^2 (1, N = 1275) = 7.094 p < .0001].

Compared with the 4-Factor post hoc model 3, the CFI and TLI remained unchanged, and RMSEA was reduced by .001. The results supported the Norm A 4-Factor post hoc model 4 as preferred even though the overall test of fit showed that the Norm A 4-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (454, N = 1275) = 718.774 p < .0001].

Review of the modification indices for item residuals found no additional pairs of item correlations that were both empirically and theoretically justified. However modification indices suggested moving item 10 (work ability) from Factor-1 to Factor-4 may improve model fit.

Norm A 4-Factor post hoc model 5

This model followed the same structure as Norm A 4-Factor post hoc model 4 with item 10 moved from Factor-1 to Factor-4. The 4-Factor post hoc model 5 and 4-Factor post hoc model 4 were not nested under each other, precluding a χ^2 difference test.

Compared with the 4-Factor post hoc model 4, the CFI decreased by .003, TLI decreased by .003, and RMSEA was increased by .001. The results supported rejecting the Norm A 4-Factor post hoc model 5, and maintaining the Norm A 4-Factor post hoc model 4 as preferred.

Review of the modification indices from Norm A 4-Factor post hoc model 4 suggested moving item 47 (heart and chest pains) from Factor-4 to Factor-1 may improve model fit.

Norm A 4-Factor post hoc model 6

This model followed the same structure as Norm A 4-Factor post hoc model 4 with item 47 moved from Factor-4 to Factor-1. The 4-Factor post hoc model 6 and 4-Factor post hoc model 4 were not nested under each other, precluding a χ^2 difference test.

Compared with the 4-Factor post hoc model 4, the CFI increased by .001, TLI increased by .001, and RMSEA was unchanged. The results supported the Norm A 4-Factor post hoc model 6 as preferred even though the overall test of fit showed that the Norm A 4-Factor post hoc model 6 still displayed a significant lack of fit [WLSMV χ^2 (454, N = 1275) = 712.747 $p < .0001$].

Review of the modification indices from Norm A 4-Factor post hoc model 6 suggested moving item 45 (comparative physical health) from Factor-1 to Factor-4 may improve model fit.

Norm A 4-Factor post hoc model 7

This model followed the same structure as Norm A 4-Factor post hoc model 6 with item 45 moved from Factor-1 to Factor-4. The 4-Factor post hoc model 7 and 4-Factor post hoc model 6 were not nested under each other, precluding a χ^2 difference test.

Compared with the 4-Factor post hoc model 6, the CFI decreased by .005, TLI decreased by .006, and RMSEA was increased by .001. The results supported rejecting the Norm A 4-Factor post hoc model 7, and maintaining the Norm A 4-Factor post hoc model 6 as preferred.

Review of the modification indices found no support for further respecifications. Therefore the Norm A 4-Factor post hoc model 6 was selected as the candidate model (see Appendix 1 for details).

Candidate Model definition - Norm B Sample

EFA showed the 8-Factor model fitted the data [WLSMV χ^2 (268, N = 1273) = 299.778, p = .09]. Therefore eight factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed the 8-Factor model, 7-Factor model, and 6-Factor model were inadmissible as all contained a factor with less than three items. All remaining factor models were admissible. Table 5.2 indicates significant χ^2 difference tests were produced for the 4-Factor model compared with the 5-Factor model, and significant χ^2 difference tests were produced for each subsequent lower factor model. Therefore, the 5-Factor model was chosen as the candidate model. The RMSEA, TLI and CFI for the 5-Factor model indicated reasonable fit.

Table 5.2

Confirmatory factor analysis of potential Hs1 candidate models in the Norm B sample (n=1273)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5-Factor	730.396	454	<.0001	.022	.955	.951	
4-Factor	777.249	458	<.0001	.023	.949	.944	4 v 5 (p < .0001)
3-Factor	832.168	461	<.0001	.025	.940	.936	3 v 4 (p < .0001)
2-Factor	947.906	463	<.0001	.029	.922	.916	2 v 3 (p < .0001)
1-Factor	1123.181	464	<.0001	.033	.894	.886	1 v 2 (p < .0001)

Respecification Norm B 5-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model. The specification of item correlations was based on method effects from shared item

content among the pairs of items. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this conclusion.

Norm B 5-Factor post hoc model 1

This model followed the same structure as Norm B 5-Factor model with the addition of a residual correlation between the following item pairs: 28 with 59; 28 with 111; 45 with 141; and, 173 with 249. The resulting 5-Factor post hoc model 1 was significantly better fitting compared with the 5-Factor model [WLSMV χ^2 (4, N = 1273) = 33.928 p < .0001].

Compared with the 5-Factor model, the CFI increased by .005, TLI increased by .005 and RMSEA was reduced by .001. The results supported the Norm B 5-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm B 5-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (450, N = 1273) = 696.254 p < .0001].

Inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between item 28 with both items 111 and 59, respectively. This observation supported removing the residual correlations between these two pairs of items and reviewing the model fit.

Norm B 5-Factor post hoc model 2

This model followed the same structure as Norm B 5-Factor post hoc model 1 with the removal of residual correlations between item pairs 28 with 59, and 28 with 111. There was no significant difference in model fit comparing the 5-Factor post hoc model 2 with the 5-Factor post hoc model 1 [WLSMV χ^2 (2, N = 1273) = 4.669 p = .0969].

Compared with the 5-Factor post hoc model 1 the CFI, TLI and RMSEA were unchanged. Under the preference for model parsimony the Norm B 5-Factor post hoc model 2 was preferred as it included fewer parameters specified, even though the overall test of fit showed that the Norm B 5-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (452, N = 1273) = 699.465 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 18

and 111 may improve model fit. The content of item 18 (vomiting and nausea) and item 111 (stomach troubles) are alike which may result in method effects. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm B 5-Factor post hoc model 3

This model followed the same structure as Norm B 5-Factor post hoc model 2 with the addition of a residual correlation between items 18 and 111 added. Mplus output generated a warning message suggesting an inadmissible model identifying item 111 as the concern. Therefore this model was rejected. The modification indices showed no further respecifications that were theoretically and empirically justified. Therefore the Norm B 5-Factor post hoc model 2 was selected as the candidate model (see Appendix 1 for details).

Candidate Model definition - TBI Sample

EFA showed the 4-Factor model fitted the data [WLSMV χ^2 (374, N = 242) = 390.425, p = .27]. Thus four factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed all models were admissible. Table 5.3 shows that all χ^2 comparison tests produced significant results, concluding no model better represented the data than the 4-Factor model. Therefore the 4-Factor model was the selected candidate model. The RMSEA, CFI and TLI were best in the 4-Factor model offering additional evidence to support the decision.

Table 5.3

Confirmatory factor analysis of potential Hs1 candidate models in the TBI sample (n=242)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
4-Factor	520.848	458	=.020	.024	.973	.970	
3-Factor	550.777	461	=.003	.028	.961	.958	3 v 4 (p < .0001)
2-Factor	594.032	463	<.0001	.034	.943	.939	2 v 3 (p < .0001)
1-Factor	665.077	464	<.0001	.042	.912	.906	1 v 2 (p < .0001)

Respecification TBI 4-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model, above. The specification of item correlations was based on method effects from shared item content among the pairs. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this prediction.

TBI 4-Factor post hoc model 1

This model followed the same structure as TBI 4-Factor model with the addition of a residual correlation between the following item pairs: 28 with 59; 28 with 111; 45 with 141; and, 173 with 249. There was no significant difference found between the resulting 4-Factor post hoc model 1 compared with the 4-Factor model [WLSMV χ^2 (4, N = 242) = 5.345 p = .2537].

Compared with the 4-Factor model, the CFI, TLI and RMSEA were unchanged. The results supported the TBI 4-Factor model as preferred. Additionally inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between all specified pairs of residual correlations which supported removing these specified parameters from the model.

The modification indices suggested specifying a residual correlation between items 3 and 39 may improve model fit. The content of item 3 (sleep and energy) and item 39 (sleep) is alike which may result in method effects. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

TBI 4-Factor post hoc model 2

This model followed the same structure as TBI 4-Factor model (no residual correlations) with the addition of a residual correlation between items 3 and 39. The resulting 4-Factor post hoc model 2 was significantly better fitting compared with the 4-Factor model [WLSMV χ^2 (1, N = 242) = 6.858 p = .0088].

Compared with the 4-Factor model (no residual correlations), the CFI increased by .002, TLI increased by .002 and RMSEA was reduced by .001. The results supported the TBI 4-Factor post hoc model 2 as preferred even though the overall test of fit showed that the TBI 4-Factor post

hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (457, N = 242) = 515.250 p = .0306].

The modification indices for item residuals showed no additional pairs that were both empirically and theoretically justified. However the modification indices suggested moving item 45 (comparative physical health) from Factor-3 to Factor-4 may improve model fit.

TBI 4-Factor post hoc model 3

This model followed the same structure as TBI 4-Factor post hoc model 2 with item 45 moved from Factor-3 to Factor-4. The resulting 4-Factor post hoc model 3 is not nested with 4-Factor post hoc model 2 precluding a χ^2 difference test.

Compared with the TBI 4-Factor post hoc model 2, the CFI increased by .002, TLI increased by .002 and RMSEA was reduced by .001. The results supported the TBI 4-Factor post hoc model 3 as preferred even though the overall test of fit showed that the TBI 4-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (457, N = 242) = 510.547 p = .0420].

The modification indices suggested moving item 53 (bodily perceptions) from Factor-2 to Factor-4 may improve model fit.

TBI 4-Factor post hoc model 4

This model followed the same structure as TBI 4-Factor post hoc model 3 with item 53 moved from Factor-2 to Factor-4. The resulting 4-Factor post hoc model 4 is not nested with 4-Factor post hoc model 3 precluding a χ^2 difference test.

Compared with the TBI 4-Factor post hoc model 3, the CFI increased by .002, TLI increased by .002 and RMSEA was reduced by .001. The results supported the TBI 4-Factor post hoc model 4 as preferred. Furthermore, the non-significant χ^2 result for overall model fit supported the decision to accept the TBI 4-Factor post hoc model 4 as the candidate model [WLSMV χ^2 (457, N = 242) = 505.529 p = .0578]. See Appendix 1 for details of the selected candidate model.

Candidate Model Replication Analysis

Each of the three candidate models were then examined for fit in additional samples. Five samples were included in this phase of analysis. The three samples used to determine the candidate models along with the Female Norm and Male Norm samples.

Table 5.4 indicates that the Norm A and Norm B candidate models, described above, performed similarly across all samples with averages on fit indices of RMSEA .023 for both, CFI .955 and .956 respectively, and TLI .951 and .952 respectively. This was surprising as the Norm A model has four factors while the Norm B model has five factors and it was anticipated the replication process would find a clear empirical preference.

The TBI candidate model produced average fit indices of RMSEA .024, CFI, .946 and TLI .941. In absolute terms these indices suggested a worse fitting model. However with four of the five samples analysed being non-clinical, the performance of the TBI model was impressive. The RMSEA for all candidate models overlapped at the 90 per cent confidence interval on the Female Norm, Male Norm and TBI samples. The RMSEA for the Norm A and Norm B models also overlapped in the Norm A and Norm B samples; whilst in these samples the RMSEA for the TBI model did not overlap, indicating poorer fit than in the other three samples. Thus, for two of the four non-clinical samples there was no significant difference in the RMSEA confidence intervals observed across the three candidate models. Additionally, across the three candidate models in the TBI sample no difference in RMSEA confidence intervals was observed, which is noteworthy for a diagnostic instrument.

Therefore the replication process was unable to clearly define a preferred model. However, the procedure supported the stability of all three candidate models across a variety of samples as seen in Table 5.4.

Item explained variance

Item R-squared values indicate the amount of factor variance explained by each of the items loading on the factor. The value for each candidate model generated from analysis in the TBI sample was reviewed. The TBI sample was chosen because all candidate models displayed the best fit in this sample. Additionally the sample is a clinical sample which is the focus of the use of the MMPI-2. The average item Fisher's Z was 0.720 in the Norm A candidate model; 0.769

in the Norm B candidate model; and, 0.774 in the TBI candidate model. Paired samples t-tests found a difference between Norm A and Norm B, $t(31) = -2.126$, $p < .05$ and between Norm A and TBI, $t(31) = -2.458$, $p < .05$. However no difference was found between Norm B and TBI, $t(31) = -.395$, $p > .05$. These results suggest the Norm B and TBI candidate models are better performed than the Norm A model in terms of item variance explained by the model.

Construct reliability

Kline (2010) and Raykov (2004) refer to the factor rho coefficient which is as a ratio of explained item variance over total factor variance. Calculation of the factor rho coefficient facilitates a comparison of the reliability of the factors, or constructs, within each candidate model. Again the TBI sample was used to calculate the values shown in Table 5.5.

With no standard errors available from the factor rho formula comparisons are arbitrary (Kline, 2010). Additionally it is difficult to compare the individual coefficients due to differences in the item to factor structure across models. One exception was for Norm A Factor-2, TBI Factor-1 and Norm B Factor-3 which cover similar items (see Appendix 1). Comparison of the coefficients in these factors found the TBI candidate model was .046 greater than the coefficient for the Norm B candidate model, which was in turn 0.123 greater than the Norm A candidate model when compared in the TBI sample. Overall the average for the factor coefficient was comparable in the TBI (0.578) and the Norm B (0.575) candidate models, with both observed to be higher than the Norm A (0.498) candidate model.

Review of the factor structure for each candidate model

To complete a review of the factor structure within each candidate model, domains of items were identified. Domains were defined as groups of 3 or more items which were allocated in a consistent pattern across the candidate models (see Table 5.6). There was no requirement that the items within a domain be similar in content. Domains were determined purely based on similarity of the candidate models' structures, defined from the factor analysis procedure describe above. For example Table 5.6 shows item 101 (thrill seeking behaviour) is included in the same domain as items that refer to head and neck pain because. These four items are allocated to the same domain because of the consistent empirical relationship between these items in all three candidate models, despite any difference in content. Table 5.6 shows there were six domains which consisted of 23 items that were observed in all candidate models, with nine

Table 5.4

Confirmatory factor analysis with the Hs1 candidate models in all samples

Candidate	Sample	χ^2	*df*	*p*	RMSEA	CFI	TLI
Norm A 4-Factor	Female Norm	848.880	454	< .0001	.025	.950	.946
				90% CI	.022 - .027		
Norm A 4-Factor	Male Norm	625.166	454	< .0001	.018	.956	.952
				90% CI	.015 - .022		
Norm A 4-Factor	Norm A	712.747	454	< .0001	.021	.954	.950
				90% CI	.018 - .024		
Norm A 4-Factor	Norm B	751.352	454	< .0001	.023	.952	.948
				90% CI	.020 - .026		
Norm A 4-Factor	TBI	540.979	454	= .0031	.028	.962	.958
				90% CI	.017 - .037		
TBI 4-Factor	Female Norm	954.499	457	< .0001	.028	.937	.932
				90% CI	.025 - .030		
TBI 4-Factor	Male Norm	696.796	457	< .0001	.022	.938	.933
				90% CI	.018 - .025		
TBI 4-Factor	Norm A	837.057	457	< .0001	.026	.933	.927
				90% CI	.023 - .028		
TBI 4-Factor	Norm B	819.129	457	< .0001	.025	.942	.937
				90% CI	.022 - .028		
TBI 4-Factor	TBI	505.529	457	= .0578	.021	.979	.977
				90% CI	.000 - .031		
Norm B 5-Factor	Female Norm	846.422	452	< .0001	.025	.950	.946
				90% CI	.022 - .027		
Norm B 5-Factor	Male Norm	613.027	452	< .0001	.018	.958	.954
				90% CI	.014 - .021		
Norm B 5-Factor	Norm A	746.587	452	< .0001	.023	.948	.943
				90% CI	.020 - .025		
Norm B 5-Factor	Norm B	699.465	452	< .0001	.021	.960	.956
				90% CI	.018 - .024		
Norm B 5-Factor	TBI	538.433	452	= .0032	.028	.962	.959
				90% CI	.017 - .037		

Table 5.5

The factor rho coefficients for the Hs1 candidate models

	Factor Rho Coefficient				
Candidate Model	1	2	3	4	5
Norm A	0.395	0.684	0.496	0.416	n/a
TBI	0.853	0.654	0.409	0.395	n/a
Norm B	0.608	0.465	0.807	0.335	0.660

items not allocated to a domain.

Review of the domain to factor structure was unable to find a preferred theoretical relationship when comparing the candidate models. Therefore a review of the nine items not allocated to a domain was undertaken.

Item 2 (appetite) was allocated to same factor as Domain A items in the Norm A candidate model. Domain A item content reflects gastrointestinal symptoms which supported the allocation of this item to the same factor. Item 2 in the Norm B and TBI candidate models was allocated to a factor that included sleep items (3 and 39) which may be considered to influence appetite. However, the Norm A candidate model was slightly preferred.

Item 3 (sleep and energy) in the Norm A candidate model was associated with Domain B items which include neck pain and headache. In the Norm B candidate model the item was not associated with any domain but was associated with sleep, body warmth and appetite items. In the TBI candidate model item 3 was associated with Domain C items (tiredness, recent health, eyesight perceptions and work ability) along with the sleep and appetite items. A cogent argument could be made for the allocation of item 3 found in each candidate model and therefore none was preferred.

Item 8 (body warmth) was associated with Domain B items in the Norm A and Norm B candidate models, and with items from Domains D, E and F in the TBI candidate model. There are two other items with content referring to the body (53 and 175). While these two items are allocated to separate domains only in the TBI candidate model are all three items (8, 53 and 175) found to load onto a single factor. Therefore the TBI candidate model was preferred.

Item 18 (vomiting and nausea) was associated with the Domain A items in the Norm A and TBI candidate models, but not in the Norm B candidate model. Items from Domain A refer to gastrointestinal symptoms which supported the loading of item 18 to the same factor. Therefore the Norm A and TBI candidate models were preferred over the Norm B candidate model.

Item 20 (constipation) was associated with Domain B in the Norm A candidate model, with Domain A in the Norm B candidate model, and with Domains D, E and F in the TBI candidate model. Items from Domain A refer to gastrointestinal symptoms which supported the allocation of item 20 with Domain A on the same factor. This was observed in the Norm B candidate model. In the TBI model, the allocation of item 20 to the same factor as the general health symptoms from Domains, D, E and F appears theoretically reasonable. The association of item 20 with Domain B, which referred to head and neck pain, was a weakness in the Norm A candidate model.

Item 39 (sleep) was associated with item 3 in all samples as would be expected. Thus the same comments apply as those covered when item 3 was reviewed.

Item 149 (head pain) was associated with Domain B in the TBI candidate model and with Domain F in the Norm A and Norm B candidate models. The content of item 149 appeared better related to Domain B items. Therefore the TBI candidate model was preferred.

Item 173 (eye fatigue whilst reading) was related to item 249 (eyesight perceptions) from Domain C in Norm A and TBI candidate model as would be expected. The separation of items 173 and 249 to separate factors was a weakness of the Norm B candidate model. Only in the TBI candidate model the item was associated with sleep symptoms which could be explained by a relationship between fatigue and sleep. Therefore the TBI candidate model was preferred.

Item 224 (general pains) was associated with Domains C, D and E in the Norm A candidate model, with Domains C and D in the Norm B candidate model, and with Domain B in the TBI model. The general nature of this item meant an argument could be made for all candidate models. However the TBI candidate model allocated item 224 to the same factor as other items which refer to pain, items 57 (neck pain) and 149 (head pain). Therefore the TBI candidate model was preferred.

Table 5.6

Item content and domain allocations for Hs1 candidate models

Items	Content	Domain	Factor Norm A	Factor Norm B	Factor TBI
28	Stomach complaints	A	1	1	1
59	Stomach discomfort	A	1	1	1
111	Stomach troubles	A	1	1	1
57	Neck pain	B	2	5	2
97	Head and nasal perceptions	B	2	5	2
101	Thrill seeking behaviour	B	2	5	2
176	Headaches	B	2	5	2
10	Work ability	C	3	3	3
141	Recent health	C	3	3	3
152	Tiredness	C	3	3	3
249	Eyesight perceptions	C	3	3	3
45	Comparative physical health	D	3	3	4
47	Heart and chest pains	D	3	3	4
175	Body weakness	D	3	3	4
143	Weight gain/loss	E	3	4	4
164	Dizziness	E	3	4	4
179	Walking difficulty	E	3	4	4
208	Shortness of breath/heart pounding	E	3	4	4
53	Bodily perceptions	F	4	4	4
91	Muscular perceptions	F	4	4	4
117	Coughing or vomiting blood	F	4	4	4
247	Numbness	F	4	4	4
255	Tinnitus symptoms	F	4	4	4
2	Appetite	**	1	2	3
3	Sleep & energy	**	2	2	3
8	Body warmth	**	2	2	4
18	Vomiting and nausea	**	1	4	1
20	Constipation	**	2	1	4
39	Sleep	**	2	2	3
149	Head pain	**	4	4	2
173	Eye fatigue whilst reading	**	3	4	3
224	General pains	**	3	3	2

** *Items not allocate to a domain*

Summary of the theoretical review

In support of the TBI model the review found no item with a clearly identifiable factor misallocation and items 8, 18, 149, 173 and 224 are loaded onto factors with strong conceptual association. Both the Norm A candidate model (8 and 20) and the Norm B candidate model (8, 18 and 173) allocated items to factor structures that appeared to have conceptual weaknesses.

Selection of the 4-Factor TBI candidate model as the baseline model

In conclusion the TBI candidate model appears the preferred model after reviewing the item to factor structure assignments. The Norm B and TBI candidate models were preferred over the Norm A candidate model when reviewing the construct validity measures (factor rho coefficient). Otherwise the candidate model statistical analyses were unable to find clear support for one model, with all the models generating acceptable and comparable results. Therefore basing the final decision on theoretical grounds is warranted. Table 5.7 shows the selected baseline model comprises four factors of 'Gastrointestinal complaints', 'Orofacial symptomology', 'Sleep quality/ energy levels' and 'General health'.

Table 5.7

The item to factor structure of the Hs1 4-Factor baseline model

Factor Name	Items included in the factor
1. 'Gastrointestinal complaints'	18, 28, 59 and 111.
2. 'Orofacial symptomology'	57, 97, 101, 149, 176 and 224
3. 'Sleep quality/ energy levels'	2, 3, 10, 39, 141, 152, 173 and 249
4. 'General health'	8, 20, 45, 47, 53, 91, 117, 143, 164, 175, 179, 208, 247 and 255

5.2 Measurement Invariance Testing

Measurement invariance testing followed a hybrid of the approaches recommended Millsap and Yun Tein (2004) and, Muthen and Muthen (1998-2010) when analysing dichotomous data. In Test 1 the residuals are constrained to equality across groups with loadings and thresholds allowed to vary (except for loadings on marker variables). In Test 2 all residuals, loadings and threshold are constrained to equality across groups, a test of strict invariance. A decrease by .002

or less in the CFI value was selected as the criteria to meet invariance as proposed by Meade et al. (2008). If the model fails to meet the requirement of strict invariance, then item parameters (loadings and thresholds) will be freed in unison, one item at a time, applying the backwards elimination procedure to find a model with the maximum constraint that meets the requirements of partial invariance (Cheung & Rensvold, 1999). The two samples included for the invariance testing were the TBI sample (n=242) and a Community sample (the gender matched MMPI-2 normative sample, n=1786, male = 1116 and female = 670).

Invariance tests detailed in Table 5.8 shows Test 1 produced acceptable fit indices with the residuals constrained to equivalence. Test 2 generated a decrease in the CFI value of .015, thus failing the test of strict invariance. Parameters of the factor model were freed sequentially and allowed to vary across groups based on the modification indices. This procedure showed that sequentially freeing loadings and thresholds for items 10, 47, 3 and 39 was sufficient to meet the requirements for partial invariance (Test 3 in Table 5.8). That is the RMSEA was unchanged and the CFI decreased by .001 when compared with the fit indices generated in Test 1. Originally item 3 was the marker variable on Factor-3, this was changed to item 173 during this procedure to free item loadings. See Table 5.12 for details of the standardized factor loadings for the TBI and Community samples in the partial invariance model.

Table 5.8

Measurement invariance testing of the Hs1 4-Factor model across the Community (n=1786) and TBI (n=242) samples.

Invariance Model	*WLSMV*χ^2	df	p	χ^2 diff	RMSEA	CFI	TLI
Test 1 residuals equal	1402.221	914	< .0001		.023	.947	.943
Test 2 residuals, loadings and thresholds equal	1600.758	970	< .0001	p <.0001[a]	.025	.932	.930
Test 3 Partial invariance	1457.460	962	< .0001	p =.0002[a]	.023	.946	.945

[a]Note this is the χ^2 difference test compared with the Test 1 model

5.3 Practical impact analysis

The finding of partial invariance makes further investigation necessary to determine whether there is any practical impact from finding non-invariant items (Millsap & Kwok, 2004). To investigate the practical impact of the partially invariant model described above, an evaluation of sensitivity and specificity values under conditions for Test 2 and Test 3 was completed. Test 2 represents the hypothetical case of strict invariance and Test 3 represents the observation of partial invariance. As previously noted in Chapter Four, the procedure described by Millsap and Kwok helps ascertain what impact on diagnostic accuracy with the MMPI-2 Hs1 scale is found from the failure to find strict invariance. Therefore a comparison of the sensitivity and specificity values in the partial invariance (Test 3) and strict invariance (Test 2) condition is required. The item with loadings and thresholds freed to facilitate partial invariance are associated with Factor-3 and Factor-4. Therefore the practical impact analysis of the failure to find strict invariance is necessary for these factors but not for Factor-1 or Factor-2.

In Table 5.9 the observed score and factor score cut-points for each invariance condition are shown. The observed scores remain unchanged across invariance conditions as only the psychometric properties of the measurement model vary with consequences for the factor scores. Information about the procedure followed to calculate these values is provided in Chapter Four (see Section 4.5). To briefly reiterate an observed composite score that was as close to the 93.32 percentile rank in the Community sample was used at the cut-point to represent a diagnosis of psychopathology. The actual percentile rank of the observed composite score from the Community sample was used as the cut-point for the factor score to represent the presence or absence of psychopathology.

Summary of Practical Impact Analysis

Table 5.10 shows the sensitivity and specificity for the TBI sample on the factors of interest. All values generated in Table 5.10 were calculated using the VassarStats calculator (Lowry, 1998-2012). Millsap and Kwok (2004) recommend reviewing the changes in sensitivity and specificity values between the two invariance conditions. This review is necessary for both Factor-3 and Factor-4.

A comparison of the sensitivity and specificity values for the TBI sample found no instances

where the value in the partial invariance condition was below the lower bound of the 95% confidence interval (CI) value in the strict invariance condition. The practical impact analysis supports retaining all items from Factor-3 and Factor-4 when assessing persons with a TBI, because there is no significant loss of sensitivity or specificity. In other words there was no significant impact on the sensitivity or specificity values from the failure to find strict invariance. Therefore retaining the invariant items and assuming full invariance does not compromise criterion related validity. Further support for this conclusion was the significant and large correlation between factor scores in the partial invariance condition, and the factor scores in the strict invariance condition observed on both Factor-3 (Spearman's rho = .988, $p < .01$) and Factor-4 (Spearman's rho = .999, $p < .01$).

The invariance of items on Factor-1 and Factor-2 combined with the finding of no practical impact from the non-invariance of items on Factor-3 and Factor-4 supports retaining all MMPI-2 Hs1 items when assessing patients suffering a TBI.

5.4 Comparison of threshold parameters

The threshold values can be compared for the items that failed the test of strict invariance. The threshold can be related to the proportion of positive responses, commonly referred to as the difficulty of item in classical test theory (Glöckner-Rist & Hoijtink, 2003). As previously explained, items were rescored to account for polarity which allows the threshold comparison to identify whether the probability of a response changing from not endorsing to endorsing psychopathology differs between the groups. Items with higher threshold are endorsed less often than items with a lower threshold (Steinberg & Thissen, 1995). Therefore a lower threshold in the TBI group reflects an increased likelihood that an item with be endorsed for psychopathology. While a higher threshold in the TBI group reflects a decreased likelihood for item endorsement compared with the normative sample.

Borsboom (2006) explains that potentially, failure to find strict invariance may have no impact when scale scores are calculated. He clarifies that when the sources of item variance between the samples cancel each other out the result is no impact from the failure of strict invariance when. Whilst it is clearly preferable that no items have a bias, a comparison of the threshold parameters can help determine whether the number of lower and higher thresholds for the TBI group cancels

each other out. Alternatively there may be a bias in the direction of higher or lower threshold parameters in the TBI group. A bias towards higher threshold scores would suggest bias towards lower scale scores, and vice versa.

Table 5.11 shows the threshold values for the TBI and the Community samples, for the items that failed the test of strict invariance. Comparisons between the two threshold scores were completed by using the geometric mean of the standard error for each sample (Howell, 2002). Table 5.11 shows items 10 and 39 are equivalent in their threshold values across groups. Item 3 is more likely to be endorsed by the TBI sample, while item 47 is less likely to be endorsed by the TBI sample. This situation appears to reflect the scenario identified by Borbsoom (2006) where the sources of item variance between the samples may cancel each other out. The result is no effect from the failure of invariance when factor scores are calculated. This conclusion is further supported by finding of no practical impact for the items that failed the test of strict invariance.

5.5 Results of Gass (1991) correction procedure analysis

There are five items from the Gass (1991) correction procedure that are part of Hs1. These items are 101, 149, 175, 179 and 247. The model definition procedure found the Gass items were allocated across two factors, with items 101 and 149 forming part of the 'Orofacial symptomology' factor, while items 175, 179 and 247 load onto the 'General health' factor. The allocation of the five Gass items to multiple factors is noteworthy because the model selected was defined initially using the TBI sample. If these items reflected specific variance due to the neurologic content of items it might have been expected that they would load onto a single factor. All five items from the Gass (1991) correction procedure met the criteria of strict invariance. This finding supports the continued use of the items with persons suffering a TBI.

Table 5.9

Observed score, factor score and percentile rank cut points used to calculated sensitivity and specificity values for invariance conditions in each Hs1 factor.

		Partial Invariance		Strict Invariance	
	Percentile rank	Observed Score	Factor Score	Observed Score	Factor Score
Factor-3	87.2	3	0.382	3	0.467
Factor-4	91.3	5	1.404	5	1.468

Table 5.10

Hs1 sensitivity and specificity analysis for Factor-3 and Factor-4 across invariance conditions in the TBI sample

	Factor 3		Factor 4	
	Sensitivity 95% CI	Specificity 95% CI	Sensitivity 95% CI	Specificity 95% CI
Partial Invariance	.877 .822 - .917	.949 .814 - .991	.789 .718 - .847	.987 .919 - .999
Strict invariance	.833 .775 - .879	1.000 .840 – 1.000	.799 .728 - .856	.987 .919 - .999

Table 5.11

Mean threshold value and standard errors (SE) for the Community and the TBI samples for items freed when defining the Hs1 partial invariance model

	Community Thresholds		TBI Thresholds			TBI compared with Combined
Item	Mean	SE	Mean	SE	Geometric Mean of SE	Norm
3	0.472	0.034	1.468	0.517	0.366	>
10	1.289	0.056	0.980	0.431	0.307	=
39	1.382	0.062	1.784	0.510	0.363	=
47	1.166	0.054	0.828	0.171	0.127	<

Table 5.12

Standardized factor loadings for the four factor CFA model of MMPI-2 Hs1 in traumatic brain injury (TBI, n=242) and Community (Norm, n=1786) samples.

Parameter estimates (PE) and standard errors (SE) are derived from the partially invariant measurement model (Invariance Test 3 in Table 5.8). Also shown is the variance explained by the model in each item (R^2)

	Factor 1		Factor 2		Factor 3		Factor 4		R^2	
Item	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM
59	.838(.036)	.862(.026)							.298(.061)	.257(.044)
18	.668(.065)	.705(.055)							.554(.087)	.503(.078)
28	.870(.031)	.891(.023)							.243(.055)	.207(.041)
111	.891(.033)	.908(.024)							.206(.058)	.175(.044)
101			.763(.038)	.687(.040)					.417(.058)	.527(.055)
57			.677(.040)	.593(.027)					.542(.054)	.648(.032)
97			.552(.047)	.469(.038)					.695(.052)	.780(.035)
149			.582(.047)	.498(.037)					.661(.055)	.752(.037)
176			.665(.041)	.580(.030)					.558(.054)	.663(.035)
224			.872(.028)	.819(.026)					.240(.049)	.329(.043)
2					.471(.059)	.560(.053)			.778(.056)	.687(.059)
3					.757(.072)	.374(.036)			.427(.109)	.860(.027)
10					.709(.080)	.542(.038)			.497(.114)	.706(.041)
39					.734(.068)	.554(.041)			.461(.099)	.694(.045)
141					.629(.054)	.716(.035)			.604(.068)	.488(.051)

	Factor 1		Factor 2		Factor 3		Factor 4		R^2	
Item	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM
152					.527(.055)	.617(.027)			.722(.058)	.619(.033)
173					.375(.046)	.456(.031)			.859(.035)	.792(.028)
249					.238(.038)	.297(.032)			.943(.018)	.912(.019)
175							.818(.031)	.751(.035)	.331(.050)	.436(.053)
8							.314(.043)	.256(.032)	.901(.027)	.934(.017)
20							.418(.043)	.345(.032)	.826(.035)	.881(.022)
45							.755(.032)	.677(.029)	.430(.048)	.541(.039)
47							.356(.082)	.589(.033)	.873(.058)	.653(.039)
53							.672(.033)	.588(.026)	.549(.045)	.655(.031)
91							.643(.037)	.557(.031)	.587(.048)	.689(.035)
117							.402(.044)	.331(.034)	.839(.035)	.890(.023)
143							.347(.039)	.284(.031)	.879(.027)	.919(.017)
164							.647(.037)	.562(.031)	.582(.047)	.685(.035)
179							.685(.038)	.601(.036)	.531(.052)	.638(.043)
208							.677(.033)	.592(.024)	.542(.045)	.649(.029)
247							.581(.042)	.496(.035)	.663(.048)	.754(.034)
255							.384(.041)	.316(.032)	.852(.032)	.900(.020)

Chapter Six – Results Hy3

6.1 Defining a Baseline model

Candidate Model definition - Norm A Sample

After removing the 20 items previously analysed with Hs1, there are 40 items from Hy3 remaining for analysis. Using these 40 items from the Hy3 scale, EFA showed that a 13-Factor model fitted the data [WLSMV χ^2 (338, N = 1249) = 368.526, p = .1218]. Therefore 13 factors were set as the upper limit for the candidate model in the Norm A sample. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed all models with greater than five factors were inadmissible due to one or more factors failing the requirement to have a minimum of three items. All remaining simple structure CFA models met the admissibility criteria.

Table 6.1 shows significant χ^2 difference tests were produced for the 4-Factor model and each subsequent lower factor model supporting the selection of the 5-Factor model as the candidate model. Noteworthy is that all admissible models produced non-significant loadings for items 14 and 230. The RMSEA for the 5-Factor model suggested a model that acceptably represents the data. While the CFI and TLI were on the borderline of acceptability. However, after completing the respecification procedure (see below) the CFI and TLI improved sufficiently, by .026 and .027 respectively, to meet the criteria for adequate model fit.

Table 6.1

Confirmatory factor analysis of potential Hy3 candidate models in the Norm A sample (n=1249)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5-Factor	1476.387	730	<.0001	.029	.898	.891	
4-Factor	1517.410	734	<.0001	.029	.893	.887	4 v 5 (p < .0001)
3-Factor	1556.794	737	<.0001	.030	.888	.882	3 v 4 (p < .0001)
2-Factor	1896.422	739	<.0001	.035	.842	.834	2 v 3 (p < .0001)
1-Factor	2873.970	740	<.0001	.048	.709	.694	1 v 2 (p < .0001)

Respecification Norm A 5-Factor model

To generate modification indices for item residual correlations in Mplus it is necessary to specify at least one item residual correlation parameter in the factor model. Items 161 and 185 loaded onto the same factor in all candidate models and both have content related to shyness. Therefore these items were specified as having correlated residuals in the factor model.

Norm A 5-Factor post hoc model 1

This model followed the same structure as Norm A 5-Factor model with the addition of a residual correlation between items 161 and 185. The resulting 5-Factor post hoc model 1 was significantly better fitting compared with the 5-Factor model [WLSMV χ^2 (1, N = 1249) = 52.359 p < .0001].

Compared with the 5-Factor model, the CFI increased by .008, TLI increased by .008 and RMSEA was reduced by .001. The results supported the Norm A 5-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (729, N = 1249) = 1419.048 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 81 and 110 may improve the model performance. Review of the item content found items referring people lying for advantage and people behaving unfairly for advantage respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 5-Factor post hoc model 2

This model followed the same structure as Norm A 5-Factor post hoc model 1 with the addition of a residual correlation between items 81 and 110. The resulting 5-Factor post hoc model 2 was significantly better fitting compared with the 5-Factor post hoc model 1 [WLSMV χ^2 (1, N = 1249) = 73.625 p < .0001].

Compared with the 5-Factor post hoc model 1, the CFI increased by .010, TLI increased by

.011 and RMSEA was reduced by .002. The results supported the Norm A 5-Factor post hoc model 2 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (728, N = 1249) = 1346.797 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 44 and 159 may improve the model performance. Review of the item content found items referring to hot flushes and fainting respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 5-Factor post hoc model 3

This model followed the same structure as Norm A 5-Factor post hoc model 2 with the addition of a residual correlation between items 44 and 159. The resulting 5-Factor post hoc model 3 was significantly better fitting compared with the 5-Factor post hoc model 2 [WLSMV χ^2 (1, N = 1249) = 29.232 p < .0001].

Compared with the 5-Factor post hoc model 2, the CFI increased by .003, TLI increased by .003 and RMSEA was unchanged. The results supported the Norm A 5-Factor post hoc model 3 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (727, N = 1249) = 1324.687 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 7 and 14 may improve the model performance. Review of the item content found items referring to interest in crime and interest in detective/mystery stories respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 5-Factor post hoc model 4

This model followed the same structure as Norm A 5-Factor post hoc model 3 with the

addition of a residual correlation between items 7 and 14. The resulting 5-Factor post hoc model 4 was significantly better fitting compared with the 5-Factor post hoc model 3 [WLSMV χ^2 (1, N = 1249) = 25.026 p < .0001].

Compared with the 5-Factor post hoc model 3, the CFI increased by .002, TLI increased by .003 and RMSEA was reduced by .001. The results supported the Norm A 5-Factor post hoc model 4 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (726, N = 1249) = 1305.113 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 29 and 213 may improve the model performance. Review of the item content found items referring to wanting to swear and quickness to anger respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 5-Factor post hoc model 5

This model followed the same structure as Norm A 5-Factor post hoc model 4 with the addition of a residual correlation between items 29 and 213. The resulting 5-Factor post hoc model 5 was significantly better fitting compared with the 5-Factor post hoc model 4 [WLSMV χ^2 (1, N = 1249) = 16.111 p = .0001].

Compared with the 5-Factor post hoc model 4, the CFI increased by .001, TLI increased by .002 and RMSEA was unchanged. The results supported the Norm A 5-Factor post hoc model 5 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 5 still displayed a significant lack of fit [WLSMV χ^2 (725, N = 1249) = 1294.121 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 167 and 265 may improve the model performance. Review of the item content found items referring to talking to new people and speaking to others respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the

modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 5-Factor post hoc model 6

This model followed the same structure as Norm A 5-Factor post hoc model 5 with the addition of a residual correlation between items 167 and 265. The resulting 5-Factor post hoc model 6 was significantly better fitting compared with the 5-Factor post hoc model 5 [WLSMV χ^2 (1, N = 1249) = 12.687 p = .0004].

Compared with the 5-Factor post hoc model 5, the CFI increased by .002, TLI increased by .001 and RMSEA was unchanged. The results supported the Norm A 5-Factor post hoc model 6 as preferred even though the overall test of fit showed that the Norm A 5-Factor post hoc model 6 still displayed a significant lack of fit [WLSMV χ^2 (724, N = 1249) = 1284.172 p < .0001].

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported and being specified within the model for method effects. However, the modification indices did suggest improvement if item 157 was moved to Factor-2. This item is currently the indicator variable for Factor-5 and will be replaced with item 29 (the next highest EFA loading item).

Norm A 5-Factor post hoc model 7

This model followed the same structure as Norm A 5-Factor post hoc model 6 with item 157 moved from Factor-5 to Factor-2. The resulting 5-Factor post hoc model 7 is not nested with 5-Factor post hoc model 6 precluding a χ^2 difference test.

Compared with the Norm A 5-Factor post hoc model 6, the CFI decreased by .005, TLI decreased by .005 and RMSEA was increased by .001. The results supported rejecting the Norm A 5-Factor post hoc model 7, and maintaining the Norm A 4-Factor post hoc model 6 as preferred.

Reviewing the modification indices found no further respecifications. Therefore the Norm A 5-Factor post hoc model 6 is the preferred candidate model (see Appendix 2 for details).

Candidate Model definition - Norm B Sample

EFA showed the 9-Factor model fitted the data [WLSMV χ^2 (456, N = 1248) = 492.927, p = .1125]. Therefore nine factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed all models with greater than five factors were inadmissible due to one or more factors failing the requirement to have a minimum of 3 items. All remaining factor models were admissible. Table 6.2 indicates significant χ^2 difference tests were produced for the 3-Factor model and each subsequent lower factor model. As the 4-Factor model was not nested under the 5-Factor model a χ^2 difference test was unable to be completed. However, Table 6.2 shows the fit statistics for the 4-Factor model were equal to or better than the 5-Factor model supporting selection of the more parsimonious 4-Factor model. Therefore the 4-Factor model was chosen as the candidate model. The RMSEA, TLI and CFI for the 4-Factor model indicate reasonable fit under the model fit criteria. Noteworthy is that all admissible models produced non-significant loading for item 14.

Table 6.2

Confirmatory factor analysis of potential Hy3 candidate models in the Norm B sample (n=1248)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5 Factor	1580.673	730	<.0001	.031	.905	.898	
4 Factor	1547.675	734	<.0001	.030	.909	.903	4 v 5 (not nested)
3 Factor	1610.678	737	<.0001	.031	.902	.896	3 v 4 (p < .0001)
2 Factor	1965.244	739	<.0001	.036	.862	.855	2 v 3 (p < .0001)
1 Factor	2923.969	740	<.0001	.049	.755	.742	1 v 2 (p < .0001)

Respecification Norm B 4-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model. The specification of item correlations was based on method effects from shared item content among the pairs of items. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this conclusion.

Norm B 4-Factor post hoc model 1

This model followed the same structure as Norm B 4-Factor model with the addition of a residual correlation between the following item pairs: 161 and 185; 81 and 110; 44 and 159; 7 and 14; 29 and 213; 167 and 265. The resulting 4-Factor post hoc model 1 was significantly better fitting compared with the 4-Factor model [WLSMV χ^2 (6, N = 1248) = 204.444 p < .0001].

Compared with the 4-Factor model, the CFI increased by .021, TLI increased by .022 and RMSEA was reduced by .004. The results supported the Norm B 4-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (728, N = 1248) = 1349.901 p < .0001].

Inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between item pair 44 and 159. This observation supported removing the residual correlations between this item pair and reviewing the model fit.

Norm B 4-Factor post hoc model 2

This model followed the same structure as Norm B 4-Factor post hoc model 1 with the removal of a residual correlation between items 44 and 159. There was no significant difference in model fit between the resulting 4-Factor post hoc model 2 compared with the 4-Factor post hoc model 1 [WLSMV χ^2 (1, N = 1248) = 1.019 p = .3127]. Under the preference for model parsimony Norm B 4-Factor post hoc model 2 is preferred as it has less specified parameters, even though the overall test of fit showed that the Norm B 4-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (729, N = 1248) = 1349.795 p < .0001].

Review of the modification indices suggested specifying a residual correlation between item 58 and 81 may improve the model performance. Review of the item content found items referring to people manipulating for sympathy and people lying to get ahead, respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation

between these items and reviewing the factor model fit.

Norm B 4-Factor post hoc model 3

This model followed the same structure as Norm B 4-Factor post hoc model 2 with the addition of a residual correlation between items 58 and 81. The resulting 4-Factor post hoc model 3 was significantly better fitting compared with the 4-Factor post hoc model 2 [WLSMV χ^2 (1, N = 1248) = 26.783 p < .0001].

Compared with the 4-Factor post hoc model 2, the CFI increased by .002, TLI increased by .002 and RMSEA was unchanged. The results supported the Norm B 4-Factor post hoc model 3 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ2 (728, N = 1248) = 1331.996 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 58 and 110 may improve the model performance. Review of the item content found items referring to people manipulating for sympathy and people behaving unfairly for self-gain respectively which supports a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm B 4-Factor post hoc model 4

This model followed the same structure as Norm B 4-Factor post hoc model 3 with the addition of a residual correlation between items 58 and 110. The resulting 4-Factor post hoc model 4 was significantly better fitting compared with the 4-Factor post hoc model 3 [WLSMV χ^2 (1, N = 1248) = 26.301 p < .0001].

Compared with the 4-Factor post hoc model 3, the CFI increased by .002, TLI increased by .003 and RMSEA was reduced by .001. The results supported the Norm B 4-Factor post hoc model 4 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (727, N = 1248) = 1312.382 p < .0001].

Reviewing the modification indices found no further respecifications. As recommended by Kline (2010), a review of the residual correlation matrix suggested high correlations between item 243 and the following items; 161, 167, and 185. All four items refer to shyness and speaking with other people, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm B 4-Factor post hoc model 5

This model followed the same structure as Norm B 4-Factor post hoc model 4 with the addition of a residual correlation between items 161 and 243. There was no significant difference in model fit between the resulting 4-Factor post hoc model 5 compared with the 4-Factor post hoc model 4 [WLSMV χ^2 (1, N = 1248) = 2.262 p = .1326]. Under the preference for model parsimony Norm B 4-Factor post hoc model 4 is preferred as it has less specified parameters.

In the next model residual correlation between items 243 and 167 will be specified (as suggested in review of 4-Factor post hoc model 4).

Norm B 4-Factor post hoc model 6

This model followed the same structure as Norm B 4-Factor post hoc model 4 with the addition of a residual correlation between items 167 and 243. The resulting 4-Factor post hoc model 6 was significantly better fitting compared with the 4-Factor post hoc model 4 [WLSMV χ^2 (1, N = 1248) = 4.334 p < .0374].

Compared with the 4-Factor post hoc model 4, the CFI, TLI and RMSEA were unchanged. The results supported the Norm B 4-Factor post hoc model 6 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 6 still displayed a significant lack of fit [WLSMV χ^2 (726, N = 1248) = 1309.898 p < .0001].

In the next model residual correlation between items 243 and 185 will be specified (as suggested in review of 4-Factor post hoc model 4).

Norm B 4-Factor post hoc model 7

This model followed the same structure as Norm B 4-Factor post hoc model 6 with the addition of a residual correlation between items 185 and 243. The resulting 4-Factor post hoc model 7 was significantly better fitting compared with the 4-Factor post hoc model 6 [WLSMV $\chi 2$ (1, N = 1248) = 7.160 p = .0075].

Compared with the 4-Factor post hoc model 6, the CFI increased by .001, TLI and RMSEA were unchanged. The results supported the Norm B 4-Factor post hoc model 7 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 7 still displayed a significant lack of fit [WLSMV χ^2 (725, N = 1248) = 1306.097 p < .0001].

There were no further modifications suggested by the output. Therefore Norm B 4-Factor post hoc model 7 is the preferred model (see Appendix 2 for details).

Candidate Model definition - TBI Sample

EFA showed the 6-Factor model fitted the data [WLSMV χ^2 (555, N = 233) = 598.439, p = .0983]. Thus six factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed the 6-Factor model was inadmissible due to having a factor with less than three items. All remaining factor models were admissible. Table 6.3 shows all χ^2 comparison tests produced significant results, concluding no model better represented the data than the 5-Factor model. Therefore the 5-Factor model was the selected candidate model. The RMSEA, CFI and TLI were best in the 5-Factor model offering additional evidence to support this selection. The RMSEA, TLI and CFI for the 5-Factor model indicate reasonable fit under the model fit criteria. Noteworthy is that all admissible models produced non-significant loadings for items 7, 14, 115, 230, 253 and 263.

Respecification TBI 5-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model, above. The specification of item correlations was based on method effects from shared item content among the pairs. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this prediction.

Table 6.3

Confirmatory factor analysis of potential Hy3 candidate models in the TBI sample (n=233)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5 Factor	857.220	730	=.0008	.027	.944	.940	
4 Factor	884.762	734	=.0001	.030	.933	.929	4 v 5 ($p < .0001$)
3 Factor	915.264	737	<.0001	.032	.921	.916	3 v 4 ($p < .0001$)
2 Factor	942.520	739	<.0001	.034	.910	.905	2 v 3 ($p < .0001$)
1 Factor	1090.670	740	<.0001	.045	.845	.836	1 v 2 ($p < .0001$)

TBI 5-Factor post hoc model 1

This model followed the same structure as TBI 5-Factor model with the addition of a residual correlation between the following item pairs: 161 and 185; 81 and 110; 44 and 159; 7 and 14; 29 and 213; 167 and 265. The resulting 5-Factor post hoc model 1 was significantly better fitting compared with the 5-Factor model [WLSMV χ^2 (6, N = 233) = 61.532 $p < .0001$].

Compared with the 5-Factor model, the CFI increased by .014, TLI increased by .015 and RMSEA was reduced by .003. The results supported the TBI 5-Factor post hoc model 1 as preferred even though the overall test of fit showed that the TBI 5-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (724, N = 233) = 819.269 p = .0078].

Inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between item pairs: 44 and 159; 29 and 213; as well as, 167 and 265. This observation supported removing the residual correlations between these three pairs of items and reviewing the model fit.

TBI 5-Factor post hoc model 2

This model followed the same structure as TBI 5-Factor post hoc model 1 with the removal of residual correlations between item pairs 44 and 159; 29 and 213; as well as, 167 and 265. There was no significant difference in model fit comparing the 5-Factor post hoc model 2 with the 5-Factor post hoc model 1 [WLSMV χ^2 (3, N = 233) = 7.117 p = .0683]. Under the

preference for model parsimony the TBI 5-Factor post hoc model 2 was preferred as it included fewer parameters specified

Compared with the 5-Factor post hoc model 1 the CFI and TLI decreased by .001, and RMSEA remained unchanged. Despite the minor decrease in CFI and TLI, the non-significant χ^2 difference test between models supports selecting the TBI 5-Factor post hoc model 2 as preferred, even though the overall test of fit showed that the TBI 5-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (727, N = 233) = 823.825 p = .0071].

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported and being specified within the model for method effects. However, the modification indices did suggest improvement if item 218 was moved to Factor-4. This item will be respecified accordingly and the model performance reviewed.

TBI 5-Factor post hoc model 3

This model followed the same structure as TBI 5-Factor post hoc model 2 with item 218 moved from Factor-3 to Factor-4. The resulting 5-Factor post hoc model 3 is not nested with 5-Factor post hoc model 2 precluding a χ^2 difference test.

Compared with the TBI 5-Factor post hoc model 2, the CFI increased by .002, TLI increased by .002 and RMSEA was reduced by .001. The results supported the TBI 5-Factor post hoc model 3 as preferred even though the overall test of fit showed that the TBI 5-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (727, N = 233) = 819.811 p = .0093].

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported and being specified within the model for method effects. However, the modification indices did suggest improvement if item 26 was moved to Factor-2. This item will be respecified accordingly and the model performance reviewed.

TBI 5-Factor post hoc model 4

This model followed the same structure as TBI 5-Factor post hoc model 3 with item 26

moved from Factor-5 to Factor-2. The resulting 5-Factor post hoc model 4 is not nested with 5-Factor post hoc model 3 precluding a χ^2 difference test.

Compared with the TBI 5-Factor post hoc model 3, the CFI increased by .003, TLI increased by .003 and RMSEA was reduced by .001. The results supported the TBI 5-Factor post hoc model 4 as preferred even though the overall test of fit showed that the TBI 5-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (727, N = 233) = 812.391 p = .0148].

Reviewing the modification indices found no further respecifications. Therefore the TBI 5-Factor post hoc model 4 is the preferred candidate model (see Appendix 2).

Candidate Model Replication Analysis

Next each of the three candidate models were examined for fit in other samples. Five samples were included in this phase of analysis. The three samples used to determine the candidate models along with the Female Norm and Male Norm samples.

Simple structure CFA showed the Norm B candidate model and the TBI candidate model were both admissible on all the replication samples. Simple structure CFA showed the Norm A candidate model to be inadmissible in the Norm B sample, item 65 had loading greater than 1. Additionally the Norm A candidate model was inadmissible in the TBI sample as Factor-5 failed the factor covariation admissibility criteria with excessive correlations between Factor-2, Factor-3 and Factor-4. Therefore Norm A candidate model was excluded from further comparisons and not considered for selection as the baseline model.

Table 6.4 indicates that the Norm B and TBI candidate models performed similarly across all samples with averages on fit indices of RMSEA .027 for both, CFI .925, and TLI .919. After the removal of each sample used to generate the two remaining candidate models (i.e. the Norm B and TBI samples), the Norm B candidate model produced slightly better fit indices than the TBI candidate model. In this comparison respective Norm B and TBI candidate model averages on CFI were .923 against .915; on TLI .917 against .909; and on RMSEA .026 against .027.

The RMSEA for both candidate models overlapped at the 90% CI on the Female Norm, Male Norm and Norm A samples. Compared with the TBI model the RMSEA for the Norm B models was greater in the Norm B sample with the RMSEA for the TBI sample being greater in the TBI model. This observation reflects the bias for a candidate models with EFA and CFA run in the same sample. The CFI and TLI suggest a slight preference for the Norm B model, while the RMSEA find no preference. Importantly the replication procedure has supported the stability of all the remaining two candidate models across a variety of samples.

Item explained variance

Item R-squared values indicate the amount of factor variance explained by each of the items loading onto the factor. The values for the Norm B and TBI candidate models when analysed in the Female Norm sample were reviewed. This sample is chosen because there is no shared single sample in which each candidate model produced their best fit indices. After removing each sample used to generate the remaining candidate models (i.e. the Norm B and TBI samples), both models were better fitting in the Female Norm sample than the Male Norm and Norm A samples.

Inspection found no clear difference in the performance of single items in both correlational and standardized values. The average item Fisher's z was 0.581 in the Norm B candidate model and, 0.595 in the TBI candidate model. Paired samples t-tests found no difference between Norm B and TBI, $t(39) = -1.18$, $p > .05$.

Construct reliability

Kline (2010) and Raykov (2004) refer to the factor rho coefficient, which can be calculated from the Mplus output, as a ratio of explained variance over total variance. This allows us to compare the reliability of the constructs, or factors within each candidate model. Again the Female Norm sample was used to calculate the values in Table 6.5 due to the reasons described above.

With no standard errors calculated during application of the factor rho formula, comparisons are arbitrary. Additionally it is difficult to compare the individual coefficients

Table 6.4

Confirmatory factor analysis with the Hy3 candidate models in all samples

Candidate	Sample	χ^2	df	p	RMSEA	CFI	TLI
Norm B 4-Factor	Female Norm	1376.995	725	< .0001	.025	.935	.930
				90% CI	.023 - .027		
Norm B 4-Factor	Male Norm	1220.276	725	< .0001	.025	.921	.915
				90% CI	.022 - .027		
Norm B 4-Factor	Norm A	1361.310	725	< .0001	.027	.913	.907
				90% CI	.024 - .029		
Norm B 4-Factor	Norm B	1306.097	725	< .0001	.025	.935	.930
				90% CI	.023 - .028		
Norm B 4-Factor	TBI	906.565	725	< .0001	.033	.920	.914
				90% CI	.025 - .039		
TBI 5-Factor	Female Norm	1475.651	727	< .0001	.027	.925	.919
				90% CI	.025- .029		
TBI 5-Factor	Male Norm	1293.829	727	< .0001	.027	.910	.903
				90% CI	.024 - .029		
TBI 5-Factor	Norm A	1385.415	727	< .0001	.027	.910	.904
				90% CI	.025 - .029		
TBI 5-Factor	Norm B	1472.530	727	< .0001	.029	.916	.910
				90% CI	.027 - .031		
TBI 5-Factor	TBI	812.391	727	= .0148	.022	.962	.959
				90% CI	.011 - .031		

due to differences in the item to factor structure across models. One exception is for Norm B Factor-4, TBI Factor-5 which covers similar items (see Appendix 2). Comparison of the coefficients in these factors found the Norm B candidate model was .103 greater than the coefficient for the TBI candidate model. Overall the average for the factor coefficient is higher than the Norm B candidate model (0.313) than in the TBI candidate model (0.279). The factor rho coefficient results subjectively prefer the structure defined by the Norm B candidate model.

Table 6.5

Factor rho coefficients for the Hy3 candidate models

	Factor Rho Coefficient				
Candidate Model	1	2	3	4	5
Norm B	.370	.245	.166	.782	n/a
TBI	.455	.254	.278	.128	.679

Review of the factor structure for each candidate model

To facilitate a review of the factor structure within each candidate model domains of items were identified. Domains were defined as groups of 3 or more items which were allocated in a consistent pattern across the Norm B and TBI candidate models (see Table 6.6). There was no requirement that the items within a domain be similar in content. Domains were defined purely based on similarity of the candidate models' structures, defined from the factor analysis procedure describe above. There were six domains which consisted of 32 items that were found in both candidate models. This left 8 items not allocated to a domain.

Review of the domain to factor structure finds a preference for the Norm B candidate model. Domains A and B load onto a single factor in the Norm B candidate model but separate factors in the TBI candidate model. Domains C and D load onto a single factor in the TBI candidate model but separate factors in the Norm B candidate model. In the Norm B model Domain D and Domain E load onto a single factor, whilst Domain E loads independently from the other domains onto a factor in the TBI model. Domain F loads independently onto a factor in both models.

The primary model differences are around the allocation to factors of Domains C, D and E. Reviewing the content of these domains may inform as to the preferred structure of the factor model. Content in Domain C reflects beliefs about other people. Content in Domain D reflects responses to other people. The content in Domain E reflects anger and frustration. Using a cognitive model of psychopathology where thoughts precede emotions and behaviours the allocation of Domains D and E onto a single factor is logical. This is the case for the Norm B candidate model but not the TBI candidate model. Additionally the

association between Domains A and B in the Norm B model generates a larger set of items that appear thematically mood related, whilst these domains are separated in the TBI model. Thus a review of the content domains finds a theoretical preference for the Norm B candidate model. A review of the 8 items not allocated to a domain was next completed.

Item 7 (reading crime articles in newspaper) is allocated with Domain C items in the Norm B model, and Domain E items in the TBI model. Neither Domain C content (beliefs about other people) nor Domain E content (anger and frustration) appear more suitably related to item 7.

Item 40 (regular cranial pain) is allocated with Domains D and E in the Norm B model, and Domain B in the TBI model. The content of item 40 is not clearly related to that of Domain D (response to other people) or the content of Domain B which is eclectic with a non-specific relation to worry and anxiety. An argument could be made that head pain, such as headaches, would relate to symptoms of stress like anger and frustration found in Domain E. Item 40 is allocated to the same factor as Domain E in the Norm B candidate model.

Item 44 (Flushes) is allocated with Domain C in the Norm B model, and Domain B in the TBI model. Item 44 content is not clearly related to that of Domain C (beliefs about other people). However, the item content does reflect Domain B's theme of non-specific anxiety and worry as seen in the TBI candidate model.

Item 129 (my behaviour is in response to that of others) is allocated with Domains D and E in the Norm B model, and Domain F in the TBI model. The item content relates well to that of Domain E (anger and frustration), and importantly it strengthens the theme suggested for Domain D (response to other people). An argument could be made that the item also relates to Domain F content of shyness, however as behavioural response can take many forms a preference is for the Norm B model structure for this item.

Item 157 (I do not care what others think of me) is allocated with Domains D and E in the Norm B model, and Domain A in the TBI model. An argument could be made for the relation of this item to both Domain A (present mood) and Domain E (anger and

frustration). However the content appears better suited to that of Domain D (responses to other people) as is found in the Norm B model.

Item 172 (my hand shakes when I do something) is allocated with Domain C in the Norm B model, and Domain B in the TBI model. The item appears better suited to Domain B (non-specific anxiety and worry), than Domain C (beliefs about other people), as found in the TBI model.

Item 230 (friendliness) is allocated with Domains A and B in the Norm B model, and Domain E in the TBI model. The notion of friendliness appears better suited to the contents of Domains A (present mood) and B (non-specific anxiety and worry), than to Domain E (anger and frustration), as seen in the Norm B model.

Item 263 (disgust when lawyers free a criminal) is allocated with Domain C in the Norm B model, and Domain A in the TBI model. An argument could be made for an association between the item content and Domain C (beliefs about other people), or Domain A (present mood), providing no clear model preference.

Table 6.6

Item content and Domain allocations for Hy3 candidate models

Items	Content	Domain	Factor Norm B	Factor TBI
9	Days full with activities of interest	A	1	1
65	Often feels blue	A	1	1
95	Often feels happy	A	1	1
125	Happiness at home is equivalent to others	A	1	1
148	Best I've felt in my life	A	1	1
11	Often a lump in the throat	B	1	3
31	Difficulty focusing on tasks	B	1	3
159	Never fainted	B	1	3
166	Sexual concerns	B	1	3
26	Best to keep quiet when in trouble	C	2	2
58	People overstate to get sympathy	C	2	2

Items	Content	Domain	Norm B	TBI
76	It is hard to convince people the truth	C	2	2
81	People lie to get ahead	C	2	2
110	People will behave unfairly to benefit themselves	C	2	2
124	Suspicious of people who are nice	C	2	2
135	Indecision has cost them opportunities	C	2	2
193	Do not step on sidewalk cracks	C	2	2
241	Safer to trust no-one	C	2	2
253	Drinks large amounts of water daily	C	2	2
14	Likes detective or mystery stories	D	3	2
98	Wants to do the opposite of what a bossy person requests, even though they are right	D	3	2
151	Resents having to admit I was fooled	D	3	2
29	Sometimes wants to swear	E	3	4
115	Unconcerned about the sight of blood	E	3	4
116	Gets upset without reason	E	3	4
213	Quick to anger, quick to soothe	E	3	4
218	Gets so restless cannot stay seated	E	3	4
161	Tries to hide shyness	F	4	5
167	Difficulty talking to new people	F	4	5
185	Shyness bothers them	F	4	5
243	Not sure what to say in a groups of people	F	4	5
265	Will not speak to someone first	F	4	5
7	Likes reading crime articles in newspaper	**	2	4
40	Regular pain all over the head	**	3	3
44	Hot flushes	**	2	3
129	Behaviour is in response to that of others	**	3	5
157	Does not care what others think of them	**	3	1
172	Hand shakes when they do something	**	2	3
230	Friendliness	**	1	4
263	Disgust when lawyers free a criminal	**	2	1

** Items not allocated to a Domain

Summary of the theoretical review

The review of the content domains found a theoretical preference for the Norm B candidate model. Reviewing the items not allocated to a domain also found a preference for the Norm B candidate model. No model preference was determined for items 7 and 263. While four of the remaining six items had preferable factor allocations in the Norm B candidate model (40, 129, 157 and 230). With only two preferred in the TBI candidate model (44 and 172).

Selection of the 4-Factor Norm B candidate model as the baseline model

In conclusion the Norm B candidate model is the preferred model. The Norm A candidate model was found to be inadmissible during the replication analysis and was thus removed from further consideration. During the replication analysis the Norm B candidate model produced measures of better model performance than the TBI candidate model. The Norm B model produced slight empirical preference over the TBI candidate model when comparing a measure of construct validity (factor rho coefficient). Finally the Norm B candidate model was preferred over the TBI model during the theoretical review. Table 6.7 shows the selected baseline model comprises four factors of 'Mood', 'Beliefs about others, 'Reactions to others' and 'Shyness'.

Table 6.7

Item to factor structure in the Hy3 4-Factor baseline model

Factor Name	Items included in the factor
1. 'Mood'	9, 11, 31, 65, 95, 125, 148, 159, 166 and 230.
2. 'Beliefs about others'	7, 26, 44, 58, 76, 81, 110, 124, 135, 172, 193, 241, 253 and 263
3. 'Reactions to others'	14, 29, 40, 98, 115, 116, 129, 151, 157, 213 and 218
4. 'Shyness'	161, 167, 185, 243 and 265

6.2 Measurement Invariance Testing

Measurement invariance testing followed a hybrid of the procedures recommended by Millsap and Yun Tein (2004) and, Muthen and Muthen (1998-2010) for the analysis of dichotomous data. In Test 1 the residuals are constrained to equality across groups with

loadings and thresholds allowed to vary (except for loadings on marker variables). In Test 2 all residuals, loadings and threshold are constrained to equality across groups, this is a test of strict invariance. As proposed by Meade et al. (2008) a decrease by .002 in the CFI value was selected as minimum the criteria to meet invariance. Should the model fail to meet the criteria of strict invariance, then sequentially one item at a time, parameters (loadings and thresholds) will be freed to find a model with the maximum constraint that meets the requirements of partial invariance. This approach follows the backwards elimination procedure (Cheung & Rensvold, 1999). The two samples included for the invariance testing were the TBI sample (n=233) and a Community sample (the gender matched MMPI-2 normative sample, n=1766, male = 1104 and female = 662).

Invariance tests detailed in Table 6.8 show Test 1 produced acceptable fit indices even with the residuals constrained to equivalence. Test 2 generated a decrease in the CFI value of .154, thus failing the test of strict invariance. Parameters of the factor model were freed sequentially and allowed to vary across groups based on the modification indices. This procedure established that sequentially freeing loadings and thresholds for items 44, 193, 172, 29, 11, 253, 40, 263, 148, 241, 31, 14, 115, 124, 161, 185, 230, 129, 76 and135 was sufficient to meet the requirements for partial invariance (Test 3 in Table 6.8). That is the RMSEA was unchanged and the CFI decreased by .002 when compared with the fit indices generated in Test 1. See Table 6.14 for details of the standardized factor loadings for the TBI and Community samples in the partial invariance model.

6.3 Practical Impact Analysis

The finding of partial invariance makes further investigation necessary to determine whether there is any practical impact from finding non-invariant items (Millsap & Kwok, 2004). To investigate the practical impact of the partially invariant measurement model described above, an evaluation of sensitivity and specificity values under conditions for Test 2 and Test 3 was completed. Test 2 represents the hypothetical case of strict invariance and Test 3 represents the observation of partial invariance. As previously noted in Chapter Four, the procedure described by Millsap and Kwok helps establish what impact on diagnostic accuracy with the MMPI-2 Hy3 is observed from the failure to find strict invariance. Therefore a comparison of the sensitivity and specificity values in the partial

Table 6.8

Measurement invariance testing of the Hy3 4-Factor model across the Community (n=1766) and TBI (n=233) samples.

Invariance Model	*WLSMV*χ^2	df	p	χ^2 diff	RMSEA	CFI	TLI
Test 1 residuals equal	2395.741	1450	< .0001		.026	.921	.915
Test 2 residuals, loadings and thresholds equal	4322.065	1522	< .0001	p <.0001*	.043	.767	.761
Test 3 Partial invariance	2452.469	1482	< .0001	p =.0001*	.026	.919	.915

**Note this is the χ^2 difference test compared with the Test 1 model*

invariance (Test 3) and strict invariance (Test 2) condition is required. The item loadings freed to facilitate partial invariance are associated with Factor-1, Factor-2, Factor-3 and Factor-4. Therefore the practical impact analysis of the failure to find strict invariance is necessary for all these factors.

Table 6.9 lists the observed score and factor score cut-points for each invariance condition. The observed scores remain unchanged across invariance conditions as only the psychometric properties of the measurement model vary with consequences for the factor scores. Information about the procedure followed to calculate these values is provided in Chapter Four (see Section 4.5). To briefly reiterate an observed composite score that was as close to the 93.32 percentile rank in the Community sample was used at the cut-point to represent a diagnosis of psychopathology. The actual percentile rank of the observed composite score was used as the cut-point for the factor score from the Community Sample to represent the presence or absence of psychopathology.

In Factor-2 the observed composite was initially found with a percentile rank of 89.2. This percentile rank was then employed to calculate the factor scores and the sensitivity values. The data was able to generate a sensitivity value in the partial invariance condition but not

the strict invariance, which precluded a comparison. This situation arose because the highest calculated factor score for the TBI sample in the strict invariance condition was 0.384, which was equivalent to the 80.5 percentile rank for the Community sample. Therefore no TBI sample cases produced a factor score above the cut-point. As a result there were zero cases that represented the presence of psychopathology. The distribution of factor scores in the TBI sample resulted in the true-positive and false-positive quadrants having a value of zero, and not being able to calculate a sensitivity value. This problem was overcome by altering the percentile rank to 62.9 as seen in Table 6.9.

Summary of Practical Impact Analysis

Table 6.10 shows the sensitivity and specificity values for the TBI sample in both the partial invariance and strict invariance conditions. All values generated in Table 6.10 were calculated using the VassarStats calculator (Lowry, 1998-2012). Millsap and Kwok (2004) recommend reviewing the changes in sensitivity and specificity values between the two invariance conditions. For each factor, if the absolute value of sensitivity or specificity for the TBI sample in the partial invariance condition is less than the lower bound of the 95% CI in the strict invariance condition, then the non-invariant items fail the tests of no practical impact.

Table 6.9

Observed score, factor score and percentile rank cut points used to calculated sensitivity and specificity values for invariance conditions in each Hy3 factor.

		Partial Invariance		Strict Invariance	
	Percentile rank	Observed Score	Factor Score	Observed Score	Factor Score
Factor-1	90.5	4	1.954	4	1.981
Factor-2	62.9	8	0.172	8	0.128
Factor-3	85.2	6	0.620	6	0.568
Factor-4	66.9	4	0.679	4	0.481

In none of Factor-1, Factor-2 or Factor-3 were the sensitivity or specificity values for the partial invariance condition less than the lower bound 95% CI in the strict invariance condition (see Table 6.10). In other words for these three factors there was no significant impact on the sensitivity or specificity values from the failure to find strict invariance. Therefore retaining the invariant items and assuming full invariance does not compromise criterion relate validity. The practical impact analysis supported retaining all items from Factor-1, Factor-2 and Factor-3 when assessing persons with a TBI.

In Factor-4 the sensitivity value in the partial invariance condition was less than the lower bound 95% CI in the strict invariance condition. Whereas the specificity value in the partial invariance condition was above the upper bound 95% CI in the strict invariance condition. The findings suggested the non-invariant items from Factor-4 (161 and 185) fail the tests of no practical impact. Only the Factor-4 items that met the criteria of strict invariance (167, 243 and 265) can be confidently retained. Tables 6.11 and 6.12 show the significant reduction in the specificity value, as represented by the increase in false-negative cases seen in the partial invariance condition compared with the strict invariance condition. The false-negative cases are the values where the condition is present and the test is negative. This value increased from zero in the strict invariance condition to 40 in the partial invariance condition.

In contrast support for retaining all items was the significant correlation between factor scores in the partial invariance condition and the factor scores in the strict invariance condition observed on Factor-1 (Spearman's rho = .998, $p < .01$), Factor-2 (Spearman's rho = .757, $p < .01$), Factor-3 (Spearman's rho = .965, $p < .01$) and Factor-4 (Spearman's rho = .979, $p < .01$).

6.4 Comparison of threshold parameters

Following the procedure outlined in the results section for Hs1 (see Section 5.4), the threshold values were compared for the items that failed the test of strict invariance. Borbsoom (2006) proposes that when the sources of item variance between the samples cancel each other out the result is no impact from the failure of strict invariance. If a lower threshold was observed in the TBI group, this reflects an increased likelihood that an item

with be endorsed for psychopathology, and vice versa. A comparison of the threshold parameters can help determine whether the number of lower and higher thresholds for the TBI group cancels each other out. Or whether there is a bias in the direction of higher or lower threshold parameters in the TBI group.

Table 6.13 shows the threshold values for the TBI and the Community samples, for the items that failed the test of strict invariance. Comparisons between the two threshold scores were completed by using the geometric mean of the standard error for each sample (Howell, 2002). As can be seen in Table 6.13, item 185 is equally likely to be endorsed by either sample. In total 11 of the 20 items show higher thresholds in the TBI sample, namely, items 31, 76, 124, 129, 135, 148, 161, 193, 230, 241 and 253. Eight items show lower thresholds in the TBI sample, namely, items 11, 14, 29, 40, 44, 115, 172 and 263.

Therefore the difference in threshold values across the groups does not uniformly increase the likelihood a person would endorse an item. The mixture of 11 higher and 8 lower thresholds shown by the TBI sample reduces concerns for consistent bias in Hy3 scale scores.

Items 161 and 185 failed both the test of strict invariance and the tests for no practical impact. Therefore it is worth considering the analysis of threshold value for these items in particular. Item 185 is equally likely to be endorsed by each sample suggesting there is no concern with endorsement of this item falsely inflating a scale score. Item 161 shows a higher threshold in the TBI sample, which again suggests there is no concern with endorsement of this item falsely inflating scale scores. Importantly the analysis to assess difference in likelihood of item endorsement by the TBI sample found no evidence that either item would inflate scale scores.

Table 6.10

Hy3 sensitivity and specificity analysis (with 95% confidence intervals: CI) for factors across invariance conditions in TBI sample.

	Factor 1		Factor 2		Factor 3		Factor 4	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Partial Invariance	.750	1.000	.475	1.000	.688	.657	.365	.994
	.681 – .808	.893 – 1.000	.396 – .555	.938 – 1.000	.499 – .833	.586 – .721	.250 – .496	.963 – .999
Strict Invariance	.750	1.000	1.000	.689	.667	.628	1.000	.912
	.681 – .808	.893 – 1.000	.463 – 1.000	.623 – .747	.387 – .870	.560 – .692	.396 – 1.000	.866 – .945

Note – values were calculated using the VassarStats calculator (Lowry, 1998-2012)

Table 6.11

Quadrant frequencies for the TBI sample in Hy3 Factor-4 partial invariance condition

	Condition Present	Condition Absent
Test Positive	23	1
Test Negative	40	169

Table 6.12

Quadrant frequencies for the TBI sample in Hy3 Factor-4 strict invariance condition

	Condition Present	Condition Absent
Test Positive	4	20
Test Negative	0	209

6.5 Results of Gass (1991) correction procedure analysis

Two of the 40 items analysed as part of Hy3 form part of the Gass (1991) correction procedure, they are item 31 on Factor-1 and item 172 on Factor-2. Both items required their thresholds and loading to be freed when defining the partial invariance model. This result identifies these items as sources of measurement model variance between the samples.

The practical impact analysis supported retaining all items from Factor-1 and Factor-2. This finding suggests the removing items 31 and 172 from assessments of person with a TBI is inappropriate. Additionally the review of item thresholds (see Table 6.13) observed item 31 as being more likely to be endorsed by the TBI sample, item 172 was less likely to be endorsed. With one item being more likely for endorsement and one less likely there is no clear bias towards an increase in scale scores.

Table 6.13

Mean threshold value and standard errors (SE) for the Community and the TBI samples for items freed when defining the Hy3 partial invariance model

	Community Thresholds		TBI Thresholds		Geometric	TBI compared
Item	Mean	SE	Mean	SE	Mean of SE	with Community
11	2.025	0.100	0.162	0.222	0.172	<
14	0.449	0.031	-0.222	0.089	0.067	<
29	1.059	0.038	-1.399	0.170	0.123	<
31	1.307	0.062	3.261	0.699	0.496	>
40	1.993	0.106	-0.261	0.111	0.109	<
44	1.977	0.088	-1.241	0.196	0.152	<
76	-0.425	0.039	0.534	0.138	0.101	>
115	0.480	0.031	-0.498	0.092	0.069	<
124	-0.803	0.052	1.081	0.226	0.164	>
129	-0.613	0.038	0.319	0.115	0.086	>
135	-0.668	0.048	0.706	0.220	0.159	>
148	-0.022	0.032	2.965	0.602	0.426	>
161	-0.582	0.049	0.548	0.096	0.076	>
172	1.532	0.059	-1.052	0.184	0.137	<
185	-0.231	0.054	-0.105	0.105	0.083	=
193	-1.629	0.059	1.138	0.149	0.113	>
230	-0.619	0.032	0.224	0.186	0.133	>
241	-1.133	0.053	1.051	0.178	0.131	>
253	-1.230	0.042	1.111	0.143	0.105	>
263	0.697	0.034	-0.786	0.116	0.085	<

Table 6.14

Standardized factor loadings for the 4-Factor CFA model of MMPI-2 Hy3 in traumatic brain injury (TBI, n=233) and Community (Norm, n=1766) samples.

Parameter estimates (PE) and standard errors (SE) are derived from the partially invariant measurement model (Test 3 in Table 6.8). Also shown is the variance explained by the model in each item (R^2)

	Factor 1		Factor 2		Factor 3		Factor 4		R^2	
Item	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM
65	.919(.020)	.858(.027)							.156(.038)	.745(.063)
9	.741(.036)	.622(.030)							.451(.053)	.748(.032)
11	.489(.087)	.505(.063)							.761(.085)	.989(.008)
31	.790(.069)	.576(.042)							.375(.110)	.651(.035)
95	.826(.029)	.726(.028)							.318(.048)	.513(.038)
125	.736(.036)	.616(.033)							.459(.054)	.621(.040)
148	.546(.138)	.395(.039)							.702(.151)	.763(.034)
159	.250(.044)	.183(.031)							.938(.022)	.525(.038)
166	.660(.042)	.534(.034)							.564(.056)	.998(.003)
230	-.019(.098)	-.096(.044)							1.00(.004)	.844(.031)
81			.496(.055)	.528(.030)					.754(.055)	.863(.024)
7			.190(.040)	.206(.033)					.964(.015)	.984(.008)
26			.372(.050)	.400(.031)					.861(.037)	.967(.011)
44			-.508(.085)	-.517(.052)					.742(.087)	.508(.045)
58			.504(.056)	.536(.031)					.746(.056)	.714(.037)
76			.460(.079)	.604(.028)					.789(.073)	.384(.051)
110			.471(.054)	.502(.032)					.778(.051)	.790(.042)
124			.777(.057)	.698(.027)					.396(.089)	.343(.053)

	Factor 1		Factor 2		Factor 3		Factor 4		R^2	
Item	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM	TBI	NORM
135			.750(.060)	.690(.028)					.438(.090)	.858(.039)
172			-.590(.072)	-.459(.046)					.652(.085)	.853(.024)
193			.372(.092)	.377(.052)					.862(.068)	.785(.029)
241			.605(.069)	.609(.032)					.634(.084)	.991(.008)
253			.141(.118)	.256(.046)					.980(.033)	.629(.039)
263			.071(.099)	.214(.040)					.995(.014)	.262(.059)
116					.719(.039)	.591(.029)			.483(.057)	.934(.023)
14					-.102(.092)	.043(.039)			.990(.019)	.840(.025)
29					.577(.087)	.171(.048)			.667(.100)	.954(.017)
40					-.548(.075)	-.586(.053)			.700(.082)	.633(.040)
98					.533(.046)	.407(.032)			.716(.049)	.971(.016)
115					.034(.095)	-.104(.039)			.999(.006)	.668(.048)
129					.647(.067)	.487(.035)			.582(.087)	.656(.062)
151					.490(.047)	.370(.032)			.759(.046)	.733(.053)
157					-.175(.047)	-.125(.033)			.969(.017)	.713(.033)
213					.506(.048)	.383(.031)			.744(.048)	.264(.046)
218					-.595(.045)	-.464(.031)			.646(.053)	.957(.014)
185							.874(.044)	.785(.032)	.774(.088)	.635(.034)
161							.475(.092)	.810(.033)	.905(.059)	.721(.032)
167							.309(.096)	.702(.032)	.236(.076)	.613(.037)
243							.922(.033)	.859(.034)	.149(.061)	.474(.041)
265							.735(.065)	.606(.033)	.460(.095)	.834(.026)

Chapter Seven – Results Sc8

7.1 Defining a Baseline model

Candidate Model definition - Norm A Sample

After removing 10 items that were previously analysed with Hs1 or Hy3, there are 68 items remaining for analysis from Sc8. Using these 68 items from the Sc8 scale, EFA showed that a 12-Factor model fitted the data [WLSMV χ^2 (1528, N = 1228) = 1612.559, p = .065]. Therefore 12 factors were set as the upper limit for the candidate model in the Norm A sample. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed all models with greater than seven factors were inadmissible due to one or more factors failing the requirement to have a minimum of three items. All remaining simple structure CFA models met the admissibility criteria.

As the 6-Factor model was not nested under the 7-Factor model a χ^2 difference test was unable to be completed. Table 7.1 shows the fit statistics for the 6-Factor model were equal to or better than the 7-Factor model which supported selecting of the more parsimonious model. Table 7.1 shows significant χ^2 difference tests were produced for the 5-Factor model and each subsequent lower factor model supporting the selection of the 6-Factor model as the candidate model. The exception to this was the 3-Factor model which was not nested under the 4-Factor model. The 3-Factor model produced the same fit statistics as the 4-Factor model and would be preferred, however the 4-Factor model produced a significant χ^2 difference tests compared with the 5-Factor model. The results from the χ^2 difference tests support the selection of the 6-Factor model as the candidate model. The RMSEA, CFI and TLI were best in the 6-Factor model offering additional evidence to support this selection. The RMSEA, TLI and CFI for the 6-Factor model indicate reasonable fit under the model fit criteria.

Respecification Norm A 6-Factor model

To generate modification indices for residual correlation in pair of items in Mplus it is necessary to specify at least one item residual correlation parameter in the factor model.

Items 168 and 229 loaded onto the same factor in all candidate models and both have content related to blank spells. These items were specified as having correlated residuals in the factor model.

Table 7.1

Confirmatory factor analysis of potential Sc8 candidate models in the Norm A sample (n=1228)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
7 Factor	2858.994	2189	<.0001	.016	.912	.909	
6 Factor	2850.086	2195	<.0001	.016	.914	.911	Not nested
5 Factor	2899.441	2200	<.0001	.016	.909	.905	5 v 6 (p < .0001)
4 Factor	2968.819	2204	<.0001	.017	.900	.897	4 v 5 (p < .0001)
3 Factor	2971.218	2207	<.0001	.017	.900	.897	Not nested
2 Factor	3280.036	2209	<.0001	.020	.860	.856	2 v 3 (p < .0001)
1 Factor	3463.931	2210	<.0001	.021	.836	.831	1 v 2 (p < .0001)

Norm A 6-Factor post hoc model 1

This model followed the same structure as Norm A 6-Factor model with the addition of a residual correlation between items 168 and 229. The resulting 6-Factor post hoc model 1 was significantly better fitting compared with the 6-Factor model [WLSMV χ^2 (1, N = 1228) = 6.564 p = .0104].

Compared with the 6-Factor model, the CFI increased by .001, TLI increased by .001 and RMSEA was unchanged. The results supported the Norm A 6-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (2194, N = 1228) = 2846.220 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 299 and 325 may improve the model performance. Review of the item content found items referring to concentration and focus difficulties supporting a theoretical basis of

residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 2

This model followed the same structure as Norm A 6-Factor post hoc model 1 with the addition of a residual correlation between items 299 and 325. The resulting 6-Factor post hoc model 2 was significantly better fitting compared with the 6-Factor post hoc model 1 [WLSMV χ^2 (1, N = 1228) = 33.641 p < .0001].

Compared with the 6-Factor post hoc model 1, the CFI increased by .002, TLI increased by .002 and RMSEA was reduced by .001. The results supported the Norm A 6-Factor post hoc model 2 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (2193, N = 1228) = 2825.857 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 6 and 90 may improve the model performance. Review of the item content found items referring to father as being a good man and love for father respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 3

This model followed the same structure as Norm A 6-Factor post hoc model 2 with the addition of a residual correlation between items 6 and 190. The resulting 6-Factor post hoc model 3 was significantly better fitting compared with the 6-Factor post hoc model 2 [WLSMV χ^2 (1, N = 1228) = 10.525 p = .0012].

Compared with the 6-Factor post hoc model 2, the CFI increased by .002, TLI increased by .002 and RMSEA was unchanged. The results supported the Norm A 6-Factor post hoc model 3 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (2192, N = 1228) =

2810.465 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 221and 287 may improve the model performance. Review of the item content found items referring to having dreams best kept to them and often dreaming about sex respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 4

This model followed the same structure as Norm A 6-Factor post hoc model 3 with the addition of a residual correlation between items 221 and 287. The resulting 6-Factor post hoc model 4 was significantly better fitting compared with the 6-Factor post hoc model 3 [WLSMV χ^2 (1, N = 1228) = 33.408 p < .0001].

Compared with the 6-Factor post hoc model 3, the CFI increased by .002, TLI increased by .002 and RMSEA was unchanged. The results supported the Norm A 6-Factor post hoc model 4 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (2191, N = 1228) = 2797.517 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 32 and 316 may improve the model performance. Review of the item content found items referring to peculiar and strange experiences, and peculiar and strange thoughts supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 5

This model followed the same structure as Norm A 6-Factor post hoc model 4 with the addition of a residual correlation between items 32 and 316. The resulting 6-Factor post hoc model 5 was significantly better fitting compared with the 6-Factor post hoc model 4 [WLSMV χ^2 (1, N = 1228) = 36.198 p < .0001].

Compared with the 6-Factor post hoc model 4, the CFI increased by .001, TLI increased by .001 and RMSEA was unchanged. The results supported the Norm A 6-Factor post hoc model 5 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 5 still displayed a significant lack of fit [WLSMV χ^2 (2190, N = 1228) = 2784.690 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 16 and 316 may improve the model performance. Review of the item content found items that referred to having occasional bad thoughts and having strange and peculiar thoughts respectively, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 6

This model followed the same structure as Norm A 6-Factor post hoc model 5 with the addition of a residual correlation between items 16 and 316. The resulting 6-Factor post hoc model 6 was significantly better fitting compared with the 6-Factor post hoc model 5 [WLSMV χ^2 (1, N = 1228) = 26.388 p < .0001].

Compared with the 6-Factor post hoc model 5, the CFI increased by .001, TLI increased by .001 and RMSEA was unchanged. The results supported the Norm A 6-Factor post hoc model 6 as preferred even though the overall test of fit showed that the Norm A 6-Factor post hoc model 6 still displayed a significant lack of fit [WLSMV χ^2 (2189, N = 1228) = 2774.957 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 276 and 192 may improve the model performance. Review of the item content found items referring to love for their mother and believing their mother was a good woman, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm A 6-Factor post hoc model 7

This model followed the same structure as Norm A 6-Factor post hoc model 6 with the addition of a residual correlation between items 192 and 276. There was no significant difference in model fit comparing the resulting 6-Factor post hoc model 7 with the 6-Factor post hoc model 6 [WLSMV χ^2 (1, N = 1228) = 2.618 p = .1057]. Under the preference for model parsimony Norm A 6-Factor post hoc model 6 is preferred as it has less specified parameters.

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported and being specified within the model for method effects. Reviewing the modification indices found no further respecifications. Therefore the Norm A 6-Factor post hoc model 6 is the preferred candidate mode (see Appendix 3 for details).

It is important to note that there were additional pairs of items considered for specification of residual correlation. However, despite thematic similarities, it was decided the item content was sufficiently different. For example item 12 (satisfactory sex life) and items 268 (wishes they were not bothered by sexual thoughts) met the modification indices criterion for specification but despite a similar theme, knowing if one were bother by sexual thoughts does not inform as to whether the same person may consider their sex life adequate. The subjective nature of this approach is an accepted limitation and will only be resolved by further research.

Candidate Model definition - Norm B Sample

EFA showed the 13-Factor model fitted the data [WLSMV χ^2 (1472, N = 1227) = 1554.239, p = .0667]. Thus 13 factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed all models with greater than five factors were inadmissible due to one or more factors failing the requirement to have a minimum of three items. Additionally the 3-Factor model and the 4-Factor model were inadmissible due to excessive correlations between factors. Therefore only the 1-Factor, 2-Factor and 5-Factor models were admissible. Table 7.2 indicates significant χ^2 difference tests were produced for the 2-Factor model compared with the 5-Factor model, and the 1-Factor model compared with the 2-factor model. The

results from the χ^2 difference tests supported selecting the 5-factor model as the candidate model. The RMSEA, CFI and TLI were best in the 5-Factor model offering additional evidence to support this selection. The RMSEA, TLI and CFI for the 5-Factor model indicate reasonable fit under the model fit criteria.

Table 7.2

Confirmatory factor analysis of potential Hy3 candidate models in the Norm B sample (n=1227)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
5 Factor	2731.029	2200	<.0001	.014	.933	.930	
2 Factor	3021.716	2209	<.0001	.017	.897	.894	2 v 5 ($p < .0001$)
1 Factor	3253.227	2210	<.0001	.020	.868	.864	1 v 2 ($p < .0001$)

Respecification Norm B 5-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model. The specification of item correlations was based on method effects from shared item content among the pairs of items. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this prediction.

Norm B 5-Factor post hoc model 1

This model followed the same structure as Norm B 5-Factor model with the addition of a residual correlation between the following item pairs: 168 and 229; 299 and 325; 6 and 90; 221 and 287; 32 and 316; 16 and 316. The resulting 5-Factor post hoc model 1 was significantly better fitting compared with the 5-Factor model [WLSMV χ^2 (6, N = 1227) = 64.869 $p < .0001$].

Compared with the 5-Factor model, the CFI increased by .006, TLI increased by .007 and RMSEA was reduced by .001. The results supported the Norm B 5-Factor post hoc model 1 as preferred even though the overall test of fit showed that the Norm B 5-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (2194, N = 1227) = 2674.571 $p < .0001$].

Inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between item pairs 168 with 229; and, 16 with 316. This observation supported removing the residual correlations between these two pairs of items and reviewing the model fit.

Norm B 5-Factor post hoc model 2

This model followed the same structure as Norm B 5-Factor post hoc model 1 with the removal of residual correlations between item pairs 168 and 229; as well as, 16 and 316. There was no significant difference in model fit comparing the 5-Factor post hoc model 2 with the 5-Factor post hoc model 1 [WLSMV χ^2 (2, N = 1227) = 4.707 p = .0950]. Under the preference for model parsimony the Norm B 5-Factor post hoc model 2 was preferred as it included fewer parameters specified

Compared with the 5-Factor post hoc model 1 the CFI, TLI and RMSEA remained unchanged. The Norm B 5-Factor post hoc model 2 is preferred as the candidate model though the overall test of fit showed that the Norm B 5-Factor post hoc model 2 still displayed a significant lack of fit [WLSMV χ^2 (2196, N = 1227) = 2677.225 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 233 and 38 may improve the model performance. Review of the item content found items referring to problems initiating task and extended periods of apathy, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm B 5-Factor post hoc model 3

This model followed the same structure as Norm B 5-Factor post hoc model 2 with the addition of a residual correlation between items 38 and 233. The resulting 5-Factor post hoc model 3 was significantly better fitting compared with the 5-Factor post hoc model 2 [WLSMV χ^2 (1, N = 1227) = 51.679 p < .0001].

Compared with the 5-Factor post hoc model 2, the CFI increased by .003, TLI increased by .003 and RMSEA was unchanged. The results supported the Norm B 5-Factor post hoc

model 3 as preferred even though the overall test of fit showed that the Norm B 5-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (2195, N = 1227) = 2654.222 p < .0001].

Review of the modification indices suggested specifying a residual correlation between items 276 and 192 may improve the model performance. Review of the item content found items referring to love for their mother and believing their mother was a good woman, supporting a theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

Norm B 5-Factor post hoc model 4

This model followed the same structure as Norm B 5-Factor post hoc model 3 with the addition of a residual correlation between items 192 and 276. The resulting 5-Factor post hoc model 4 was significantly better fitting compared with the 5-Factor post hoc model 3 [WLSMV χ^2 (1, N = 1227) = 4.706 p = .0301].

Compared with the 5-Factor post hoc model 3, the CFI, TLI and RMSEA were unchanged. The results supported the Norm B 5-Factor post hoc model 4 as preferred even though the overall test of fit showed that the Norm B 5-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (2194, N = 1227) = 2649.038 p < .0001].

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported and being specified within the model for method effects. Reviewing the modification indices found no further respecifications. Therefore the Norm B 5-Factor post hoc model 4 is the preferred candidate model (see Appendix 3 for details).

Candidate Model definition - TBI Sample

EFA showed the 5-Factor model fitted the data [WLSMV χ^2 (1948, n = 229) = 2044.106, p = .0635]. Therefore five factors were set as the upper limit for the candidate model. When this best-fitting EFA was converted to a simple structure CFA, WLSMV estimation showed the 5-Factor model was inadmissible due to having item 192 producing a standardized

loading value greater than 1 on Factor-4. The 4-Factor model was inadmissible due to having a factor with less than three items. The 3-Factor model was inadmissible due to item 6 producing a standardized factor loading greater than one on Factor-2. Additionally, Factor-1 and Factor-3 were excessively correlated in the 3-Factor model. Table 7.3 shows the χ^2 difference test found the 1-Factor model significantly different compared with the 2-Factor model. Therefore the 2-Factor model was the selected candidate model. The RMSEA, CFI and TLI were best in the 2-Factor model offering additional evidence to support this selection. The RMSEA, TLI and CFI for the 2-Factor model indicate reasonable fit under the model fit criteria. Noteworthy is that both admissible models produced non-significant loadings for items 34, 210, 242 and 343.

Table 7.3

Confirmatory factor analysis of potential Sc8 candidate models in the TBI sample (n=233)

Model	χ^2	df	p	RMSEA	CFI	TLI	χ^2 difference test
2 Factor	2414.237	2209	=.0013	.020	.954	.943	
1 Factor	2554.907	2210	<.0001	.026	.923	.921	1 v 2 (p < .0001)

Respecification TBI 2-Factor model

The first respecification included the residual correlations specified in the Norm A candidate model, above. The specification of item correlations was based on method effects from shared item content among the pairs. Therefore it is reasonable to expect the method effects are present for all factor models comprising the same items. The first post hoc model tested this prediction.

TBI 2-Factor post hoc model 1

This model followed the same structure as TBI 2-Factor model with the addition of a residual correlation between the following item pairs: the following item pairs: 168 and 229; 299 and 325; 6 and 90; 221 and 287; 32 and 316; 16 and 316. The resulting 2-Factor post hoc model 1 was significantly better fitting compared with the 2-Factor model [WLSMV χ^2 (6, N = 229) = 18.733 p = .0046].

Compared with the 2-Factor model, the CFI increased by .002, TLI increased by .002 and RMSEA was unchanged. The results supported the TBI 2-Factor post hoc model 1 as preferred even though the overall test of fit showed that the TBI 2-Factor post hoc model 1 still displayed a significant lack of fit [WLSMV χ^2 (2203, N = 229) = 2398.337 p = .0021].

Inspection of the ratio between the mean correlation and standard error for each item pair, showed a non-significant relationship between item pairs: 168 and 229; as well as, 32 and 316. This observation supported removing the residual correlations between these pairs of items and reviewing the model fit.

TBI 2-Factor post hoc model 2

This model followed the same structure as TBI 2-Factor post hoc model 1 with the removal of residual correlations between item pairs; 168 and 229, and 32 and 316.

A significant difference was observed in the model fit between the resulting TBI 2-Factor post hoc model 2 compared with TBI 2-Factor post hoc model 1 [WLSMV χ^2 (2, N = 229) = 7.188 p = .0275]. The χ^2 difference test suggested the TBI 2-Factor post hoc model 1 is preferred. However, neither of the item pairs removed (168 and 229; 32 and 316) was suggested under the modification indices in the output from TBI 2-Factor post hoc model 2. This observation supported the excluding these correlated residuals from the specified model. Compared with the 2-Factor post hoc model 1 the CFI, TLI and the RMSEA remained unchanged in the TBI 2-Factor post hoc model 2.

The results supported the TBI 2-Factor post hoc model 2 as preferred even though the overall test of fit showed that the Norm B 4-Factor post hoc model 4 still displayed a significant lack of fit [WLSMV χ^2 (2205, N = 229) = 2401.429 p = .0020]. This decision was based on the finding of equal fit indices with a preference for the more parsimonious model. Additionally the modification indices failed to identify the removed item pairs as potentially improving the factor model in the modification indices output.

Review of the modification indices suggested specifying a residual correlation between items 276 and 192 may improve the model performance. Review of the item content found items referring to love for mother, and believing mother was a good woman, supporting a

theoretical basis of residual correlation. The potential for method effects along with the modification indices supported specifying a residual correlation between these items and reviewing the factor model fit.

TBI 2-Factor post hoc model 3

This model followed the same structure as TBI 2-Factor post hoc model 2 with the addition of a residual correlation between items 192 and 276. The resulting 2-Factor post hoc model 3 was significantly better fitting compared with the 2-Factor post hoc model 2 [WLSMV χ^2 (1, N = 229) = 8.856 p = .0029].

Compared with the 2-Factor post hoc model 3, the CFI increased by .002, TLI increased by .003 and RMSEA was reduced by .001. The results supported the TBI 2-Factor post hoc model 3 as preferred even though the overall test of fit showed that the TBI 2-Factor post hoc model 3 still displayed a significant lack of fit [WLSMV χ^2 (2204, N = 229) = 2392.538 p = .0028].

Review of the modification indices found no further pairs of items that were both empirically and theoretically supported as being specified within the model for method effects. Reviewing the modification indices found no further respecifications. Therefore the TBI 2-Factor post hoc model 3 is the preferred candidate mode (see Appendix 3 for details).

Candidate Model Replication Analysis

Each of the three candidate models were next examined for fit in other samples. Five samples were included in this phase of analysis. The three samples used to determine the candidate models along with the Female Norm and Male Norm samples.

Simple structure CFA showed the Norm A candidate model to be inadmissible on the Female Norm sample, with item 276 producing a standardised loading greater than 1. Additionally simple structure CFA showed the Norm A candidate model was inadmissible on the TBI sample as item 6 had a standardised loading greater than 1. The Norm B candidate model and the TBI candidate model were both admissible on all the replication

samples. Therefore Norm A candidate model was excluded for selection as the baseline model.

Table 7.4 indicates that the Norm B candidate model generated average fit indices indicative of a better fitting model then the TBI candidate model; CFI .933 and .910 respectively, TLI .930 and .907 respectively and RMSEA .015 and .017 respectively. After the removal of each sample used to generate the two remaining candidate models (i.e. the Norm B and TBI samples), the Norm B candidate model produced clearly better fit indices than the TBI candidate model. In this comparison respective Norm B and TBI candidate model averages on CFI were .920 against .894; on TLI .917 against .890; and on RMSEA .015 against .017.

It is important to consider that the TBI candidate model was defined from a clinical sample whilst the remaining four samples, and the Norm B candidate model, were non-clinical participants. However the replication process finds a clear preference for the Norm B candidate model. Noteworthy is that the Norm B candidate model generated fit indices in the TBI sample indicating a preferred solution compared with the TBI candidate model in the same sample.

Selection of the Norm B candidate model as the baseline model

The Norm B candidate model is selected as the baseline model with the additional comparisons on item residuals, construct validity and a theoretical review of the factor structure not required. In addition to the better performance of the Norm B candidate model in the replication analysis the fit indices generated by the TBI candidate model were outside the criteria designated as indicating reasonable fit in the male sample and the Norm A sample. This further supports rejecting the TBI candidate model and selecting the Norm B candidate model as the baseline model. Table 7.5 shows the selected baseline model comprises five factors of 'Unusual experiences and thoughts', 'Beliefs about family', 'Unusual thoughts', 'Beliefs about self and others' and 'Cognitive symptoms'.

Table 7.4

Confirmatory factor analysis with Sc8 candidate models in all samples

Candidate	Sample	χ^2	df	p	RMSEA	CFI	TLI
Norm B 5-Factor	Female Norm	2810.399	2194	< .0001	.014	.937	.935
				90% CI	.013 - .016		
Norm B 5-Factor	Male Norm	2772.739	2194	< .0001	.015	.912	.908
				90% CI	.013 - .017		
Norm B 5-Factor	Norm A	2876.664	2194	< .0001	.016	.911	.907
				90% CI	.014 - .018		
Norm B 5-Factor	Norm B	2649.038	2194	< .0001	.013	.942	.940
				90% CI	.011 - .015		
Norm B 5-Factor	TBI	2363.436	2194	= .0061	.018	.962	.961
				90% CI	.011 - .024		
TBI 2-Factor	Female Norm	3029.843	2204	< .0001	.017	.916	.913
				90% CI	.015 - .018		
TBI 2-Factor	Male Norm	2928.889	2204	< .0001	.017	.879	.875
				90% CI	.016 - .019		
TBI 2-Factor	Norm A	3067.692	2204	< .0001	.018	.887	.883
				90% CI	.016 - .019		
TBI 2-Factor	Norm B	2910.794	2204	< .0001	.016	.910	.907
				90% CI	.015 - .018		
TBI 2-Factor	TBI	2392.538	2204	= .0028	.019	.958	.957
				90% CI	.012 - .025		

Table 7.5

Item to factor structure for Sc8 5-factor baseline model

Factor Name	Items included in the factor
1. Unusual experiences & fearful thoughts	23, 38, 42, 48, 138, 145, 168, 170, 177, 180, 182, 210, 229, 234, 268, 274, 292, 295, 296, 298, 299, 307, 311, 319, 322, 329, 333 and 355
2. Beliefs about family	6, 90, 192, 276 and 343
3. Unusual thoughts	16, 32, 34, 35, 85, 106, 221, 242, 287, 290, 316, 320, 323 and 332
4. Beliefs about self and others	12, 17, 21, 22, 46, 92, 190, 256, 273, 277, 278, 279, 280, 281, 291 and 303
5. Cognitive symptoms	147, 165, 233, 252 and 325

7.2 Measurement Invariance Testing

Measurement invariance testing followed a hybrid of procedures recommended by Millsap and Yun Tein (2004) and, Muthen and Muthen (1998-2010) for the analysis of dichotomous data. In Test 1 the residuals are constrained to equality across groups with loadings and thresholds allowed to vary (except for loadings on marker variables). In Test 2 all residuals, loadings and threshold are constrained to equality across groups, this is a test of strict invariance. As proposed by Meade et al. (2008) a decrease by .002 or less in the CFI value was selected as the criteria to meet invariance. Should the model fail to meet the criteria of strict invariance, then sequentially one item at a time, parameters (loadings and thresholds) will be freed to find a model with the maximum constraint that meets the requirements of partial invariance. This approach follows the backwards elimination procedure (Cheung & Rensvold, 1999). The two samples included for the invariance testing were the TBI sample (n=229) and a Community sample (the gender matched MMPI-2 normative sample, n=1726, male = 1079 and female = 647).

Invariance tests detailed in Table 7.6 shows Test 1 produced acceptable fit indices even with the residuals constrained to equivalence. Test 2 generated a decrease in the CFI value of .121, thus failing the test of strict invariance. Parameters of the factor model were freed

sequentially and allowed to vary across groups based on the modification indices.

This procedure found that sequentially freeing loadings and thresholds for items 252, 210, 299, 355, 38, 274, 303, 290, 138, 322, 234, 170, 323, 332, 177, 242, 42, 182, 291, 92, 190, 281, 16, 287, 329, 278, 34 and 17 was sufficient to meet the requirements for partial invariance (Test 3 in Table 7.6). That is the RMSEA was unchanged and the CFI decreased by.002 when compared with the fit indices generated in Test 1. See Table 7.12 for details of the standardized factor loadings for the TBI and Community samples in the partial invariance model.

Table 7.6

Measurement Invariance testing of the Sc8 5-Factor model across the Community (n=1726) and TBI (n=229) samples.

Invariance Model	*WLSMV*χ^2	Df	p	χ^2 diff	RMSEA	CFI	TLI
Test 1 residuals equal	5312.084	4388	< .0001		.015	.934	.932
Test 2 residuals, loadings and thresholds equal	6790.834	4514	< .0001	p <.0001*	.023	.838	.837
Test 3 Partial invariance	5415.618	4458	< .0001	p =.0001*	.015	.932	.930

**Note this is the χ^2 difference test compared with the Test 1 model*

7.3 Practical impact analysis

The finding of partial invariance makes further investigation necessary to determine whether there is any practical impact from finding non-invariant items (Millsap & Kwok, 2004). To investigate the practical impact of the partially invariant measurement model described above, an evaluation of sensitivity and specificity values under conditions for Test 2 and Test 3 was completed. Test 2 represents the hypothetical case of strict invariance and Test 3 represents the observation of partial invariance. As previously noted in Chapter

Four, the procedure described by Millsap and Kwok helps ascertain what impact on diagnostic accuracy with the MMPI-2 Sc8 scale is found from the failure to find strict invariance. Therefore a comparison of the sensitivity and specificity values in the partial invariance (Test 3) and strict invariance (Test 2) condition is required. All items that are part of Factor-2 met the criteria of strict invariance. All other factors had at least one item parameter required to be freed to define the partial invariance model. Therefore the practical impact analysis of the failure to find strict invariance is necessary for Factor-1, Factor-3, Factor-4 and Factor-5.

In Table 7.7 the observed score and factor score cut-points for each invariance condition are shown. The observed scores remain unchanged across invariance conditions as only the psychometric properties of the measurement model vary with consequences for the factor scores. Information about the procedure followed to calculate these values is provided in Chapter Four (see Section 4.5). To briefly reiterate an observed composite score that was as close to the 93.32 percentile rank from the Community sample was used at the cut-point to represent a diagnosis of psychopathology. The actual percentile rank of the observed composite score was used as the cut-point for the factor score in the Community sample to represent the presence or absence of psychopathology.

In Factor-1 the observed composite cut-point was initially found with a percentile rank of 92.1. This percentile rank was then employed to determine the factor score cut-point. The data was able to generate a specificity value in the partial invariance condition but not the strict invariance, which precluded a comparison. This situation arose because the lowest calculated factor score for the TBI sample in the strict invariance condition was 1.305, which was equivalent to the 96.3 percentile rank for the Community sample. Therefore the lowest TBI sample factor score is greater than the cut-point factor score, resulting in zero cases that represent the absence of psychopathology. The distribution of factor scores in the TBI sample resulted in the true-negative and false-negative quadrants having a value of zero, and not being able to calculate a specificity value. This problem was overcome by altering the percentile rank to 98.2 as seen in Table 7.7.

Summary of Practical Impact Analysis

Table 7.8 shows the sensitivity and specificity values for the TBI sample in both the partial

invariance and strict invariance conditions. All values generated in Table 7.8 were calculated using the VassarStats calculator (Lowry, 1998-2012). Millsap and Kwok (2004) recommend reviewing the changes in sensitivity and specificity values between the two invariance conditions.

If the absolute value for the TBI sample in the partial invariance condition is less than the lower bound of the 95% CI in the strict invariance condition, then the non-invariant items from that factor fail the tests of no practical impact. One exception to this rule is if the sensitivity or specificity value in the partial invariance condition is considered acceptable (Millsap & Kwok, 2004). In these cases the decision would be no practical impact.

Table 7.7

Observed score, factor score and percentile rank cut points used to calculated sensitivity and specificity values for invariance conditions in each factor.

		Partial Invariance		Strict Invariance	
	Percentile rank	Observed Score	Factor Score	Observed Score	Factor Score
Factor-1	97.8	10	1.530	10	1.484
Factor-3	89.6	6	1.368	6	1.364
Factor-4	89.6	5	1.269	5	1.163
Factor-5	92.3	2	1.498	2	1.353

In none of Factor-1, Factor-3, Factor-4 or Factor-5 was the sensitivity values for the partial invariance conditions less than the lower bound 95% CI for the strict invariance conditions. In Factor-1 and Factor-3 the specificity value for the partial invariance condition was greater than the lower bound 95% CI for the strict invariance condition. These findings suggest for Factor-1 and Factor-3 there was no significant impact on the sensitivity or specificity values from the failure to find strict invariance. In Factor-5 the specificity value in the partial invariance condition was less than the lower bound 95% CI for the strict invariance condition. However, the absolute value of 0.901 is strong. This strong specificity value supported concluding the items from Factor-5 pass the tests of no practical impact.

Therefore retaining the all items from Factor-1, Factor-2 and Factor-5 and assuming full invariance does not compromise criterion related validity.

The specificity value in Factor-4 did suggest an important practical impact from the non-invariant items. The specificity value in the partial invariance condition is less than the lower bound of the 95% CI in the strict invariance condition. In addition to the significant reduction in the specificity value (a reduction of .737) the absolute value is low.

Therefore the practical impact analysis supported retaining all items from Factor-1, Factor-3 and Factor-5. The measurement invariance testing found all items from Factor-2 met the criteria of strict invariance which supported retaining all items from Factor-2. The items from Factor-4 failed the tests of no practical impact. This finding suggested the non-invariant items from Factor-4 (17, 92, 190, 278, 281, 291 and 303) have an important practical impact on the utility of the MMPI-2 with persons suffering a TBI. Only the items from Factor-4 (12, 21, 22, 46, 256, 273, 277, 279 and 280) that passed the test of strict invariance can be confidently retained. Tables 7.9 and 7.10 show the significant reduction in the sensitivity value, as represented by the increase in false-positive cases seen in the partial invariance condition compared with the strict invariance condition. False-positive cases are those shown in column 'Condition Absent' and row 'Test Positive' for both tables. Under strict invariance there are zero false-positive cases. While under partial invariance there are 59 false-positive cases. In contrast, support for retaining all items was the significant correlation between factor scores in the partial invariance condition and the factor scores in the strict invariance condition observed on Factor-1 (Spearman's rho = .999, $p < .01$), Factor-3 (Spearman's rho = .996, $p < .01$), Factor-4 (Spearman's rho = .970, $p < .01$) and Factor-5 (Spearman's rho = .951, $p < .01$).

7.4 Comparison of threshold parameters

Following the procedure outlined in the results section for Hs1 (see Section 5.4), the threshold values were compared for the items that failed the test of strict invariance. Borbsoom (2006) proposes that when the sources of item variance between the samples cancel each other out the result is no impact from the failure of strict invariance. If a lower threshold was observed in the TBI group, this reflects an increased likelihood that an item

with be endorsed for psychopathology, and vice versa. A comparison of the threshold parameters can help determine whether the number of lower and higher thresholds for the TBI group cancel each other out. Or whether there is a bias in the direction of higher or lower threshold parameters in the TBI group.

Table 7.11 shows the threshold values for the TBI and the Community samples, for the items that failed the test of strict invariance. Comparisons between the two threshold scores were completed by using the geometric mean of the standard error for each sample (Howell, 2002). As can be seen in Table 7.11, items 34, 177, 242 and 287 are equally likely to be endorsed by either sample. In total six items are observed to have higher thresholds in TBI sample, namely 16, 38, 170, 274, 290 and 299. While 18 items are observed to have lower thresholds in the TBI sample, namely 17, 42, 92, 138, 182, 190, 210, 234, 252, 278, 281, 291, 303, 322, 323, 329, 332 and 355. Therefore 18 of the 28 items compared suggesting an increase in likelihood for endorsement by the TBI sample. This review suggested a bias towards inflated scale scores in the TBI sample.

Items 17, 92, 190, 278, 281, 291 and 303 failed both the test of strict invariance and the tests for no practical impact. Therefore it is worth considering the analysis of threshold value for these items in particular. All seven items showed lower thresholds in the TBI group, suggesting that the items are more likely to be endorsed by the TBI sample. Again, this review suggested there is a bias to inflated scores in the TBI sample on Sc8.

7.5 Results of Gass (1991) correction procedure analysis

Items 106, 147, 165, 170, 180, 295 and 325 are part of the Gass (1991) correction procedure. Items 106, 147, 165, 180, 295 and 325 met the criteria of strict invariance. Only item 170 from the Gass (1991) correction procedure was identified as a specific source of measurement model variance on the Sc8 scale. The practical impact analysis supported retaining item 170. Table 7.8 shows the factor to which item 170 is allocated (Factor-1) passed the tests of no practical impact. The review of item thresholds (see Table 7.11) observed item 170 as being less likely to be endorsed by the TBI sample. The results fail to support the contention that use of these items with a person suffering a TBI would falsely inflate an MMPI-2 profile.

Table 7.8

Sc8 sensitivity and specificity analysis (with 95% confidence intervals: CI) for factors across invariance conditions in the TBI sample.

	Factor 1		Factor 3		Factor 4		Factor 5	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Partial Invariance	.991	.706	.985	.500	1.000	.263	.985	.901
	.963 – .998	.440 – .886	.953 – 0.996	.322 – 0.678	.969 – 1.000	.173 – .375	.909 – .999	.842 – .941
Strict Invariance	.960	1.000	.967	.706	.933	1.000	.717	.991
	.923 – .980	.463 – 1.000	.930 – 0.985	.440 – 0.886	.889 – .961	.517 – 1.000	.623 – .796	.946 – .999

Note – values were calculated using the VassarStats calculator (Lowry, 1998-2012)

Table 7.9

Quadrant frequencies for the TBI sample in Sc8 Factor-4 in the partial invariance condition

	Condition Present	Condition Absent
Test Positive	149	59
Test Negative	0	21

Table 7.10

Quadrant frequencies for the TBI sample in Sc8 Factor-4 in the strict invariance condition

	Condition Present	Condition Absent
Test Positive	208	0
Test Negative	16	5

Table 7.11

Mean threshold value and standard errors (SE) for the Community and the TBI samples for items freed when defining the Sc8 partial invariance model

	Community Thresholds		TBI Thresholds			
Item	Mean	SE	Mean	SE	Geometric Mean of SE	TBI compared with Community
16	0.215	0.038	1.987	0.313	0.223	>
17	2.372	0.156	0.691	0.211	0.186	<
34	0.930	0.038	0.888	0.239	0.171	=
38	0.770	0.042	3.088	0.466	0.331	>
42	2.080	0.086	0.593	0.336	0.245	<
92	2.020	0.100	0.143	0.190	0.152	<
138	2.372	0.112	0.347	0.298	0.225	<
170	1.760	0.086	2.923	0.487	0.350	>
177	1.376	0.045	1.560	0.316	0.226	=
182	2.012	0.086	0.791	0.317	0.232	<
190	2.029	0.105	0.152	0.189	0.153	<
210	1.941	0.066	1.068	0.374	0.269	<
234	2.293	0.115	0.023	0.348	0.259	<
242	0.453	0.035	0.543	0.203	0.146	=
252	2.584	0.196	1.313	0.147	0.173	<
274	0.892	0.046	2.277	0.404	0.288	>
278	1.166	0.044	0.310	0.156	0.115	<
281	2.077	0.103	0.355	0.157	0.133	<
287	0.645	0.034	0.711	0.217	0.155	=
290	0.127	0.034	1.707	0.269	0.192	>
291	2.049	0.098	1.461	0.206	0.161	<
299	1.379	0.063	3.492	0.622	0.442	>
303	2.640	0.175	0.122	0.251	0.216	<
322	1.957	0.077	0.401	0.424	0.305	<
323	1.994	0.094	0.554	0.291	0.216	<
329	2.827	0.246	0.936	0.356	0.306	<
332	2.063	0.105	0.887	0.262	0.200	<
355	2.654	0.146	0.660	0.357	0.273	<

Table 7.12

Standardized factor loadings for the five factor CFA model of MMPI-2 Sc8 in traumatic brain injury (TBI, n=229) and Community (Norm, n=1726) samples.

Parameter estimates (PE) and standard errors (SE) are derived from the partially invariant measurement model (Test 3 in Table 7.6). Also shown is the variance explained by the model in each item (R^2)

Item	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5		R^2	
	TBI	Norm	TBI	Norm	TBI	Norm	TBI	Norm	TBI	Norm	TBI	Norm
145	.719(.034)	.617(.027)									.483(.049)	.619(.033)
23	.568(.037)	.464(.030)									.677(.043)	.785(.028)
38	.706(.055)	.551(.031)									.501(.078)	.697(.034)
42	.551(.081)	.437(.058)									.696(.089)	.809(.050)
48	.626(.036)	.520(.030)									.608(.045)	.730(.031)
138	.310(.096)	.467(.063)									.904(.060)	.782(.059)
168	.593(.037)	.487(.031)									.649(.044)	.762(.030)
170	.728(.054)	.619(.041)									.469(.078)	.617(.051)
177	.595(.065)	.231(.047)									.646(.078)	.947(.022)
180	.762(.031)	.666(.030)									.419(.047)	.556(.041)
182	.574(.070)	.446(.060)									.670(.080)	.801(.053)
210	.107(.134)	.134(.076)									.989(.029)	.982(.020)
229	.597(.038)	.492(.032)									.643(.045)	.758(.032)
234	.439(.096)	.469(.070)									.807(.084)	.780(.065)
268	.603(.036)	.498(.029)									.636(.044)	.752(.029)
274	.661(.059)	.585(.030)									.563(.077)	.658(.036)
292	.448(.040)	.355(.027)									.800(.035)	.874(.019)
295	.380(.040)	.298(.029)									.855(.030)	.911(.017)
296	.574(.038)	.469(.030)									.670(.044)	.780(.028)

298	.606(.042)	.500(.032)							.633(.050)	.750(.032)
299	.737(.061)	.631(.033)							.457(.089)	.601(.041)
307	.595(.040)	.490(.033)							.646(.048)	.760(.033)
311	.769(.029)	.674(.027)							.408(.045)	.545(.036)
319	.740(.035)	.640(.034)							.453(.051)	.590(.043)
322	.476(.128)	.354(.065)							.774(.122)	.875(.046)
329	.629(.074)	.654(.075)							.605(.093)	.573(.098)
333	.718(.037)	.616(.034)							.484(.053)	.620(.041)
355	.382(.114)	.492(.072)							.854(.087)	.758(.070)
276			.521(.143)	.633(.050)					.729(.149)	.599(.063)
6			.374(.114)	.476(.045)					.860(.085)	.774(.043)
90			.362(.111)	.462(.044)					.869(.080)	.786(.040)
192			.517(.140)	.630(.053)					.732(.145)	.603(.066)
343			.314(.102)	.406(.043)					.901(.064)	.835(.035)
316					.826(.029)	.761(0.026)			.317(.049)	.420(.039)
16					.719(.053)	.578(0.029)			.482(.076)	.665(.034)
32					.626(.041)	.541(0.031)			.608(.051)	.708(.033)
34					.095(.121)	.302(0.040)			.991(.023)	.909(.024)
35					.331(.043)	.270(0.031)			.890(.028)	.927(.017)
85					.707(.038)	.625(0.029)			.499(.054)	.609(.036)
106					.436(.045)	.361(0.034)			.810(.039)	.869(.025)
221					.704(.037)	.622(0.027)			.505(.053)	.614(.034)
242					.073(.100)	.437(0.034)			.995(.015)	.809(.029)
287					.341(.101)	.236(0.039)			.884(.069)	.944(.018)
290					-.509(.075)	-.430(0.034)			.741(.077)	.815(.029)
320					.606(.047)	.520(0.035)			.633(.057)	.729(.036)
323					.642(.108)	.564(0.049)			.588(.138)	.682(.055)

332			.473(.120)	.553(0.056)					.776(.114)	.694(.062)
277					.844(.025)	.747(.024)			.287(.042)	.442(.036)
12					.385(.047)	.285(.033)			.852(.036)	.919(.019)
17					.716(.053)	.726(.042)			.488(.077)	.472(.060)
21					.660(.038)	.531(.029)			.564(.050)	.718(.031)
22					.851(.025)	.756(.030)			.276(.043)	.429(.045)
46					.562(.043)	.436(.032)			.684(.048)	.810(.028)
92					.663(.065)	.559(.053)			.561(.087)	.688(.059)
190					.673(.072)	.599(.050)			.547(.097)	.641(.059)
256					.589(.042)	.461(.030)			.654(.049)	.788(.027)
273					.788(.029)	.674(.028)			.379(.046)	.545(.038)
278					.464(.085)	.372(.042)			.785(.079)	.862(.031)
279					.381(.045)	.282(.031)			.854(.035)	.920(.017)
280					.530(.045)	.407(.033)			.719(.048)	.835(.027)
281					.488(.080)	.583(.050)			.762(.078)	.660(.058)
291					.360(.115)	.478(.064)			.870(.083)	.772(.061)
303					.827(.049)	.652(.057)			.316(.081)	.576(.074)
325							.858(.034)	.770(.034)	.264(.058)	.407(.053)
147							.748(.046)	.632(.039)	.440(.068)	.601(.049)
165							.631(.059)	.507(.046)	.602(.074)	.743(.047)
233							.748(.045)	.632(.036)	.441(.067)	.601(.045)
252							.472(.108)	.463(.117)	.777(.102)	.786(.109)

Chapter Eight – Discussion

To quickly reiterate the primary controversy is that misinterpretation of neurological symptoms in individuals with TBI may generate spuriously inflated scores on diagnostic measures (Nelson et al., 1989). In line with this concern, the Gass (1991) correction procedure suggests removing 14 items from the MMPI-2 when generating a profile for a person suffering a TBI. There is continued use of this procedure despite the inconsistent results from validation studies (Arbisi & Ben-Porath, 1999; Brulot et al., 1997; Edwards et al., 2003; Gass & Wald, 1997; Glassmire et al., 2003; Rayls et al., 1997). Measurement invariance analysis and practical impact analysis, using the Millsap and Kwok (2004) procedure, were complete on Hs1, Hy3 and Sc8. The discussion section will commence with an interpretation of the results and their implications for each clinical scale separately.

In all three clinical scales the defined factor-model failed the test of strict invariance but was found to meet the criteria for partial invariance. Items that failed both the test of strict invariance and the tests of no practical impact are not supported for retention when assessing a person suffering a TBI. In Hs1 the practical impact analysis supported retaining all items from these scales in assessment of a person who has suffered a TBI. In Hy3, two items (161 and 185) failed the test of strict invariance and failed the tests of no practical impact. In Sc8, seven items (17, 92. 190, 278, 281, 291 and 303) failed both the test of strict invariance and the tests of no practical impact. While none of these nine items were included in the Gass (1991) correction procedure, the implications from these findings along with potential areas for future research are considered.

Next will be the General Discussion. In this section the integration of the findings across all three scales is interpreted. It will be shown how the analyses do not support removal of the items suggested by the Gass (1991) correction procedure. Additionally the findings fail to support the underlying hypothesis that bias ensues from MMPI-2 items which contain neurological content. This finding is valuable as the neurological content hypothesis is central to correction procedures with other populations such as

spinal cord injury, multiple sclerosis and temporal lobe epilepsy. Furthermore the neurological content hypothesis is difficult to integrate with understanding the injury severity paradox as an outcome from differences in insight associated with differences in TBI severity. With the data failing to support the neurological content hypothesis the interpretive problem is avoided. This conclusion facilitates use of the Stress-Appraisal-Coping model to continue as the simplest explanation for the injury severity paradox. The support for the Stress-Appraisal-Coping model is discussed and the value of therapeutic information provided by this model is considered.

The benefit from using measurement invariance and practical impact analysis is reviewed. The backwards elimination approach which can define specific sources of variance is highlighted as strength of the study. Finally some limitations of the study and future directions are provided.

8.1 Discussion for Hs1

EFA and CFA showed a 4-Factor model best represented the 32 items from Hs1. Initially EFA and CFA were used to define three candidate models, one from each of the Norm A, Norm B and TBI samples. The replication procedure was unable to find a clearly preferred candidate model. In this step an all-female and an all-male sample were included to assess the generalisability of the candidate model across genders. Importantly in all five samples each candidate model met admissibility and model fit criteria. Review of the item R-squared values found the Norm B and the TBI candidate model were preferable over the Norm A candidate model. Comparisons of the factor rho coefficient for each candidate model found the same preference for the Norm B and TBI candidate models. A theoretical review of the item to factor relationships defined in each candidate model found a preference for the TBI model. Therefore the TBI candidate model was selected as the preferred factor model. The item to factor structure for this baseline model, named Hs1 4-Factor model is provided in Table 5.7. The factors were labelled; ‘Sleep quality/ energy levels’, ‘Orofacial symptomology’, ‘General health’ and ‘Gastrointestinal complaints’.

Notwithstanding the value in replication studies, confidence in the Hs1 4-Factor model

is increased as the model selected was tested across a variety of samples. Additionally the non-significant χ^2 result for the factor model in the TBI sample is an unusual finding and a strong criterion for preferring this model. Large samples can be a double-edged sword when it comes to conducting CFA as the power to detect small differences in calculated fit indices can lead to rejection of acceptable models (Borsboom, 2006; Brown, 2006). However, adequate sample size is also a necessary precondition to obtain confident conclusions from any CFA procedure. The successful replication of the Hs1 4-Factor model across diverse samples that range between 242 and 1462 participants improves the confidence in the conclusions. As shown in Table 5.4 the Hs1 4-Factor model (the TBI candidate model in the Table) produced good measures of fit in a variety of samples. The consistency of these findings provides strong empirical support for the factor model proposed.

The next step was to complete measurement invariance testing across the TBI and a gender-matched MMPI-2 normative sample (the Community sample). Table 5.8 shows that the measurement invariance analysis found the Hs1 4-Factor model failed to meet the requirements of strict invariance. Using the backwards elimination procedure the requirements for partial invariance were met when loadings and thresholds were freed on items 3, 10, 39 and 47. The finding of partial invariance required further investigation to determine the clinical impact from finding non-invariant items. Therefore the practical impact analysis as described by Millsap and Kwok (2004) was completed (see Section 4.5 and Table 5.10).

Practical impact analysis supported retaining all Hs1 items when assessing patients who had suffered a TBI. All the items from Factor-1 and Factor-2 met the criteria of strict invariance. Therefore the items allocated to these factors are appropriate for assessing psychopathology in a person suffering a TBI. Factor-3 and Factor-4 from the model of Hs1 were observed to contain items that failed the test of strict invariance. Therefore these two factors required evaluation through the practical impact analysis. Table 5.10 shows the sensitivity and specificity values for Factor-3 and Factor-4 are not significantly reduced in the partial invariance condition when compared with the strict invariance condition. Therefore the practical impact analysis supported continuing to use all items from Factor-3 and Factor-4 when using the MMPI-2 with a person

suffering a TBI. The combination of measurement invariance testing and practical impact analysis supported retaining all items from Hs1 when assessing patients suffering a TBI.

The thresholds of the four items (3, 10, 39 and 47) that failed the test of strict invariance were reviewed. If there was a consistent bias towards increased likelihood for endorsement by the TBI sample (i.e. lower threshold values), this would suggest potential inflation of scale scores from the failure of strict invariance. Table 5.11 shows that items 10 and 39 have no significant difference in threshold values. Item 3 is identified as more likely to be endorsed by the TBI sample with item 47 less likely to be endorsed by the TBI sample.

Borbsoom (2006) explains that potentially failure to find strict invariance may have no impact when scale scores are calculated. He clarifies that when the sources of item variance between the samples cancel each other out the result is no impact from the failure of strict invariance. The review of the threshold values identified one item which increased probability of endorsement and one item which decreased the probability for endorsement in the TBI sample. It is reasonable to expect that one increase and one decrease in likelihood of item endorsement would cancel each other out when total scores are calculated for Hs1. The result would be no impact from the failure of strict invariance. This conclusion is further supported by finding no practical impact for the items that failed the test of strict invariance.

The correction procedure proposed by Gass (1991) includes items 101, 149, 175, 179 and 247 from Hs1. If the hypothesis that these items bias MMPI-2 profile interpretation were correct, it would be expected that these items specifically would be the source of any failure of invariance between a non-TBI and a TBI sample. All five of the items included in the Gass (1991) correction procedure met the requirements of strict invariance. This finding supports concluding that that removal of these five items from an MMPI-2 assessment is unwarranted.

These findings concur with previous research that determined the Gass (1991) correction procedure (1991) is inappropriate for clinical diagnosis (Arbisi & Ben-

Porath, 1999; Brulot et al., 1997; Edwards et al., 2003). However, none of the previous research directly evaluated item level effects using a method such as measurement invariance, which is a specific strength of this study. Additionally, with dichotomous data, the results of the measurement invariance testing with theta parameterisation are equivalent to a test of differential item functioning under item response theory (Glöckner-Rist & Hoijtink, 2003). Measurement invariance testing or evaluation of differential item functioning is recommended to determine if a test instrument is appropriate in a different population from those employed during the instruments design (Muniz & Bartram, 2007). Measurement invariance testing to assess the Gass (1991) correction procedure follows this recommendation. No previous research into the Gass (1991) correction procedure has employed measurement invariance testing. The results from the current study suggest that with the TBI population, items in the MMPI-2 that may refer to neurological symptoms associated with the injury do not diminish the validity of an assessment on Hs1 scale. This conclusion is similar to Glassmire et al. (2003) who, despite not using the suggested analytic method, argued that endorsement of the Gass (1991) correction items can be also explained by a psychological illness.

There is further evidence from the analysis on Hs1 that removing the Gass (1991) correction items is unnecessary. The correction items are split between two factors, with items 101 and 149 forming part of the 'Orofacial symptomology' factor, while items 175, 179 and 247 load onto the 'General health' factor. The factor analysis suggests the items identified by Gass that are included in Hs1 do not form a single factor representing the neurological symptoms of a TBI. This is noteworthy as the Hs1 4-Factor model was originally defined using a TBI sample. If the neurologic items were spuriously inflating scale scores it might be expected that they produced a separate factor from the remaining influences on the measurement model. The failure to find a single factor of neurologic items further supports concluding that the items do not unduly influence the scoring of Hs1.

In summary the 32 items in Hs1 were found to be best represented with a 4-Factor model. The tests of measurement invariance were conducted by comparing the endorsement patterns in a sample of patients suffering a TBI and endorsement patterns of a gender matched subset of the MMPI-2 normative sample. The Hs1 4-Factor model

met the criterion for partial invariance. Items 3, 10, 39 and 47 were identified as failing the test of strict invariance. All items that failed the test of strict invariance passed the tests for no practical impact. Comparison of the threshold parameters across groups for the items that failed strict invariance found no evidence to support a bias towards inflated scores for persons who have suffered a TBI. All items from the Gass (1991) correction procedure met the criteria of strict invariance. These findings support retaining all Hs1 items when assessing patients suffering a TBI.

8.2 Discussion Hy3

Twenty items that are also part of Hs1 were removed from the analysis of Hy3. It was necessary to define a baseline model with the remaining 40 items for the measurement invariance analysis. EFA and CFA were used to define three candidate models from the Norm A, Norm B and TBI samples. The candidate models were compared in a replication procedure. During this stage an all female and an all male sample were included to assess the generalisability of the candidate model across genders. The Norm A candidate model was excluded as a potential factor model as it failed the admissibility criteria in the Norm B and TBI samples. The replication procedure established a marginal preference for the Norm B candidate model over the TBI candidate model. However as can be seen in Table 6.4 both candidate models generated acceptable fit statistics across a variety of samples. Therefore it was decided that further comparisons were warranted.

Review of the item R-squared values found no difference between the Norm B and the TBI candidate models. Review of the factor rho coefficient for each candidate model found a subjective preference for the Norm B candidate model. A theoretical review found a preference for the Norm B candidate model. The Norm B candidate model was preferred over the TBI candidate model during the replication procedure, a comparison of construct reliability and a theoretical review. Therefore the Norm B candidate model was selected as preferred factor model. The stringent procedure followed to define and select the Hy3 4-Factor model as representative of the 40 items provides confidence in the model chosen. Table 6.7 provides details of the item to factor structure for the Hy3 4-Factor model. Hy3 factors were labelled ‘Mood’, ‘Beliefs about others’, ‘Reactions to

others' and 'Shyness'.

The next step was to complete measurement invariance testing across the TBI and a gender-matched MMPI-2 normative sample (the Community sample). Table 6.8 shows that the measurement invariance analysis found Hy3 4-Factor model failed to meet the requirements of strict invariance. The Hy3 4-Factor model was able to meet the requirement for partial invariance when loadings and thresholds for 20 items were freed using the backwards elimination procedure. The items with parameters freed sequentially were 44, 193, 172, 29, 11, 253, 40, 263, 148, 241, 31, 14, 115, 124, 161, 185, 230, 129, 76 and 135. The finding of partial invariance required further investigation to determine the clinical impact from finding non-invariant items. Therefore the practical impact analysis as described by Millsap and Kwok (2004) was completed (see Table 6.10).

Practical impact analysis was completed on each of the four factors as they all contained at least one item that failed the test of strict invariance. Comparisons of the sensitivity and specificity values for the TBI sample in the partial and strict invariance conditions found one indication of a substantial practical impact. The sensitivity value for Factor-4 was significantly less in the partial invariance condition when compared with the strict invariance condition. The result is that items 161 and 185 which failed the test of strict invariance and failed the tests of no practical impact are not supported for retention. The practical impact analysis supported retaining the remaining 38 items from Hy3.

Borbsoom (2006) suggests that when the sources of item variance between the samples cancel each other there is no impact on scale scores from the failure of strict invariance. The review of the threshold values for the items that failed the test for strict in variance in Hy3 is shown in Table 6.13. As can be seen 11 items are more likely to be endorsed by the TBI sample with eight items less likely to be endorsed. Therefore the difference in threshold values across the groups does not uniformly increase the likelihood a person would endorse an item. The mixture of increases and decreases of item endorsement likelihood by the TBI sample reduces concerns for undue inflation of scale scores.

Items 161 and 185 failed both the test of strict invariance and the tests for no practical impact. Therefore it is worth considering the analysis of threshold value for these items in particular. Item 185 is equally likely to be endorsed by each sample suggesting there is no concern with endorsement of this item falsely inflating a scale score. Item 161 is less likely to be endorsed by the TBI sample, which again suggests there is no concern with endorsement of this item falsely inflating scale scores.

Twenty items from the Hy3 scale were excluded from the analysis as these items are also part of Hs1. The analysis of Hs1 concluded all items, including the 20 duplicate items, were appropriate for assessment with persons suffering a TBI. Therefore including the analysis for Hs1 the results find only two items from the 60 items on Hy3 failed both the tests of strict invariance and of no practical impact. In other words 58 of the 60 items from the Hy3 scale are suitable for assessment of persons with a TBI. However the question remains what to do with items 161 and 185?

Items 161 and 185 have been identified in the tests for no practical impact as potentially unsuitable due to the significant reduction in the sensitivity value found in the partial invariance condition. The tests of no practical impact assess the capacity of the observed scores to correctly classify the factor scores. The factor scores can be interpreted to measure underlying psychopathology, and observed scores measures the assessment of psychopathology. Therefore the interpretation of the findings is that items 161 and 185 are diminished in their capacity to correctly identify the psychopathology when it exists in a person who has suffered a TBI. Evidence of this is seen in Tables 6.11 & 6.12, which show an increase in false-negative diagnosis from zero in the strict invariance condition to 40 in the partial invariance condition. Failure to identify existing psychopathology potentially denies a TBI sufferer the opportunity to benefit from therapy. This in turn can negatively impact the rehabilitation process. Interestingly, failure to identify psychopathology could occur when a test has a bias to reducing scores in persons with a TBI, which is contrary to the concern that neurological content in items will falsely inflate MMPI-2 scores.

One option in response to these findings is do nothing and maintain using the all Hy3 items with patients suffering a TBI. Only two items (161 and 185) from the 60 items in

Hy3 failed the tests of strict invariance and no practical impact. When reviewing the threshold values for these two items in the TBI sample and the Community sample, only item 161 was significantly different. The TBI sample was less likely to endorse item 161. A reduced likelihood for endorsement could lead to a bias for lower scores. The reduction in the sensitivity value for the factor related to item 161 observed in the tests for no practical impact also suggests a bias for lower scores. As a result any person with a TBI who generates scores of clinical importance with Hy3 has done so despite a bias in reducing scores for persons with a TBI. This should provide confidence in any assessment that identifies psychopathology using Hy3.

From a statistical perspective it could be argued that with four factors being subjected to the post hoc practical impact analyses a Bonferroni adjustment to critical p value is necessary. Applying the .05/4 adjustment would drop critical p from .05 to .0125. This would in turn alter the confidence intervals around the comparative sensitivity or specificity values to be closer to 99% (rather than the 95% CI employed in the analyses). Under this assumption the 99% CI for sensitivity in the strict invariance condition is .376 to 1.000. With the revised confidence interval the sensitivity value in the partial invariance condition remains below the lower bound value in the strict invariance condition. Therefore even if a Bonferroni adjustment were applied, items 161 and 185 would still fail the tests of no practical impact.

At a practical level the risk from degraded sensitivity is that assessment will overlook true cases of psychopathology. There are multiple options available to remediate the psychometrically underperforming scale. The scale could have the two items removed and re-scored with 58 questions. However, if one were to follow this approach it would be a prerequisite to also have the *T*-scores adjusted to reflect the altered scale length. Additionally removal of items fundamentally alters the underlying measurement model. In this situation Millsap and Kwok (2004) recommend repeating the factor model definition, measurement invariance testing and practical impact analysis completed in this study with the remaining items. It would be necessary to repeat this process until the analysis supported retaining all the items in the scale.

On Hy3, raw scores (or numbers of items endorsed) of 19 for males and 20 for females

are clinically important as these scores represent a *T*-score of 65 which is used to indicate clinical significance. Using these scores as a guide a clinician could review the responses for a TBI patient if their profile score fell just below the suggested values. When a profile has a raw score of 17-18 for males or 18-19 for females then the answers to questions 161 and 185 can be reviewed. That is because if these items have been scored as 0 then if they were scored as 1 the person would generate a *T*-score of 65 or more. If the review finds that different responses to these two items would result in a profile that reflects psychopathology, then a prudent method would be to approach treatment or rehabilitation assuming the psychopathology exists. The risk from this approach is the potential for a false positive diagnosis which could lead to incorrect medication and associated damaging consequences (Spitzer & Frances, 2011). However, if rehabilitation focused on psychotherapy while simultaneously gathering additional diagnostic information these risks may be safely mitigated.

The practical impact analysis fails to support the hypothesis that the neurological and physical symptoms associated with a TBI falsely inflate the Hy3 scale. Two of the 40 items analysed in this scale form part of the Gass (1991) correction procedure, namely, items 31 and 172. When defining the partial invariance model both items required the respective loadings and thresholds to be freed. At first, this failure of strict invariance at the item level supports the contention by Gass that items 31 and 172 measure psychopathology differently in persons with a TBI. The review of item thresholds observed item 31 as being more likely to be endorsed by the TBI sample, while item 172 was less likely to be endorsed, thus cancelling each other out. The practical impact analysis found support for retaining all items from Factor-1 (including item 31) and Factor-2 (including item 172). These observations fail to support the contention that the items identified by the Gass (1991) correction procedure lead to falsely inflated clinical scale score in the TBI population.

The contents of the two items that failed both the test of strict invariance and of no practical impact are not neurological symptoms; 161 (hides shyness) and 185 (bothered by shyness). Furthermore, as discussed above reduced sensitivity is related to a bias for lower scale scores in the TBI sample. Whereas the neurologic hypothesis is concerned that a TBI will lead to falsely inflated scale scores. Therefore while the analysis

identified items 161 and 185 as failing the tests of strict invariance and no practical impact, neither of the content of, or the impact on the MMPI-2 reflect the concerns noted in the neurologic content bias hypothesis.

Previous research into the Gass (1991) correction procedure found some support for the proposal to remove 14 items when completing an MMPI-2 assessment with patients suffering a TBI (Gass & Wald, 1997; Rayls et al., 1997). Other papers reviewing the Gass (1991) correction procedure concluded that the removal of items when assessing persons with a TBI is inappropriate (Arbisi & Ben-Porath, 1999; Brulot et al., 1997; Edwards et al., 2003). A limitation of the previous research is that no previous study applied the direct evaluation method of measurement invariance the recommended by Muniz and Bartram (2007) to investigate as the appropriateness of tests in various populations. The studies conducted in this thesis have established that removing the items identified by Gass (1991) from the Hy3 scale when assessing patients with a TBI is unwarranted. The findings are not a rejection of the suggestion that MMPI-2 items may refer to neurological symptoms associated with a TBI. However, the findings do not support the proposition that the content of items on the Hs1 and Hy3 scale falsely inflate scale scores in persons with a TBI.

In summary 20 items from Hy3 were excluded from analysis as these items were previously assessed with Hy1. The analysis found a 4-Factor model best represented the remaining 40 items from Hy3. Tests of measurement invariance were completed by comparing the endorsement patterns in a sample of patients suffering a TBI and endorsement patterns of a gender matched subset of the MMPI-2 normative sample. The Hy3 4-Factor model met the criteria of partial invariance. Items 11, 14, 29, 31, 40, 44, 76, 115, 124, 129, 135, 148, 161, 172, 185, 193, 230, 241, 253 and 263 were identified as failing the test of strict invariance. A practical impact analysis supported retaining 18 of the 20 items found to fail the test of strict invariance. Items 161 and 185 failed both the test of strict invariance and the tests of no practical impact. Comparison of the threshold parameters across groups for the items that failed strict invariance found no evidence to support a bias towards inflated scores for persons who have suffered a TBI. Importantly neither the content of items 161 and 185, nor the impact on the MMPI-2 assessment reflects the concerns outlined by Gass (1991). Two items from the Gass

(1991) correction procedure (31 and 172) failed the test of strict invariance. However the factors which comprised both these items met the criteria for no practical impact which supports their continued use when assessing persons with a TBI.

8.3 Discussion Sc8

Ten items from Sc8 are also part of either Hs1 or Hy3. The four items from Hs1 (91, 179, 247 and 255) and six items from Hy3 (9, 31, 44, 65, 166 and 218) were removed from the analysis of Sc8. It was necessary to define a baseline model with the remaining 68 items to facilitate the measurement invariance analysis. EFA and CFA were used to define three candidate models from the Norm A, Norm B and TBI samples. The candidate models were compared in a replication procedure. During this stage an all female and an all male sample was included to assess the generalisability of the candidate model across genders. The Norm A candidate model was excluded as a potential factor model as it failed the admissibility criteria in the all female and TBI samples. The replication procedure found a clear preference for the Norm B candidate model over the TBI candidate model. In the Norm A and Male Norm samples the TBI candidate model generated fit indices below the threshold defined to meet acceptance criteria. Furthermore the Norm B candidate model generated indices reflecting a better fitting model in the TBI sample. This finding was observed despite the TBI candidate model having the benefit of being defined with this sample. Therefore the Norm B candidate model was selected as the preferred factor model. In summary EFA and CFA found a 5-Factor model best represented the 68 items from Sc8 (see Table 7.5 for details of the item to factor structure). Sc8 factors were labelled: 'Unusual experiences and fearful thoughts'; 'Beliefs about family'; 'Unusual thoughts'; 'Beliefs about self and others'; and, 'Cognitive symptoms'. The rigorous procedure employed to define and select the Sc8 5-Factor model as representative of the 68 items, including replication across samples, increases confidence in the chosen model.

The next step was to complete measurement invariance testing across the TBI and a gender-matched MMPI-2 normative sample (the Community sample). Table 7.6 shows the Sc8 5-Factor model failed the test of strict invariance. Freeing item threshold and loading parameters following the backwards elimination procedure was successful in

defining a model that met the requirements of partial invariance. The partial invariance model was defined with loadings and threshold sequentially freed for items 252, 210, 299, 355, 38, 274, 303, 290, 138, 322, 234, 170, 323, 332, 177, 242, 42, 182, 291, 92, 190, 281, 16, 287, 329, 278, 34 and 17. The finding of partial invariance required further investigation to determine the clinical impact from finding non-invariant items. Therefore the practical impact analysis as described by Millsap and Kwok (2004) was completed (see Table 7.8).

Practical impact analysis was completed on four of the five factors as they contained at least one item that failed the test of strict invariance. Factor-2 was not subject to practical impact analysis as all the items on this factor meet the criteria of strict invariance when defining the partial invariance model. To complete the practical impact analysis the sensitivity and specificity values in the TBI sample were compared across the partial invariance condition and the strict invariance condition. The sensitivity values in Factor-1, Factor-3, Factor-4 and Factor-5 met the requirement of no practical impact.

The specificity value in Factor-4 and Factor-5 was significantly lower in the partial invariance condition compared with the strict invariance condition. Millsap and Kwok (2004) recommend that a change in sensitivity or specificity be acceptable if the absolute value in the partial invariance condition is strong. The specificity value of .901 in the partial invariance condition on Factor-5 is strong enough to meet the requirement for no practical impact. The specificity value of .263 in the partial invariance condition for Factor-4 is too low to meet the requirement for no practical impact. The result is that items 17, 92, 190, 278, 281, 291 and 303 failed the test of strict invariance, and failed the tests of no practical impact. These seven items are not supported for retention. The practical impact analysis and measurement invariance testing supported retaining the remaining 61 items from Sc8.

Borbsoom (2006) suggests that there may be no impact from the failure of strict invariance when the sources of item variance between the samples cancel out each other. The review of the threshold values for the items that failed the test for strict in variance in Sc8 is shown in Table 7.11. The Table shows items 34, 177, 242 and 287 are

equally likely to be endorsed by either sample. In total six items are less to be endorsed by the TBI sample, they are items 16, 38, 170, 274, 290 and 299. While 18 items are more likely to be endorsed by the TBI sample, they are items 17, 42, 92, 138, 182, 190, 210, 234, 252, 278, 281, 291, 303, 322, 323, 329, 332 and 355. The considerably greater number of items suggesting an increase (n=18), compared with a decrease (n=6), in likelihood for endorsement by the TBI sample raises concerns for undue inflation of scale scores in the TBI population.

The failure of items from Factor-4 to pass the tests of no practical impact requires consideration. The items of concern are 17, 92, 190, 278, 281, 291 and 303 as these items also failed the test of strict invariance. The conclusion is based solely on the significantly reduced specificity value in the partial invariance condition. The factor scores can be interpreted to measure underlying psychopathology, and observed scores measures the assessment of psychopathology. Therefore the interpretation of the findings is that items 17, 92, 190, 278, 281, 291 and 303 are diminished in their capacity to correctly identify the absence of psychopathology when none exists in a person who has suffered a TBI. The increase in the number of false-positive cases in the partial invariance condition compared with the strict invariance condition exemplifies this interpretation (see Table 7.9 and 7.10).

Millsap and Kwok (2004) argue that specificity is important when a false-positive may lead to a person being labelled with a diagnosis unnecessarily or accessing costly clinical resources. Therefore it is necessary to consider the impact from a false-positive diagnosis of psychopathology in a person who has suffered a TBI. An incorrect diagnosis of psychopathology may lead to the use of unnecessary medication with potentially harmful consequences (Spitzer & Frances, 2011). Alternatively a false-positive diagnosis may result in unnecessary psychotherapy. Compared with unnecessary medication this latter prospect, especially under a cognitive model, does not have the same risks for the individual receiving treatment.

The Gass (1991) correction procedure was designed to remove the risk for falsely inflated clinical scale scores in persons who have suffered a TBI. If scale scores were falsely inflated, then persons without psychopathology would be incorrectly diagnosed

with an illness. In other words the test would produce an excessive number of false-positive diagnoses. Table 7.9 and Table 7.10 show the large number of additional false-positive diagnoses in the partial invariance condition compared with the strict invariance condition for Sc8 Factor-4. At first inspection this result may be seen to support the Gass (1991) correction procedure. However, none of the Gass (1991) correction procedure items were allocated to Factor-4. Furthermore, none of the items from Factor-4 contain content referring to neurologic symptoms associated with a TBI. As such the practical impact analysis has not identified items which support the hypothesis that neurological symptoms associated with a TBI may falsely inflate MMPI-2 scores.

The items from Factor-4 consistently reflect thoughts by a person about themself and others, the types of thoughts that could be the focus in cognitive-based therapies. A conservative approach to treatment for patients with a TBI would be to review the seven items from Factor-4. If these items are not contributing to an elevated score then treatment can proceed as per normal. Instead, if these items are specifically contributing to an elevated score then cognitive based therapy is recommended prior to a decision to include pharmacology during rehabilitation. During the course of therapy it is standard practice to refine or confirm an initial diagnosis, which can then inform as to the appropriateness for pharmacotherapy.

Alternatively the seven non-invariant items from Factor-4 may be removed from MMPI-2 assessment with patients suffering a TBI. However, as previously explained, removal of items from Sc8 would alter the measurement model for the scale. Therefore the remaining items would require a repeat of factor model definition, measurement invariance tests and practical impact analysis completed in this study (Millsap & Kwok, 2004). This process would need to be repeated until the analysis supported retaining all the items in the factor model. In addition it would be necessary to recalculate the raw score to *T*-score distribution for the altered Sc8 scale. Recalibration is necessary to maintain the integrity of interpretations for patients with a TBI based on Sc8 scale *T*-scores.

Finally, the full MMPI-2 could continue to be used with patients who have suffered a

TBI. Items 17, 92, 190, 278, 281, 291 and 303 failed both the test of strict invariance and the tests of no practical impact. Therefore it is worth considering the analysis of threshold value for these items in particular. All seven items are more likely to be endorsed by the TBI sample, which again suggests potential for endorsement of these items to falsely inflate scale scores. The inflation of scale scores increases the potential for false positives with the MMPI-2, which would reduce the specificity of the diagnostic instrument. An excessive reduction in the specificity of items from Factor-4 was identified as problematic in the post hoc practical impact analysis. All seven items (17, 92, 190, 278, 281, 291 and 303) from Factor-4 that were compared for threshold values across groups found a bias in the direction of increased likelihood for endorsement by the TBI group. Therefore continued use of these seven items may increase the potential for false positive diagnoses. Despite this outcome reflecting the primary concern that led to the development of the Gass (1991) correction procedure it is important to reiterate that none of the items refer to neurologic symptoms that commonly occur after a TBI.

Seven of the Gass (1991) correction procedure items were investigated in the analyses of Sc8. The results support retaining all seven items when completing a MMPI-2 assessment on a person with a TBI. Items 106, 147, 165, 180, 295 and 325 were found to meet the criteria of strict invariance which supports their retention. Item 170 failed the test of strict invariance. However as this item was allocated to Factor-1, the practical impact analysis supports its retention with patients suffering a TBI. Furthermore a review of the item thresholds found item 170 was less likely to be endorsed by the TBI sample. Potential bias towards inflating scale scores requires an item to be more likely endorsed by the TBI sample.

In summary 10 items from Sc8 were excluded from the analysis as 91, 179, 247 and 255 were part of Hs1, and 9, 31, 44, 65, 166 and 218 were part of Hy3. All ten items were supported for retention based on the investigation completed on Hs1 and Hy3. As for Sc8, the analysis found a 5-Factor model best represented the remaining 68 items. Tests of measurement invariance were completed by comparing the endorsement patterns in a sample of patients suffering a TBI and endorsement patterns of a gender matched subset of the MMPI-2 normative sample. The Sc8 5-Factor model met the criteria for partial

invariance. Items 16, 17, 34, 38, 42, 92, 138, 170, 177, 182, 190, 210, 234, 242 252, 274, 278, 281, 287, 290, 291, 299, 303, 322, 323, 329, 332 and 355 were identified as failing the test of strict invariance. Items 17, 92, 190, 278, 281, 291 and 303 also failed the tests of no practical impact. Comparison of the threshold parameters across groups for the items that failed strict invariance found evidence to support a bias towards inflated scores for persons who have suffered a TBI. However a review of the item content found no evidence to support concerns that items were endorsed due to referring to neurologic symptoms after a TBI. Including the analysis for Hs1 and Hy3, the results find seven items from the 78 items on Sc8 failed both the tests of strict invariance and of no practical impact. In other words 71 of the 78 items from the Sc8 scale are suitable for assessment of persons with a TBI.

Seven items from the Gass (1991) correction procedure were part of Sc8 analysis. Six of these items (106, 147, 165, 180, 295 and 325) met the criteria of strict invariance. Item 170 failed the test of strict invariance but passed the tests of no practical impact. These findings fail to support the continued application of the Gass (1991) correction procedure. These findings augment the earlier measurement invariance and practical impact analysis that rejected removing the items identified by the Gass (1991) correction procedure from Hs1 and Hy3.

8.4 General Discussion

Gass (1991) correction procedure

Deciding whether to use the uncorrected MMPI-2 or the Gass (1991) correction procedure is an important decision in rehabilitation for a patient with a TBI. Inappropriate use of the MMPI-2 increases the risk for misdiagnosis of psychopathology in TBI patients. Misdiagnosis can result in a person with psychopathology not receiving the treatment available, and associated gains from psychotherapy (Anson & Ponsford, 2006b, 2006c; Handel et al., 2007). The findings from this dissertation fail to support employing the MMPI-2 Gass (1991) correction procedure when assessing persons with a TBI.

The Gass (1991) correction procedure suggests removing 14 items from the MMPI-2

when completing an assessment with persons who have suffered a TBI. The 14 items for removal are proposed to artificially inflate profiles for persons suffering a TBI due to the neurologic content of the items. The fundamental rationale for a correction procedure is that the correction procedure removes items which reduce the clinical validity of the MMPI-2. Implied by the need for a correction procedure is that the 14 items identified by Gass (1991) reduce the validity of the MMPI-2 when used with patients who suffer a TBI. The findings from the measurement invariance and practical impact analysis completed on Hs1, Hy3 and Sc8 fail to support this hypothesis.

The 14 items from the Gass (1991) correction procedure are 31, 101, 106, 147, 149, 165, 170, 172, 175, 179, 180, 247, 295 and 325. Whilst none of the factor models were found to meet the criteria of strict invariance the benefit of using the backwards elimination procedure is that it can identify which items were specific sources of variance in the measurement model parameters. When defining the partial invariance model for Hs1 none of the five items (101, 149, 175, 179 and 247) from the Gass (1991) correction procedure required parameters to be freed. Therefore all five items met the requirement of strict invariance. When defining the partial invariance model for Hy3 the two items (31 and 172) from the Gass (1991) correction procedure both required parameters to be freed. Therefore both items failed the test of strict invariance. When defining the partial invariance model for Sc8, one of the seven items from the Gass (1991) correction procedure required parameters to be freed. Therefore six items (106, 147, 165, 180, 295 and 325) met the requirement of strict invariance with one item (170) failing the test of strict invariance.

In total three of the items from the Gass (1991) correction procedure failed the test of strict invariance (31, 170 and 172). Therefore excluding the practical impact analysis 11 of 14 Gass (1991) correction procedure items were found to be invariant across a Community and a TBI sample. These 11 items are therefore psychometrically equivalent when assessing persons both with and without a TBI. This equivalence suggests there is no reason for concern that responses to these 11 items by persons with a TBI will falsely inflate a scale score due to neurological content of an item. This evidence alone is sufficient to conclude that clinicians should refrain from employing the Gass (1991) correction procedure.

In addition the tests of no practical impact found evidence to support continued use of the three items from the Gass (1991) correction procedure that failed the test of strict invariance. Millsap and Kwok (2004) provide a procedure to evaluate whether the impact of non-invariant items is of clinical concern, defined in terms of changes in sensitivity and specificity between the preferred condition of strict invariance and the observed condition of partial invariance. The tests of no practical impact are required for any factor that contains an item which failed the test of strict invariance. An item fails the tests of no practical impact when the item is part of a factor where the sensitivity or specificity value in the partial invariance condition is below the lower bound of the 95% CI in the strict invariance condition. With an exception to this rule being where the absolute value for either sensitivity or specificity in the partial invariance condition is acceptable. Three items from the Gass (1991) correction procedure failed the test of strict invariance, namely 31, 170 and 172. Item 31 is part of Hy3 Factor-1, item 172 is part of Hy3 Factor-2 and item 170 is part of Sc8 Factor-1. As Table 6.10 shows Hy3 Factor-1 and Hy3-Factor-2 both met the requirements of no practical impact. Table 7.8 shows Sc8 Factor-1 also met the requirements of no practical impact. The result being that none of the three Gass (1991) correction procedure items that failed the test of strict invariance were found to result in an important practical impact for the MMPI-2. Therefore 11 of the 14 items from the Gass (1991) correction procedure meet the criteria of strict invariance. The remaining three items meet the requirement of no practical impact from failure to find strict invariance. These findings suggest that that all 14 items included in the Gass (1991) correction procedure are suitable measures of psychopathology in persons who have suffered a TBI. There is no evidence to support the proposal by Gass to remove these 14 items when scoring the MMPI-2 in a person with a TBI.

These are important findings as continued use of the Gass (1991) correction procedure may result in failure to identify psychopathology in persons who have suffered a TBI. Due to the overlap of items across scales, the Gass (1991) correction procedure includes items that form part of eight MMPI-2 clinical scales; Hs1, D2, Hy3, Pd4, Pt6, Sc8, Ma9 and Si0. These scales are important components in the identification of a wide variety of psychopathologies, such as anxiety and depressive disorders which are the most common forms of psychopathology found in persons suffering a TBI (Greene, 2000;

Kreutzer et al., 2001; Moore et al., 2006; Seel et al., 2003). Furthermore the importance of two-point codes to the interpretation of MMPI-2 profiles illustrates that the findings are more important than solely whether any single scale score moves across a diagnostic threshold or cut-off point. Even a change in the relative importance of two scales that are scored as clinically relevant can lead to important changes in the interpretation of a MMPI-2 profile. The crucial message is that removing psychometrically valid items from MMPI-2 scales risks failing to identify the types of psychopathology that commonly occur after a TBI.

Studies have established that people with comorbid TBI and psychopathology experience both physical and psychological gains from timely treatment (Anson & Ponsford, 2006c; Fann et al., 2001). However, without a diagnosis of psychopathology, the application of timely therapy is unlikely. Consequently there is an appreciation for the crucial role of accurate assessment as a precursor to providing therapy as part of rehabilitation post a TBI (Dawson et al., 2007; Handel et al., 2007; Seel et al., 2003). The findings from this dissertation suggest using the Gass (1991) correction procedure may result in a clinician missing treatable psychopathology in a patient with a TBI. Therefore using the Gass (1991) correction procedure can potentially delay rehabilitation after a TBI. At worst a person may endure years of untreated psychopathology.

Conclusion for Gass (1991) correction procedure

The findings in this dissertation are a valuable addition to the previous research into the Gass (1991) correction procedure, which has produced conflicting results. Some previous research supports the correction procedure (Gass & Wald, 1997; Rayls et al., 1997). While other studies conclude the correction procedure is either inappropriate or has important limitations (Arbisi & Ben-Porath, 1999; Brulot et al., 1997; Edwards et al., 2003; Glassmire et al., 2003). No previous studies have employed measurement invariance to assess the Gass (1991) correction procedure. Measurement invariance testing or evaluation of differential item functioning is recommended prior to using test instruments in populations that differ from those employed during the design phase (Muniz & Bartram, 2007). APA guidelines recommend some validity evidence should be part of test development and usage (American Education Research Association,

1999). Using measurement invariance testing to assess the Gass (1991) correction procedure follows these recommendations. As no previous research into the Gass (1991) correction procedure has used measurement invariance testing, the findings from this dissertation are important additions to the previous research. The findings from measurement invariance and practical impact analysis completed in this study do not support the use of the Gass (1991) correction procedure. This conclusion is concordant with previous research that concluded clinicians should refrain from using the Gass (1991) correction procedure (Brulot et al., 1997; Edwards et al., 2003; Glassmire et al., 2003).

Conclusion for the neurological content hypothesis

The Gass (1991) correction procedure evolved from the theory that items containing reference to neurologic symptoms would bias MMPI-2 profiles in patients with TBI. The MMPI-2 was designed using the criterion-keying approach. Using this methodology requires an item to be proficient at predicting psychopathology there is no requirement for a theoretical basis to item selection. A common concern is that profiles of persons with a TBI are inflated due to the content of some items reflecting neurological symptoms which often occur after injury (Alfano et al., 1991; Gass, 1991; Gualtieri & Johnson, 1999; Nelson et al., 1989). The Gass (1991) correction procedure implies that the 14 items selected for removal are sensitive to neurologic damage from a TBI and these items represent a specific source of failure of invariance (Gass, 2009). Therefore a measurement invariance assessment of the Gass (1991) correction procedure is an important study of the hypothesis that items with neurological content in the MMPI-2 inflate the profiles of TBI patients.

Measurement invariance analysis was able to determine which items were the sources of variance between a TBI and a Community sample by using the backwards elimination procedure to define the partial invariance model. Following this approach only three of the 14 items from the Gass (1991) correction procedures were observed to fail the test of strict invariance, namely items 31, 170 and 172. Despite only finding three items that were non-invariant, some may interpret this to support the hypothesis of a bias to increased scale scores from neurologic content items. The review of item thresholds observed item 31 as being more likely to be endorsed by the TBI sample, while items

170 and 172 were less likely to be endorsed by the TBI sample. Therefore only item 31 is observed to have a greater likelihood of endorsement in the TBI sample which could potentially lead to a bias of higher scale scores. However, with item 172 having a lesser likelihood of endorsement, and both item 31 and 172 are part of Hy3, the differences may cancel each other out when scores for the scale are calculated. This may result in no impact from the failure of strict invariance under the proposal by Borsboom (2006).

The crucial issue is whether a non-invariant item leads to diminished criterion-related validity of a test instrument. Reduced functioning of the MMPI-2, when used with a person suffering a TBI, is implied by the proposition that items with neurologic content inflate the clinical scales. The practical impact analysis finds that none if the three non-invariant items from the Gass (1991) correction procedure are inappropriate when assessing a person with a TBI.

There were nine items identified across Hy3 (161 and 185) and Sc8 (17, 92, 190, 278, 281, 291 and 303) as both non-invariant, and potentially causing significant reduction in the performance of the clinical scales. It may have been that these items were overlooked by Gass when defining his correction procedure, and therefore they do reflect the concerns with items containing neurologic content. However, review of these items found none that included content that resembled potential neurological symptoms found with a TBI. Therefore the only conclusion supported by the data is that items with neurologic content are not unduly biasing the profiles of patients with a TBI.

Borsboom (2006) advises that measurement invariance testing is likely to find sources of variance across samples in diagnostic instruments. Borsboom explains that item parameters are likely to differ by some degree across populations, with large samples and scales with many items these differences can lead to failing the test of strict invariance. Therefore it is not surprising to find items that are a specific source of failure of strict invariance across a TBI and a community sample. In fact 52 of the 140 items analysed across the three clinical scales were identified as sources of thresholds and/or loadings variance. Strong support for the neurological hypothesis would expect only the items with neurological content to be a specific source of variance and these items reduce the validity of clinical scales. However, only three of these 52 items

identified as failing the test of strict invariance were items identified by Gass as part of his correction procedure. Importantly only nine of the 52 items found as a source of variance were deemed to have an important practical impact and none of these items contained neurological content. The findings do not suggest that none of MMPI-2 items contain neurologic content. However, the findings do suggest that the items with neurologic content are not resulting in a bias towards inflated scale scores in people suffering a TBI.

Conclusion for correction procedures that follow the neurological content hypothesis

The Gass (1991) correction procedure is designed for persons who have suffered a TBI. However there are other MMPI-2 correction procedures based on similar concerns about items with neurological content. Correction procedures have also been suggested for persons with multiple sclerosis, spinal cord injury and temporal lobe epilepsy (Barncord & Wanlass, 2000; Kendall et al., 1978; Meyerink et al., 1988; Nelson et al., 2004). The findings in this dissertation raise doubts over the appropriateness of any correction procedure that is yet to be validated using measurement invariance analyses. Therefore the analytic approach completed in this dissertation would be a valuable investigation of all MMPI-2 correction procedures recommended for other populations.

Conclusion for items that failed both the tests of strict invariance and of no practical impact

The results therefore fail to find support for the contention that items with neurologic content should be removed from scoring MMPI-2 profiles with persons who have suffered a TBI. However there is data to suggest that some items from the MMPI-2 may be inappropriate. These are items 161 and 185 from Hy3, and 17, 92, 190, 278, 281, 291 and 303 from Sc8 which were observed as failing the both the test of strict in variance and the tests of no practical impact. This finding may be considered to support the recommendation by O'Shanick and O'Shanick (2005) that MMPI-2 interpretations are to be avoided when assessing personality disorders in patients with a comorbid TBI. However there are alternative interpretations that should be considered.

The first option is to use MMPI-2 as a filter and augment with another assessment tool. The strength of the MMPI-2 is that is provides a wealth of clinical information that can

assist in diagnosing a variety of psychopathologies. If a clinician believes pathologies related to Hy3 or Sc8 exist, or do not exist, yet the MMPI-2 profile oppose this suspicion then further analyses may be prudent. The first step would be to determine if raw scores on Hy3 or Sc8 are below, or above, diagnostic threshold by less than the number of problematic items identified (i.e. the 2 on Hy3 and 7 items on Sc8 respectively). If this were the case then assessment with another well validated instrument relevant to the suspected psychopathology should be considered. However, this approach would require any additional assessment tool utilised has been subjected to measurement invariance analysis for use with the TBI population.

Another option is to do nothing and continue to use all items. As previously discussed unnecessary pharmacological treatment and the potential harmful consequences are to be avoided and are a potential undesired consequence from an incorrect diagnosis (Spitzer & Frances, 2011). However a potentially false-positive diagnosis and unnecessary provision of psychotherapy does not entail the same risks. Therefore the clinical importance of the practical impact may be lessened by a decision to use psychotherapy while concurrently further assessing the person with a TBI for Hy3 or Sc8 related psychopathology. The clinician could rely on the clinical information derived from the therapeutic process and refine their working diagnosis if necessary, as if often the case during therapy.

The third option is to remove the problematic items from the assessment procedure with persons who have suffered a TBI. However removing these items alters the underlying measurement model of Hy3 and Sc8. Millsap and Kwok (2004) recommend that when items are removed the procedure to define a factor-model and complete measurement invariance with practical impact analysis should be repeated. This process may again lead to the conclusion that items should be removed and therefore the procedure is continually repeated until retaining all items from the scale is supported. This approach is the most prudent and a future study completing this analysis would be invaluable. Importantly any MMPI-2 scale which has an item removed following this procedure requires a *T*-score recalibration method to maintain the validity of clinical interpretations.

In the overview to the MMPI-2 in Chapter Two the project to restructure MMPI-2 clinical scales was described (Tellegen et al., 2003). To reiterate briefly clinical scales were redesigned to reduce the correlations between scales in response to concerns with discriminant validity. This was achieved by defining a new 'Demoralisation' scale which aimed to represent to overarching first factor found across the clinical scales (Frank, 1974). Eight more restructured clinical scales were developed which aimed to represent the core construct of each existing MMPI-2 clinical scale after removing the demoralisation component. Finally items from the MMPI-2 were not allowed to be allocated to multiple scales. In the process some items from the MMPI-2 were removed when designing the nine restructured clinical scales. Importantly eight of the nine items that failed the tests of strict invariance and no practical impact were removed, with only item 161 remaining on restructured scale 7. At face value the removal of these underperforming items appears to support the restructure scales as appropriate for use with a person suffering a TBI. However, it is important that these new scales be subjected to the same measurement invariance and practical impact analysis completed in this dissertation.

Support for the Stress-Appraisal-Coping model of psychopathology

The conclusion that the neurologic content of some MMPI-2 items does not falsely inflate scale scores for persons who have suffered a TBI provides clarification of the mechanisms that underpin the experience of psychological distress after injury. The results support understanding psychological distress for a person with a TBI by applying the Stress-Appraisal-Coping model. However, it is important to note that the support for this model is indirect. The support for the Stress-Appraisal-Coping model comes from removing the confusion that neurological influences are biasing the MMPI_2 profiles. This clarification helps understand the injury severity paradox as potentially explained by the Stress-Appraisal-Coping model, although this conclusion further requires explicit hypothesis testing. Furthermore the alternate hypothesis that malingering underpins the injury severity paradox was not tested and while not the focus of this section it cannot be discounted.

An important component of the Stress-Appraisal-Coping model is the role of insight. With the results failing to support concerns that items containing neurological content

bias scale scores, the injury severity paradox found in TBI samples can be more comfortably explained using the role of insight under the Stress-Appraisal-Coping model. This understanding has important clinical benefit as the Stress-Appraisal-Coping model can provide valuable clinical information to help focus therapy for a person with comorbid psychopathology and TBI. These understandings will now be discussed in more detail.

Insight and the injury severity paradox

Research consistently observes that those who suffer a mild TBI are at increased risk of psychological distress compared with those suffering moderate or severe injuries (Crowe, 2008). The situation has been labelled the injury severity paradox in this dissertation. The injury severity paradox is repeatedly observed on the MMPI-2 scales Hs1, D2, Hy3, Pt7 and Sc8 (Cripe, 1999; Kurtz et al., 2007; Youngjohn et al., 1997). The results have already been interpreted above, to show that there is no support for the hypothesis that items with neurological content unduly inflate scale scores. To reiterate, as stated in the Introduction, the term 'distress' is employed as shorthand for multiple types of psychopathology.

This finding is important because the injury severity paradox is difficult to align with the neurological content-bias hypothesis. The hypothesis implies the more severely injured would endorse more items with neurological content, which should bias scale scores to being higher with injury severity. Yet the more severely injured are found to endorse fewer items on a clinical scale. Bogod, Mateer, & MacDonald (2003) found that an increase in TBI severity was associated with a decrease in a measure of awareness of injury related deficits. Therefore an alternative explanation for the observed frequencies of item endorsements is that a person with a more severe TBI is less likely to endorse items due to a lack of insight into their deficits. There are two competing explanations as to the underlying mechanisms influencing endorsement of the neurologic items. The findings from this dissertation do not support the neurological content explanation, which further supports the role of diminished insight.

The observation that MMPI-2 profiles are not biased by items with neurological content removes concerns that the injury severity paradox is an artefactual confound specific to

the MMPI-2. The conclusion that the injury severity paradox is explained by the role of reduced insight is supported by other studies. Kurtz, Shealy and Putnam (2007) reviewed the scores for individuals who suffered a TBI on both the Personality Assessment Inventory and the MMPI-2. Kurtz and colleagues observed the injury severity paradox on both clinical instruments. This finding supports understanding the injury severity paradox on the MMPI-2 as an accurate expression of psychological distress which varies with the severity of TBI.

Persons who have suffered a TBI and are aware of their functional deficits occurring as a result of their injury are at greater risk for developing psychopathology. Morton and Wehman (1995) completed a literature review of the emotional consequences of TBI. They observed that research consistently finds persons with mild TBI are keenly aware of their deficits. In contrast, persons with severe TBI were less likely to show awareness of any functional limitations. Essentially this observation finds an injury severity paradox for insight into functional deficits post a TBI.

Therefore both insight and psychological distress reflect the injury severity paradox. From a clinical perspective it is important to determine whether insight influences psychological distress, or vice versa. Godfrey, Partridge, Knight and Bishara (1993) found the return of insight in sufferers of TBI was coupled with an increase in risk of psychological distress. Furthermore the key role of insight in the experience of psychological distress is not limited to those with only a mild TBI. Cooper-Evans and colleagues (2008) observed that in a sample of only severely injured TBI patients, higher levels of insight were related to lower measures of self-esteem which was in turn correlated with psychological distress. Consequently across the full spectrum of TBI severity the research shows insight as an underlying mechanism related to psychological distress.

From a therapeutic perspective recognising the connected between insight into deficits and the experience of psychological distress is invaluable. This recognition permits a clinician to understand psychopathology in persons with a TBI as an outcome of the Stress-Appraisal-Coping model for psychological distress. The Stress-Appraisal-Coping model proposes psychological distress is not resulting from specific characteristics of an

event. Psychological distress is an outcome from how an individual appraises their capacity to cope with the characteristics of an event. Therefore psychological distress after a TBI is related to the injured person's appraisal of their ability to cope with any adaptational challenges' arising consequent to their injury.

Insight and the Stress-Appraisal-Coping model

It may not be immediately clear how the appraisal process is related to the specific role of insight. During the appraisal stage potential coping strategies are considered with the goal to alleviate psychological distress. If an injured person lacks insight then during the appraisal stage they may be unaware of their coping abilities, which could result in experiencing psychological distress. However the Stress-Appraisal-Coping model describes the experience of psychological distress as a two-step process (Lazarus & Folkman, 1984). An important precursor to appraising one's capabilities is recognising the need for a person to employ a coping strategy. It is in this stage that the reduced insight can protect against experiencing psychological distress (Godfrey et al., 1996).

In the first stage of the Stress-Appraisal-Coping model a person evaluates whether the needs of the environment require a coping strategy (Lazarus & Folkman, 1984). Only if subjectively required does the person then evaluate whether they have the resources to successfully employ a coping strategy. If during the first stage, a situation is deemed benign then no coping strategy is considered necessary and the risk for psychological distress is absent. In the second stage, if a situation is deemed potentially harmful and a person is able to successfully employ a coping strategy, psychological distress is again absent (Lazarus & Folkman, 1984). Therefore not perceiving a situation as stressful can have the same emotional outcome as being able to successfully employ a coping strategy. The reduced insight found in persons with a severe TBI (Bogod et al., 2003) is proposed to diminish the likelihood that they will appraise a situation as potentially harmful and therefore requiring a coping strategy. The consequence from the reduction in insight is the person with severe TBI is less likely than a person with mild TBI to experience psychological distress, which could develop into psychopathology.

The key role of the appraisal stage in the Stress-Appraisal-Coping model is supported by Crowe (2008) who found appraisal of events was a better predictor of psychological

distress than the number of stressful events experienced. Findings such as this are repeatedly observed for patients who have suffered a TBI (Machulda et al., 1998; Strom & Kosciulek, 2007). Therefore the suggestion that insight into deficits is the underlying mechanism that explains the experience of psychopathology in a person with a TBI is well documented. Furthermore the increase in psychopathology levels observed in TBI samples with duration after injury is also likely reflecting the role of insight into deficits in the development of psychopathology (Godfrey et al., 1993; MacNiven & Finlayson, 1993). As the severity of the injury lessens over time, with natural recovery of the neurological structures damaged, insight into deficits improves. As such both the injury severity paradox and the differing levels of psychopathology found over duration may reflect the same underlying mechanisms outlined in the Stress-Appraisal-Coping model.

The findings from this study further supported the understanding that psychopathology in a person with a TBI is an outcome of the Stress-Appraisal-Coping process. The discordant proposition that MMPI-2 profiles reflect item endorsement due to neurological symptoms is not supported. To reiterate, there were two competing explanations as to what were the underlying mechanisms influencing endorsement of the MMPI-2 neurologic items, 1) item content or 2) a valid expression of psychological distress. The findings from this dissertation fail to support the item content proposition.

The proposal to understand psychopathology in a person who has suffered a TBI with the Stress-Appraisal-Coping model is not intended to deny the existence, or debilitating effects of symptoms associated with neurological damage. The research is unequivocal that damage to specific neurological regions can result in functional deficits. Crowe (2008) and Miller (1996) highlight the common neuropsychological deficits post a TBI include decreased speed of information processing, working memory deficits, impulsivity, interpersonal communication difficulties to name a few. Importantly a lack of insight is also identified was sequelae to neurological damage.

Neuropsychological deficits can influence the outcome of both the appraisal and the coping strategy stages in the Stress-Appraisal-Coping model. The central role of reduced insight in the appraisal stage has been discussed in detail. However, other deficits such as impulsivity may lead to quick and inaccurate judgements regarding the

potential harmfulness in a situation. Additionally coping strategies may be limited by slower information processing speed and reduced working memory. Or strategies may be incorrectly implemented due to interpersonal communication difficulties. In all the scenarios outlined, the consequence is psychological distress that can be traced directly to the neuropsychological deficits associated with a TBI.

Importantly all the scenarios outlined above reflect the interrelationship between neuropsychological symptoms and psychological distress. In Chapter One a neuropsychosocial approach to understand the development of psychopathology after a TBI was proposed. This framework incorporates an interrelationship between neuropsychological deficits and psychological distress. The Stress-Appraisal-Coping model is concordant with a neuropsychosocial approach to psychopathology in persons who have suffered a TBI. The strength of the neuropsychosocial model is that neurological, psychological and social factors are considered in combination by clinicians to understand the presentation of psychological distress in a person. The benefit of the Stress-Appraisal-Coping model is that is provides important information that can inform therapeutic strategies.

Therapy using the Stress-Appraisal-Coping model

Cognitive behavioural therapy operates on the understanding that an emotional response to an event is mediated by automatic thoughts about the event. Lazarus and Folkman (1984) describe the appraisal stage in particular from the Stress-Appraisal-Coping model as reflecting the cognitive assumption which underlies CBT. By not delineating the coping stage the authors may have undersold the synergy between the Stress-Appraisal-Coping model and CBT. Under the Stress-Appraisal-Coping model the coping process includes conscious or unconscious application of cognitive and behavioural strategies to ameliorate the experience of stress (Vosvick et al., 2010). Therefore, both the appraisal and the coping stages within the Stress-Appraisal-Coping model can be targeted using CBT techniques.

Godfrey, Knight and Partridge (1996) applied the Stress-Appraisal-Coping model to understand emotional adjustment in persons with a TBI. They found that while psychological distress is in response to neuropsychological symptoms and functional

losses, the relationship between these symptoms and losses, and distress is mediated by aspects of the Stress-Appraisal-Coping model. Specifically the appraisal stages and employing coping skills mediates the relationship between neuropsychological symptoms and psychological distress. These findings re-iterate that the Stress-Appraisal-Coping model does not deny the existence of the neurological symptoms described by MMPI-2 items, but that these symptoms can lead to psychological distress via the appraisal and coping mechanisms. The capacity for therapy to target both the appraisal and coping stages is a benefit from applying the Stress-Appraisal-Coping model to understand an individual's difficulties adjusting to changes post a TBI (Strom & Kosciulek, 2007).

Importantly the CBT techniques employed must consider any neuropsychological deficits resulting from a TBI (Folzer, 2001). CBT practitioners are endowed with a multitude of techniques at their disposal. These include identification and alteration of dysfunctional thoughts, behavioural therapy, assertiveness training, problem solving skills and peer support. Dysfunctional thoughts may lead to an inaccurate appraisal of either the potential harm from a situation or underestimating the personal resources available to handle a situation. CBT has been shown to successfully modify coping strategies in persons who have suffered a TBI (Anson & Ponsford, 2006a). While behaviour therapy, an important component of CBT, is recommended when treating Obsessive Compulsive Disorder symptoms in people who have suffered a TBI (Hiott & Labbate, 2002). Peer support can help both the person who has suffered a TBI and their family (Hibbard et al., 2002). When difficulties coping with changes in functioning after a TBI derive from a lack of understanding as to what common skills can be utilised then peer support can be especially helpful (Lazarus & Folkman, 1984).

Therapy with a person suffering a TBI will likely require adaptations to counter any specific neuropsychological symptoms. Folzer (2001) recommends using more concrete approaches to therapy with relaxation techniques to minimise the consequences from over arousal. Relaxation techniques form an integral component to the treatment of many anxiety disorders. Additionally Folzer highlights the importance for psychoeducation of both the injured person and their caregivers including family.

While appreciating the Stress-Appraisal-Coping model can inform therapy it raises a clinical conundrum. It has been shown how this model explains the observed injury severity paradox by highlighting the role of insight in experiencing psychological distress. Additionally psychoeduation is recommended as an important component of therapy (Folzer, 2001). The need for psychoeducation reflects the understanding by clinicians that without acceptance of a problem, then a patient is unlikely to follow a therapeutic program. However psychoeducation may produce increased psychological distress by creating insight. This potential outcome is exemplified by the findings of Smith and Godfrey (1995) that family education improved symptoms awareness in patients with a TBI but the increased awareness was associated with higher levels of depression. Therefore a difficult aspect of therapy is determining the appropriate time to discuss the presence of injury-related deficits.

If diminished insight protects against psychological distress, then therapy for these TBI patients may be considered unnecessary and potentially harmful. However, studies find that underreporting of deficits in person with a TBI has no long term benefit, despite the immediate reduction in psychological distress (Godfrey et al., 1996; Godfrey et al., 1993). Godfrey and colleagues highlight the need to monitor symptoms adequately when educating a person with a TBI about their capabilities. The MMPI-2 is a useful instrument for this purpose as the test covers a wide range of psychopathology. While the measurement invariance and practical impact analysis failed to support the need for a correction procedure, the analysis did identify nine items that were problematic. These are the nine items that failed both the tests of measurement invariance and no practical impact. However, eight of these nine items were removed when the restructured clinical scales were designed. Therefore notwithstanding the need for these restructured scales to be subjected to measurement invariance analysis, the MMPI-2 restructured clinical scales may be preferred.

Conclusion for the Stress-Appraisal-Coping model

The findings from the measurement invariance and practical impact analyses are shown to support the Stress-Appraisal-Coping model in explaining psychological distress in a person who has suffered a TBI. That is, failure to find evidence of any direct bias due to the neurological content hypothesis clarifies the specific role of insight in explaining the

injury severity paradox. The Stress-Appraisal-Coping model was used to show how a lack of insight can protect against experiencing psychological distress for a person with a TBI. The Stress-Appraisal-Coping model is an important clinical guide as it can inform treatment, especially as the model is suited to cognitive based therapies.

8.3 Benefits from using measurement invariance and practical impact analyses

Previous research into the Gass (1991) correction procedure has utilised a variety of different methodologies. Gass and Wald (1997) employed χ^2 analysis to assess differences in item endorsement between a TBI and non-TBI sample. However a limitation of the Gass and Wald study was that it did not control for post injury psychopathology. Edwards et al. (1998) similarly failed to control for post injury psychopathology when they measured whether a person with a TBI considered symptomology to be a result of the injury. Rayls et al., (1997) compared endorsement of correction items in a TBI and the MMPI-2 normative sample. Control for post injury psychopathology in the TBI sample was again missing from the study design. The uncontrolled confound of post injury psychopathology observed in these studies is not a limitation when using measurement invariance testing.

The existence or absence of psychopathology does not change a test of the underlying measurement model. Neither does any concern about whether psychopathology is pre-existing or consequent to a TBI. Measurement invariance analysis directly tests the hypothesis that items in the MMPI-2 measure the same constructs with the same metric relationship between items and constructs in persons with and without a TBI (Bowden et al., 2008). While acknowledging that measurement invariance does not demonstrate the same causal factors underlie the responses, it does establish that the responses equivalently measure the same types of psychopathology. An important innovation in this thesis study was the application of the practical impact analysis to assess the diagnostic effect for any items that failed the test of strict invariance.

By completing measurement invariance and practical impact analysis this study was able to assess various questions. The appropriateness of the Gass (1991) correction procedure was specifically targeted. Simultaneously, the neurological content

hypothesis could be reviewed by determining if any items that failed the tests of strict invariance and no practical impact contained neurological content. Without using the backwards elimination procedure to define a partial invariance model, then determining which items met and which failed the test of strict invariance would be unworkable. Using the practical impact analysis permitted a direct examination of the underlying concern which instigated the neurological-bias hypothesis and the Gass (1991) correction procedure. The concern being that scale scores are falsely inflated from item endorsement for reasons other than psychological distress which reduces the validity of the MMPI-2 profile. Measurement invariance allowed assessment of whether the psychometric properties of items were equivalent across samples. While practical impact analysis assessed whether any non-invariant items were unduly impacting the criterion related validity of diagnostic scales that use these items.

Additionally the results from the measurement invariance and practical impact analyses provide support for the position that insight into injury related deficits as the reason for the injury severity paradox. The support for insight as the causal mechanism provides confidence in the Stress-Appraisal-Coping model. This is valuable for clinicians because the Stress-Appraisal-Coping model provides an understanding of psychopathology which can provide valuable information for therapy. The Stress-Appraisal-Coping model provides a guide as to when cognitions or coping strategies are maladaptive. All of these conclusions were derived from the measurement invariance approach applied in this dissertation. In other words, examination of measurement invariance analysis has strong implications for construct validity.

The application of the backwards elimination procedure to define the partial invariance model facilitated a detailed examination of the items from the Gass (1991) correction procedure. All three clinical scales were observed to fail the test of strict invariance. However the method used to define the partial invariance model led to finding that 11 of the 14 items from the correction procedure did meet the requirements of strict invariance. Importantly on Hs1 all of the Gass (1991) correction procedure items met the criteria of strict invariance. As strict invariance is universally held as sufficient to support for unbiased use of a test, the Gass (1991) correction procedure must be rejected as only three items failed this requirement.

Applying the practical impact analyses was crucial to clarify the interpretation from finding a partial invariance model in the clinical scales. Traditionally if a diagnostic instrument failed the test of strict invariance in a specific population, then the instrument was deemed inappropriate for that population (Meredith, 1993). The requirement of strict invariance is beginning to be considered too conservative (Borsboom, 2006; Millsap & Kwok, 2004). Recent research suggests that it is unlikely an instrument will meet the requirements of strict invariance, especially with tests incorporating large numbers of items (Borsboom, 2006; Byrne & van de Vijver, 2010). Therefore it is becoming increasingly important to determine whether the non-invariant items from a partial invariance model can be retained.

Cheung and Rensvold (1999) consider the non-invariant items can be retained under a partial invariant model provided these items do not markedly affect cross group-comparisons. While Borbsoom (2006) proposed that when the sources of item variance between the samples cancel each other out the result is no impact from the failure of strict invariance when. Schmitt and Kuljanin (2008) reviewed the literature of measurement invariance analyses since 2000 and observed that half the studies were concluding with a partial invariance model. However none of these studies which conclude with a partial invariance model explicitly tested the assumption that the non-invariant items could be retained based on the Cheung and Rensvold (1999) or Borsboom (2006) criteria. Schmitt and Kuljanin suggest the practical impact analysis approach outlined by Millsap and Kwok (2004) would be an important addition to any measurement invariance study that concludes with a partial invariance model. The practical impact analysis facilitates an empirical evaluation of the diagnostic impact from the non-invariant items. A search of PsycINFO of the papers published since 2008 fails to reveal any invariance studies that followed this recommendation.

Applying the Millsap and Kwok (2004) practical impact analysis was a strength of this thesis. This approach was able to conclude that while three items from the Gass (1991) correction procedure were non-invariant, none of these items was biasing the clinical validity of the MMPI-2. Additionally while the study found 52 items that were non-invariant, the study was able to conclude that only nine of these items potentially have an adverse diagnostic impact. A review of these nine items found none contained items

of neurologic content which supports the decision to reject the neurologic content hypothesis. Implicit support for using the restructured clinical scales was found with only one of the nine items included in the new MMPI-2 scales. None of these conclusions would be permissible without the application of the Millsap and Kwok (2004) practical impact analysis.

Duplicate items were removed on Hy3 and Sc8 to protect against repeated testing for the same item and to minimise computational difficulties. If the items on the MMPI-2 were not dichotomous variables then the inclusion of duplicate items on subsequent scales may change the underlying measurement model to such a degree that the conclusions differ. However with dichotomous data, the results of the measurement invariance testing with theta parameterisation are equivalent to a test of differential item functioning under item response theory (Glöckner-Rist & Hoijtink, 2003). Therefore identifying differential item functioning on one scale would be expected to be replicated on any subsequent scale including the same item.

Parcelling of items is one solution to facilitate a single set of analysis on all items from the clinical scales. Little, Cunningham et al. (2002) provide an excellent overview to the costs and benefits of parcelling. Benefits include reducing the role of individual item variance which can be problematic for models with large numbers of items. Additionally parcels better reflect the constructs being measured compared to individual items. Compared with parcels, item level data has lower communality, greater likelihood of distributional violations, lower reliability and a smaller ratio of common-to-unique factor variance. The disadvantages of parceling include that the process may obscure factor model misspecification that would be found if analyses were performed on item-level data. Importantly the authors note pitfalls of parceling are most evident when attempting to identify the exact relations among the individual items comprising the constructs. Therefore parceling may hide non-invariant items because they are included in a parcel that is collectively invariant across samples. Additionally partial invariance established with item level analysis permits completion of the practical impact analysis. Practical impact analysis may be obscured by parcelling.

8.4 Limitations

On Hy3 two items and on Sc8 seven items failed the tests of strict invariance and of no practical impact. Yet none of these items were included in the Gass (1991) correction procedure. Measurement invariance and practical impact analysis of the remaining seven clinical scales not included in this dissertation may reveal more problematic items. Ideally one could investigate all 370 items which comprise the ten MMPI-2 clinical scales in a single analysis. A single analysis would make the requirement to remove duplicate items on subsequent scales unnecessary. Unfortunately software and hardware limitations make the opportunity for item level analysis of the complete set of clinical scale in a single test of invariance difficult if not impossible at the present time.

In some of the scale factors the sensitivity and specificity values were calculated with low cell counts which can be seen by the size of the confidence intervals. Having a larger sample would lead to increase cell counts and this would be a strength of future replication studies where feasible. Alternatively relaxing the cut-off points to be closer to the median score for the distribution would likely result in increased cell counts. However such a change would lead to moving the cut-points away from the clinical interpretation of MMPI-2 profile scores. It is important to keep the decision point for the analysis as best as possible to reflect the decision points in clinical practice.

The TBI sample was gathered from a forensic clinical setting. The finding by Dunn and Lees-Haley (1995) that the Gass (1991) correction procedure is inappropriate in forensic neuropsychological evaluations supports the conclusions from this dissertation. However, Dunn and Lees-Haley did not employ measurement invariance testing to derive their conclusions. A limitation noted above for the previous research on the Gass (1991) correction procedure. It has been suggested that being involved in litigation is a mediating variable between injury and item endorsement (Hoffman et al., 1999; Senior & Douglas, 2001). A TBI sample gathered from a non-forensic medical establishment which is then used to replicate the analysis from this dissertation would alleviate any concerns about the potential confound from litigation. Alternatively a non-TBI forensic sample could also control for the possible confound of litigation.

The findings from this study would benefit from replication in a TBI sample that better represent the population. That is, a TBI sample that is not solely personal-injury evaluation based and a sample also comprising some non-clinical participants. Strength of the sample employed for the current study is that is comprised TBI patients who suffered a spectrum of injury severity, not only the mildly injured patients.

8.5 Future Directions

All new research has an implicit caveat that the study requires replication. Replication studies would be strengthened by employing samples that overcome the limitations discussed previously. Measurement invariance analysis with a non-forensic TBI sample or using a non-TBI forensic sample will overcome the possible limitation of litigation as a potential confound.

Measurement invariance and practical impact analysis using a non-TBI clinical and a TBI sample are recommended. Whilst the defined factor-models were replicated in the Community and TBI samples, an additional clinical sample would further support the conclusions. Analysis of treatment outcomes based on diagnoses from a TBI sample would further support that the uncorrected MMPI-2 is appropriate for diagnosis of psychopathology in patients with a TBI. It is accepted that psychological therapy is beneficial and arguably fundamental to patient's recovery after suffering a TBI. Therefore the accuracy of MMPI-2 assessment can be tested by analysing the impact treatment has on those patients with a TBI, compared to those without a TBI, who are diagnosed with psychopathology using the MMPI-2. If the MMPI-2 diagnoses correctly, then patients with a TBI who are diagnosed with a psychological disorder would be expected to show comparable benefits from psychotherapy to patients without a TBI who are diagnosed with psychopathology. An alternative option to compare treatment outcome for persons with a TBI who are diagnosed using the MMPI-2, with treatment outcome for those diagnosed using an alternative diagnostic instrument. However, before using an alternative diagnostic instrument the tool would need to be subjected to measurement invariance analysis to support the instruments usage in the TBI population.

Two items from Hy3 and seven items from Sc8 were found to fail both the tests of strict invariance and the tests of no practical impact. Removing these items and completing the measurement invariance and practical impact analysis is the recommended approach to determine which items from Hy3 and which from Sc8 are worth removing (Millsap & Kwok, 2004). This approach may support removing additional items from the scale when assessing patients with a TBI. The final list of items from Hy3 and Sc8 that are derived can only be proposed as a correction procedure when a method to recalibrate *T*-scores for the scale is defined. The recalibration method is crucial to allow clinicians to follow contemporary guides to assessing *T*-scores and code types.

Alternatively the restructured clinical scales (Tellegen et al., 2003) may be preferred. Only one item (161) remains included in the restructured clinical scales from the nine that were observed to fail the tests of strict invariance and no practical impact. When designing the restructured clinical scales the authors removed the overarching first factor. Additionally for each clinical scale a core construct was defined with only items that measure this construct remaining. Therefore some MMPI-2 items were excluded from the restructured clinical scales. The aim of this procedure was to remove items that were not strong indicators of the constructs being measured. Eight of the nine items that failed the test of strict invariance and no practical impact were excluded from the restructured clinical scales. This finding suggests the design process appears to have achieved its goal of removing poorly performing items, at least for the TBI population. However, measurement invariance is required on the nine restructured clinical scales to support this conclusion.

Correction procedures have been proposed for patients suffering multiple sclerosis, cerebrovascular disease and spinal cord injury (Barncord & Wanlass, 2000; Gass, 1992; Kendall et al., 1978; Meyerink et al., 1988; Millsap & Kwok, 2004). None of these procedures have been subjected to measurement invariance and practical impact analysis. The finding from this study which fail to support the Gass (1991) correction procedure in the TBI population, casts doubt on other procedures that have yet to be subjected to the measurement invariance analysis. Analysis following the approach from this dissertation can assess whether the items identified should be retained or removed, and whether additional items not identified are potentially biasing assessment.

The support for the Stress-Appraisal-Coping model found in this study can be explicitly tested using the restructured clinical scales developed by Tellegen and colleagues (2003). The newly developed demoralisation scale (RcD) provides the opportunity to assess the role of appraisal and coping mechanisms in adjustment after a TBI. As Lazarus and Folkman (1984) note that coping is defined by the efforts to manage stressful demands. The RcD measures the general maladjustment in individuals with psychopathology, which can be seen as a measure of the outcome from the coping process. Understanding this process can inform clinicians as to the most appropriate first line of psychotherapy appropriate for patients suffering a TBI during their rehabilitation. The generalised distress measured by the RcD scale may be able to inform clinicians as to the appropriateness of CBT for an individual with a TBI. That is a high scored on RcD indicating maladaptive processing within the Stress-Appraisal-Coping model suggests CBT is appropriate for this individual. Lower scores on the RcD scale would suggest other factors are primarily contributing to the patient's presentation, and these would be the focus for rehabilitation.

8.6 Conclusions

Measurement invariance and practical impact analysis failed to find support for the Gass (1991) correction procedure. Therefore clinicians should avoid using this correction procedure when completing a MMPI-2 assessment on a person suffering a TBI. The analysis did however identify two items from Hy3 and seven from Sc8 that may bias assessment of people with TBI. A procedure was outlined to complete assessment of Hy3 and Sc8 to derive a final number of items that could be removed for an MMPI-2 assessment on a person with a TBI. If a correction procedure is defined for these scales then a method to recalibrate *T*-scores for the scale is mandatory.

Alternatively the restructured clinical scales (Tellegen et al., 2003) may be a preferred option. The procedure employed to design these new scales removed eight of the nine items that were identified to potentially bias assessment of a person who has suffered a TBI. This observation provides support for the restructured scales. However the restructured scales require measurement invariance testing with a TBI sample before it can be concluded they are preferred.

Finally, the findings suggest the concern that items with neurological content will bias clinical scale scores in patients with a TBI was not supported. This conclusion was shown to be important in understanding the specific role of insight using the Stress-Appraisal-Coping model in the development of psychopathology in persons with a TBI. Support for the Stress-Appraisal-Coping model is important as it provides a clinical therapeutic model to alleviate psychological distress which is at increased risk for persons who have suffered a TBI.

References

Access Economics Pty Ltd. (2009). *The economic cost of spinal cord injury and traumatic brain injury in Australia*: Victorian Neurotrauma Initiative.

Al-Adawi, S., Dorvlo, A. S. S., Al-Naamani, A., Glenn, M. B., Karamouz, N., Chae, H., et al. (2007). The ineffectiveness of the Hospital Anxiety and Depression Scale for diagnosis in an Omani traumatic brain injured population. *Brain Injury, 21*(4), 385 - 393.

Alderfer, B. S., Arciniegas, D. B., & Silver, J. M. (2005). Treatment of depression following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 20*(6), 544-562.

Alfano, D. P., Finlayson, M. A., Stearns, G. M., & MacLennan, R. N. (1991). Dimensions of neurobehavioural dysfunction. *Neuropsychology, 5*(1), 35 - 41.

Alfano, D. P., Finlayson, M. A., Stearns, G. M., & Neilson, P. M. (1990). The MMPI and neurologic dysfunction: Profile configuration and analysis. *The Clinical Neuropsychologist, 4*(1), 69 - 79.

Alfano, D. P., Neilson, P. M., Paniak, C. E., & Finlayson, M. A. (1992). The MMPI and closed-head injury. *Clinical Neuropsychologist, 6*(2), 134 - 142.

Alfano, D. P., Paniak, C. E., & Finlayson, M. A. (1993). The MMPI and closed head injury a neurocorrective approach. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 6*(2), 111 - 116.

Alkemade, N. R. (2007). Are correction models appropriate for MMPI-2 Scale 1 with patients suffering a traumatic brain injury? Unpublished Honours Thesis. University of Melbourne.

Almagor, M., & Koren, D. (2001). The adequacy of the MMPI-2 harris-ligoes subscales: a cross-cultural factor analytic study of scales D, Hy, Pd, Sc, and Ma. *Psychological Assessment., 13*(2), 199 - 215.

American Education Research Association. (1999). *Standards in educational and psychological testing: American Education Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME)*. Washington, DC: American Education Research Association.

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

Anson, K., & Ponsford, J. (2006a). Coping and emotional adjustment following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 21*(3), 248-259.

Anson, K., & Ponsford, J. (2006b). Evaluation of a coping skills group following traumatic brain injury. *Brain injury, 20*(2), 167-178.

Anson, K., & Ponsford, J. (2006c). Who benefits? Outcome following a coping skills group intervention for traumatically brain injured individuals. *Brain injury, 20*(1), 1-13.

Arbisi, P. A., & Ben-Porath, Y. S. (1997). Characteristics of the MMPI-2 F(p) scale as a function of diagnosis in an inpatient sample of veterans. *Psychological Assessment, 9*, 102–105.

Arbisi, P. A., & Ben-Porath, Y. S. (1999). The use of the Minnesota Multiphasic Personality Inventory-2 in the psychological assessment of persons with TBI: Correction factors and other clinical caveats and conundrums. *Neurorehabilitation, 13*, 117 - 125.

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2003). Empirical correlates of common MMPI-2 two-point codes in male psychiatric inpatients. *Assessment, 10*(3), 237-247.

Arbisi, P. A., & Butcher, J. N. (2004). Relationship between personality and health symptoms: use of the MMPI-2 in medical assessments. *International journal of clinical and health psychology, 4*(3), 571-595.

Archer, R. P., & Newson, C. P. (2000). Psychological test usage with adolescent clients: survey update. *Assessment, 7*(3), 227-235Arciniegas, D. B., Harris, S. N., & Brousseau. (2003). Psychosis following traumatic brain injury. *International review of psychiatry, 15*, 328-340.

Avdeyeva, T. V., Tellegen, A., & Ben-Porath, Y. (2012). Empirical correlates of low scores on MMPI-2/MMPI-2-RF restructured clinical scales in a sample of university students. *Assessment, 19*(3), 388-393. doi:http://dx.doi.org/10.1177/1073191111411675

Bachna, K., Sieggreen, M. A., Cermak, L., Penk, W., & O'Connor, M. (1998). MMPI/MMPI-2: comparisons of amnesic patients. *Archives of clinical neuropsychology, 13*(6), 535 - 542.

Barncord, S. W., & Wanlass, R. L. (2000). A correction procedure for the Minnesota Multiphasic Personality Inventory-2 for persons with spinal cord injury. *Archives of physical medicine rehabilitation, 81*, 1185 - 1190.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186-203.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Gass, C. S., Franzen, M. D., Youngjohn, J. R., et al. (1995). Over reporting of closed-head injury symptoms on the MMPI-2. *Psychological assessment, 74*(4), 517 - 523.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5*, 404–417.

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2007). Anxiety disorders and suicidal behaviours in adolescence and young adulthood: Findings from a longitudinal study. *Psychological Medicine, 37*, 431 - 440.

Bogod, N. M., Mateer, C. A., & MacDonald, S. W. S. (2003). Self-awareness after traumatic brain injury: A comparison of measures and their relationship to executive functions. *Journal of the Internation Neurospcyhological Society, 9*, 450-458.

Bolinskey, P. K., & Nichols, D. S. (2011). Construct drift in the MMPI-2 restructured clinical scales: Further evidence and a possible historic example. *Journal of Clinical Psychology, 67*(9), 907-917. doi:http://dx.doi.org/10.1002/jclp.20814

Bontempo, D. E., & Hofer, S. M. (2007). Assessing factorial invariance in cross-sectional and longitudinal studies. In A. D. Ong & M. H. M. Van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 153-175). New York: Oxford University Press.

Borsboom, D. (2006). When does measurement invariance matter? *Medical care, 44*(11), S176 - S181.

Bowden, S. C., Gregg, N., Bandalos, D., Davis, M., Coleman, C., Holdnack, J. A., et al. (2008). Latent mean and covariance differences with measurement equivalence in college students with developmental difficulties versus the Wechsler Adult Intelligence Scale-III/Wechsler Memory Scale-III normative sample. *Educational and psychological measurement, 68*(4), 621-642.

Brain Injury Association of Queensland. (2006). Australian statistics on acquired brain injury. Retrieved 15/03, 2007, from http://www.biaq.com.au/pdfs/factsheets/Australian%20statistics%20on%20acquired%20brain%20injury%20-%202006.pdf.

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

Brulot, M. M., Strauss, E., & Spellacy, F. (1997). Validity of the Minnesota Multiphasic Personality Inventory-2 for use with patients with suspected head injury. *The Clinical Neuropsychologist, 11*(4), 391 - 401.

Bryant, R. A., & Harvey, A. G. (1999). The influence of traumatic brain injury on acute stress disorder and post-traumatic stress disorder following motor vehicle accidents. *Brain Injury, 13*(1), 15 - 22.

Busch, C. R., & Alpern, H. P. (1998). Depression after mild traumatic brain injury: A review of current research. *Neuropsychology Review, 8*, 95 - 108.

Butcher, J. N. (Ed.). (2006). *MMPI-2 a practitioner's guide*. Washington, DC: American Psychological Association.

Butcher, J. N. & Rouse, S. V. (1996) Personality: Individual differences and clinical assessment, *Annual Review of Psychology, 47*, 87-111.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Lim, J., & Nezami, E. (1998). Objective study of abnormal personality in cross-cultural settings: the Minnesota Multiphasic Personality Inventory (MMPI-2). *Journal of Cross-Cultural Psychology, 29*, 189 - 214.

Byrne, B. M. (1998). *Structural equation modelling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications, and programming.* New Jersey: Lawrence Erlbaum Associates.

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: addressing the issue of nonequivalence. *International journal of testing, 10*, 107 - 132.

Calabrese, W. R., Rudick, M. M., Simms, L. J., & Clark, L. A. (2012). Development and validation of big four personality scales for the schedule for nonadaptive and

adaptive personality--second edition (SNAP-2). *Psychological Assessment, 24*(3), 751-763. doi:http://dx.doi.org/10.1037/a0026915

Caldwell, A. B. (2006). Maximal measurement or meaningful measurement: the interpretive challenges of the MMPI–2 restructured clinical (RC) scales. *Journal of personality assessment 87*(2), 193-201.

Cameron, C. M., Purdie, D. M., Kliewer, E. V., & McClure, R., J. (2008). Ten-year outcomes following traumatic brain injury: A population-based cohort. *Brain Injury, 22*(6), 437-449.

Caplan, B., & Shechter, J. (1995). *The role of nonstandard neuropsychological assessment in rehabilitation: history, rationale, and examples*. Washington, DC: American Psychological Association.

Chang, C.-H. (1996). Finding two dimensions in MMPI-2 depression. *Structural Equation Modeling, 3*(1), 41 - 49.

Chambless, D.L., Sanderson, W.C., Shoham, V., Bennett-Johnson, S., Pope, K.S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S.R., Sue, S., Beutler, L., Williams, D.A., & McCurry, S. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*, 5-18

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling, 14*(3), 464-504.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of management, 25*(1), 1-27.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233 - 255.

Chisholm, S. M., Crowther, J. H., & Ben-Porath, Y. S. (1997). Selected MMPI-2 scales' ability to predict premature termination and outcome from psychotherapy. *Journal of Personality Assessment, 69*(1), 127-144

Cooper-Evans, S., Alderman, N., Knight, C., & Oddy, M. (2008). Self-esteem as a predictor of psychological distress after severe acquired brain injury: An exploratory study. *Neuropsychological Rehabilitation, 18*(5/6), 607-626.

Corcoran, C., McAllister, T. W., & Malaspina, D. (2005). Psychotic disorders. In J. M. Silver, T. W. McAllister & S. C. Yudofsky (Eds.), *Textbook of traumatic brain injury*. Arlington, VA: American psychiatric publishing Inc.

Coyne, J. C., & Schwenk, T. L. (1997). The relationship of distress to mood disturbance in primary care and psychiatric populations. *Journal of Consulting and Clinical psychology, 65*(1), 161-168.

Cripe, L. (1999). Use of the MMPI with mild closed head injury. In N. R. Varney & R. J. Roberts (Eds.), *The evaluation and treatment of mild traumatic brain injury* (pp. pp. 291 - 314). New Jersey: Lawrence Erlbaum Associates Inc. .

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper.

Crowe, S. F. (2008). *The behavioural and emotional complications of traumatic brain injury*. Philadelphia: Taylor & Francis.

Davis, P. J., Reeves, J. L., Hastie, B. A., Graff-Radford, S. B., & Naliboff, B. D. (2000). Depression determines illness conviction and pain impact: A structural equation modeling analysis. *Pain Medicine, 1*(3), 238 - 246.

Davison, K., & Bagley, C. R. (1969). Schizophrenia-like psychoses associated with organic disorders of the central nervous system: a review of the literature. In R. N. Herrington (Ed.), *Current Problems in Neuropsychiatry: Schizophrenia, Epilepsy, the Temporal Lobe* (Vol. 1, pp. 113–184). London: Headley.

Dawson, D. R., Schwartz, M. L., Winocur, G., & Stuss, D. T. (2007). Return to productivity following traumatic brain injury: Cognitive, psychological, physical, spiritual, and environmental correlates. *Disability & Rehabilitation, 29*(4), 301-313.

Dearth, C. S., Berry, D. T. R., Vickery, C. D., Vagnini, V. L., Baser, R. E., Orey, S. A., et al. (2005). Detection of feigned head injury symptoms on the MMPI-2 in head injured patients and community controls. *Archives of clinical neuropsychology, 20*, 95-110.

Diaz, A. P., Schwarzbold, M. L., Thais, M. E., Hohl, A., Bertotti, M. M., Schmoeller, R., et al. (2012). Psychiatric disorders and health-related quality of life after severe traumatic brain injury: A prospective study. *Journal of Neurotrauma, 29*(1029-1037).

Dikmen, S. S., Bombardier, C. H., Machamer, J. E., Fann, J. R., & Temkin, N. R. (2004). Natural history of depression in traumatic brain injury. *Archives of physical medicine and rehabilitation, 85*, 1457-1464.

Dombrowski, L. K., Petrick, J. D., & Strauss, D. (2000). Rehabilitation treatment of sexuality issues due to acquired brain injury. *Rehabilitation Psychology, 45*(3), 299-309.

Dumais, A., Lesage, A. D., Boyer, R., Lalovic, A., Chawky, N., Menard-Buteau, C., et al. (2005). Psychiatric risk factors for motor vehicle fatalities in young men. *Canadian Journal of Psychiatry, 50*, 838 - 844.

Dunn, J. T., & Lees-Haley, P. R. (1995). The MMPI-2 correction factor for closed-head injury: a caveat for forensic cases. *Assessment, 2*(1), 47-51.

Edwards, D. W., Dahman, B. A., Wanless, R. L., Holmquist, L. A., Wicks, J. J., Davis, C., et al. (2003). Personality assessment in neuropsychology: The nonspecificity of MMPI-2 neurocorrection methods. *Assessment, 10*(3), 222 - 227

Edwards, D. W., Holmquist, L. A., Wanlass, R. L., Wicks, J. J., & Davis, C. (1998). Comparing three methods of "neuro-correction" for the MMPI–2. *Journal of International Neuropsychological Society, 4*, 27-28.

Epstein, R. S., & Ursano, R. J. (1994). Anxiety Disorders. In J. M. Silver, S. C. Yudofsky & R. E. Hales (Eds.), *Neuropsychiatry of Traumatic Brain Injury* (pp. 3-41). Washington, DC: American Psychiatric Press.

Fann, J. R., Katon, W. J., Uomoto, J. M., & Esselman, P. C. (1995). Psychiatric disorders and functional disability in outpatients with traumatic brain injuries. *American Journal of Psychiatry, 152*(10), 1493 - 1499.

Fann, J. R., Uomoto, J. M., & Katon, W. J. (2001). Cognitive improvement with treatment of depression following mild traumatic brain injury. *Psychosomatics, 42*, 48 - 54.

Fechner-Bates, S., Coyne, J. C., & Schwenk, T. L. (1994). The relationship of self-reported distress to depressive disorders and other psychopathology. *Journal of Consulting and Clinical psychology, 62*(3), 550-559.

Finn, S. E., & Kamphuis, J. H. (2006). The MMPI–2 restructured clinical (RC) scales and restraints to innovation, or "what have they done to my song?" *Journal of personality assessment, 87*(2), 202-210.

First, M. B. (2005). Mutually exclusive versus co-occurring diagnostic categories: the challenge of diagnostic comorbidity. *Psychopathology, 38*(4), 206-210.

Fishbain, D. A., Cole, B., Cutler, R. B., Lewis, J., Rosomoff, H. L., & Rosomoff, R. S. (2006). Chronic pain and the measurement of personality: do states influence traits? *Pain medicine, 7*(6), 509 - 529.

Fleming, J. M., & Strong, J. (1995). Self-awareness of deficits following acquired brain injury: Considerations for rehabilitation. *British Journal of Occupational Theory, 58*, 55-58.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299.

Folkman, S., & Lazarus, R. S. (1986). Stress processes and depressive symptomatology. *Journal of Abnormal Psychology, 95*(2), 107-113

Folzer, S. D. (2001). Psychotherapy with "mild" brain-injured patients. *American Journal of Orthopsychiatry, 71*(2), 245-251.

Forbes, D., Creamer, M., Allen, N., Elliot, P., McHugh, T., Debenham, P., & Hopwood, M. (2002). The MMPI-2 as a predictor of symptom change following treatment for posttraumatic stress disorder. *Journal of Personality Assessment, 79*(2), 321-336.

Frank, J. D. (1974). Psychotherapy: The restoration of morale. *American Journal of Psychiatry, 131* 271-274.

Franklin, C. L., Repasky, S. A., Thompson, K. E., Shelton, S. A., & Uddo, M. (2002). Differentiating overreporting and extreme distress: MMPI-2 use with compensation-seeking veterans with PTSD. *Journal of Personality Assessment, 79*(2), 274-285.

Fujii, D. (2005). Psychotic disorder due to the traumatic brain injury: review of the literature. In J. E. Pletson (Ed.), *Progress in schizophrenia research*. New York: Nova science publishers, Inc.

Fujii, D., & Ahmed, I. (2002). Characteristics of psychotic disorder due to traumatic brain injury: an analysis of case studies in the literature. *The Journal of Neuropsychiatry and Clinical Neurosciences, 14*, 130-140.

Gass, C. S. (1991). MMPI-2 interpretation and closed head injury: A correction factor *Psychological Assessment., 3*, 27 - 31.

Gass, C. S. (1992). MMPI-2 interpretation of patients with cerebrovascular disease: A correction factor. *Archives of Clinical Neuropsychology, 7*(1), 17 - 27.

Gass, C. S. (2000). Assessment of emotional functioning with the MMPI-2. In G. Groth-Marnaut (Ed.), *Neuropsychological Assessment in Clinical Practice*. New York: John Wiley and Sons Inc.

Gass, C. S. (2006). Use of the MMPI-2 in neuropsychological evaluations. In J. N. Butcher (Ed.), *MMPI-2 a practitioner's guide* (pp. 301 - 326). Washington, DC: American Psychological Association.

Gass, C. S. (2009). Use of the MMPI-2 in neuropsychological evaluations. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment*. New York: Oxford University Press.

Gass, C. S., & Lawhorn, L. (1991). Psychological adjustment following stroke: an MMPI study. *Psychological Assessment: A journal of Consulting and Clinical Psychology, 3*(4), 628 - 633.

Gass, C. S., Luis, C. A., Rayls, K. R., & Mittenberg, W. (1999). *MMPI-2 profiles in acute traumatic brain injury: Impact of demographic variables and neurological symptom reporting*. Paper presented at the 27th annual meeting of the International Neuropsychological Society.

Gass, C. S., & Russell, E. W. (1991). MMPI profiles of closed head trauma patients: impact of neurological complaints. *Journal of Clinical Psychology, 47*(2), 253 - 260.

Gass, C. S., & Wald, H. S. (1997). MMPI-2 interpretation and closed-head trauma: Cross-validation of a correction factor. *Archives of Clinical Neuropsychology, 12*(3), 199 - 205.

Glassmire, D. M., Kinney, D. I., Greene, R. L., Stolberg, R. A., Berry, D. R., & Cripe, L. (2003). Sensitivity and specificity of MMPI-2 neurologic correction factors: receiver operating characteristic analysis. *Assessment, 10*(3), 299 - 309.

Glöckner-Rist, A., & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling, 10*(4), 544-565.

Godfrey, H. P. D., Knight, R. G., & Partridge, F. M. (1996). Emotional adjustment following a traumatic brain injury: A stress-appraisal-coping formulation. *Journal of Head Trauma Rehabilitation, 11*(29-40).

Godfrey, H. P. D., Partridge, F. M., Knight, R. G., & Bishara, S. (1993). Course of insight disorder and emotional dysfunction following closed head injury: A

controlled cross-sectional follow-up study. *Journal of Clinical and Experimental Neuropsychology, 15*(4), 503-515.

Gori, A., Rosapia, L-G., Giannini, M. & Schulderberg, D. (2010). Predicting treatment outcomes by combining different assessment tools: Toward an integrative model of decision support in psychotherapy. *Journal of Psychotherapy Integration, 20*(2), 251-269.

Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W., et al. (2006). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders. *Alcohol Research and Health, 29*(2), 107-120.

Gray, J. A., & McNaughton, N. (1996). The neuropsychology of anxiety: Reprise. In D. A. Hope (Ed.), *Perspectives on anxiety, panic and fear* (pp. 61 - 134). Lincoln, NE: University of Nebraska Press.

Greene, R. L. (2000). *The MMPI-2 an interpretive manual.* (2nd ed.). Sydney: Allyn and Bacon.

Greene, R. L., Rouse, S. V., Butcher, J. N., Nichols, D. S., & Williams, C. L. (2009). The MMPI-2 restructured clinical (RC) scales and redundancy: Response to tellegen, ben-porath, and sellbom. *Journal of Personality Assessment, 91*(3), 222-226. doi:http://dx.doi.org/10.1080/00223890902800825

Greve, K. W., Bianchini, K. J., Love, J. M., Brennan, A., & Heinly, M. T. (2006). Sensitivity and specificity of MMPI-2 validity scales and indicators to malingered neurocognitive dysfunction in traumatic brain injury. *The Clinical Neuropsychologist, 20*(491 - 512).

Gualtieri, C. T., & Johnson, L. G. (1999). Traumatic brain injury: Special issues in psychiatric assessment. *Neurorehabilitation, 13*, 103 - 115.

Handel, S. F., Ovitt, L., Spiro, J. R., & Vani Rao, M. S. (2007). Affective disorder and personality change in a patient with traumatic brain injury. *Psychosomatics, 48*(1), 67 - 70.

Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: a critical review of the MMPI and MMPI-2. *Psychological Bulletin, 113*(3), 453 - 471.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416.

Hibbard, M. R., Bogdany, J., Uysal, S., Kepler, K., Silver, J., Gordon, W. A., et al. (2000). Axis II psychopathology in individuals with traumatic brain injury. *Brain Injury, 14*, 45-61.

Hibbard, M. R., Cantor, J., Charatz, H., Rosenthal, R., Ashman, T., Gundersen, N., et al. (2002). Peer support in the community: Initial findings of a mentoring program for individuals with traumatic brain injury and their families. *The Journal of Head Trauma Rehabilitation., 17*(2), 112-131.

Hibbard, M. R., Uysal, S., Kepler, K., Bogdany, J., & Silver, J. (1998). Axis I psychopathology in individuals with traumatic brain injury. *Journal of Head Trauma Rehabilitation, 13*(4), 24 - 39.

Hiott, D. W., & Labbate, L. (2002). Anxiety disorders associated with traumatic brain injuries. *Neurorehabilitation, 17*(4), 345.

Hoelzle, J. B., Nelson, N. W., & Arbisi, P. A. (2012). MMPI-2 and MMPI-2-Restructured Form validity scales: Complementary approaches to evaluate response validity. *Psychological Injury and Law, 5*(3-4), 174-191.

Hoffman, R. G., Scott, J. G., Emick, M. A., & Adams, R. L. (1999). The MMPI-2 and closed head injury: effects of litigation and head injury severity. *Journal of Forensic Neuropsychology, 1*(2), 3-13.

Holsinger, T., Steffens, D. C., Phillips, C., Helms, M. J., Havlik, R. J., Breitner, J. C., et al. (2002). Head injury in early adulthood and the lifetime risk of depression. *Archives of General Psychiatry, 59*(1), 17-22.

Howell, D. C. (2002). *Statistical methods for psychology.* (5th ed.). Belmont, CA: Thomson Wadsworth.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure anaylsis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1 - 55.

Huber, D., & Henrich, G. (2003). Personality traits and stress sensitivity in migraine patients. *Behavioural medicine, 29*(4-14).

IBM Corp. (2011). IBM SPSS Statistics for Windows (Version 20). Armonk, NY IBM Corp.

Iverson, G. L., & Lange, R. T. (2011). Moderate and severe traumatic brain injury. In M. R. Schoenberg & J. G. Scott (Eds.), *The little black book of neuropsychology a syndrome-based approach* (pp. 663-696). New York: Springer.

Johnson, S. K., DeLuca, J., & Natelson, B. H. (1996). Personality dimensions in the chronic fatigue syndrome: a comparison with multiple sclerosis and depression. *Journal of Psychiatric Research, 30*(1), 9 - 20.

Jorge, R. E., Robinson, R., G., Moser, D., Tateno, A., Crespo-Facorro, B., & Arndt, S. (2004). Major depression following traumatic brain injury. *Archives of General Psychiatry, 61,* 42 - 50.

Jorge, R. E., Robinson, R. G., & Arndt, S. (1993). Are there symptoms that are specific for depressed mood in patients with traumatic brain injury? *Journal of Nervous and Mental Disease, 181*(2), 91-99.

Jorge, R. E., Robinson, R. G., Arndt, S. V., Forrester, A. W., Geisler, F., & Starkstein, S. E. (1993). Comparison between acute and delayed-onset depression following traumatic brain injury. *Journal of Neuropsychiatry Clinical Neuroscience, 5*(1), 43-49.

Jorge, R. E., & Starkstein, S. E. (2005). Pathophysiologic aspects of major depression following traumatic brain injury. *Journal of Head Trauma Rehabilitation, 20*(6), 475-487.

Jorge, R. E., Starkstein, S. E., Arndt, S., Moser, D., Crespo-Facorro, B., & Robinson, R., G. (2005). Alcohol misuse and mood disorders following traumatic brain injury. *Archives of General Psychiatry, 62*, 742-749.

Kendall, P. C., Edinger, J., & Eberly, C. (1978). Taylor's MMPI correction factor for spinal cord injury: Empirical endorsement. *Journal of Consulting and Clinical Psychology, 46*(2), 370-371.

Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *The American Journal of Psychiatry, 159*(7), 1133-1145.

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). *The Journal of Neuropsychiatry and Clinical Neurosciences 19*(2), 106-127.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(2), 212-228.

Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York: The guilford press.

Knapp, S. J., & VandeCreek, L. D. (2006). *Practical ethics for psychologists: a positive approach.* Washington, DC: American Psychological Association.

Koponen, S., Taiminen, T., Portin, R., Himanen, L., Isoniemi, H., Heinonen, H., et al. (2002). Axis I and II psychiatric disorders after traumatic brain injury: a 30-year follow-up study. *American Journal of Psychiatry, 159*(8), 1315-1321.

Kortte, K. B., Wegener, S. T., & Chwalisz, K. (2003). Anosognosia and denial: Their relationship to coping and depression in acquired brain injury. *Rehabilitation Psychology, 48*(3), 131-136.

Kreutzer, J. S., Seel, R. T., & Gourley, E. (2001). The prevalence and symptom rates of depression after traumatic brain injury: A comprehensive examination. *Brain Injury, 15*, 563-576.

Kurtz, J. E., & Putnam, S. H. (2006). Patient-informant agreement on personality ratings and self-awareness after head injury. *The Clinical Neuropsychologist, 20*(3), 453-468.

Kurtz, J. E., Shealy, S. E., & Putnam, S. H. (2007). Another look at paradoxical severity effects in head injury with the personality assessment inventory. *Journal of Personality Assessment, 88*(1), 66-73.

La Chapelle, D. L., & Alfano, D. P. (2005). Revised neurobehavioral scales of the MMPI: sensitivity and specificity in traumatic brain injury. *Applied Neuropsychology, 12*(3), 143-150.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping.* New York: Springer.

Lillie, R. A., Kowalski, K., Patry, B. N., Sira, C., Tuokko, H., & Matier, C. A. (2010). Everyday impact of traumatic brain injury. In T. D. Marcotte & I. Grant (Eds.), *Neuropsychology of everyday functioning*. New York: The Guilford Press.

Lima, E. N., Stanley, S., Kaboski, B., Reitzel, L. R., Richey, A., Castro, Y., Williams. F. M., Tannenbaum, K. R., Stellrecht, N. E., Jakobsons, L. J., Wingate, L. R., Joiner, T. E. Jr. (2005). The incremental validity of the MMPI-2: When does

therapist access not enhance treatment outcome? *Psychological Assessment*, 17(4), 462-468.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the questions, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173.

Lowry, R. (1998-2012, 2012). VassarStats: website for statistical computation. 2011, from http://faculty.vassar.edu/lowry/clin1.html

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: some practical issues. *Psychological bulletin, 114*(3), 533 - 541.

Machulda, M. M., Berquist, T. F., Ito, V., & Chew, S. (1998). Relationship between stress, coping, and postconsuccion symptoms in a healthy adult population. *Archives of clinical neuropsychology, 13*, 415-424.

MacNiven, E., & Finlayson, M. A. (1993). The interplay between emotional and cognitive recovery after closed head injury. *Brain Injury, 7*(3), 241-246.

Malaspina, D., Goetz, R. R., Friedman, J. H., Kaufmann, C. A., Faraone, S. V., Tsuang, M., et al. (2001). Traumatic brain injury and schizophrenia in members of schizophrenia and bipolar disorder pedigrees. *The American Journal of Psychiatry, 158*(3), 440-446.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., & Trautwein, U. (2010) A new look at the big five factor structure through exploratory structural equation modelling. *Psychological Assessment, 22*(3), 471-491

Marsh, H. W., Nagengast, B., & Morin, A., J. S. (2013) Measurement invariance of the big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity and la dolce vita effects. *Developmental Psychology, 49*(6), 1194-1218

McAllister, T. W., & Flashman, L. A. (1999). Mild brain injury and mood disorders: causal connections, assessment and treatment. In N. R. Varney & R. J. Roberts (Eds.), *The evaluation and treatment of mild traumatic head injury* (pp. 347 - 374). New Jersey: Lawrence Erlbaum Associates Inc. .

McAvinue, L., O'Keefe, F., McMackin, D., & Robertson, I. H. (2005). Impaired sustained attention and error awareness in traumatic brain injury: Implications for insight. *Neuropsychological Rehabilitation, 15*(5), 569-587.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592.

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361 - 388.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525 - 543.

Meyerink, L. H., Reitan, R. M., & Selz, M. (1988). The validity of the MMPI with multiple sclerosis patients. *Journal of Clinical Psychology, 44*(5), 764-769.

Miller, L. (1996). Neuropsychology and pathophysiology of mild head injury and the post concussion syndrome: Clinical and forensic considerations. *Journal of Cognitive Rehabilitation, 14*, 8-23.

Miller, L. J., & Donders, J. (2001). Subjective symptomatology after traumatic head injury. *Brain Injury, 15*(4), 297-304.

Millsap, R. E. (1997). Invariance in measurement and prediction: Their relationship in the single-factor case. *Psychological Methods, 2*, 248-260.

Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 153-171). Mahwah, NJ: LEA.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9*, 93-115.

Millsap, R. E., & Meredith, W. (2007). Factorial invariance: historical perspectives and new problems. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100.* (pp. 131-152). Mahwah, NJ: LEA.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*(3), 479-515.

Moldover, J. E., Goldberg, K. B., & Prout, M. F. (2004). Depression after traumatic brain injury; a review of evidence for clinical heterogeneity. *Neuropsychology Review, 14*(3), 143 - 154.

Mooney, G., & Speed, J. (2001). The association between mild traumatic brain injury and psychiatric conditions. *Brain Injury, 15*(10), 865 - 877.

Moore, E. L., Terryberry-Spohr, L., & Hope, D. A. (2006). Mild traumatic brain injury and anxiety sequelae: A review of the literature. *Brain Injury, 20*(2), 117 - 132.

Morton, M. V., & Wehman, P. (1995). Psychosocial and emotional sequelae of individuals with traumatic brain injury: A literature review and recommendations. *Brain injury, 9*(1), 81-92.

Muniz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist, 12*(3), 206-219.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide.* (Sixth ed.). CA: Muthén & Muthén.

National Institute on Drug Abuse. (2010). Drugged driving. 3/10/2012, from http://www.drugabuse.gov/publications/drugfacts/drugged-driving

Nelson, L. D., Elder, J. T., Groot, J., Tehrai, P., & Grant, A. C. (2004). Personality testing and epilepsy: Comparison of two MMPI-2 correction procedures. *Epilepsy & Behaviour, 5*, 911 - 918.

Nelson, L. D., Satz, P., Mitrushina, M., Van Gorp, W., Cicchetti, D., Lewis, R., et al. (1989). Development and validation of the neuropsychology behavior and affect profile. *Psychological Assessment., 1*(4), 266 - 272.

Nichols, D. S. (2006). The trials of separating bath water from baby: A review and critique of the MMPI-2 restructured clinical scales. *Journal of Personality Assessment, 87*(2), 121-138.

Nielsen, A. S., Mortensen, P. B., O'Callaghan, E., Mors, O., & Ewald, H. (2002). Is head injury a risk factor for schizophrenia? *Schizophrenia Research, 55*(1-2), 93-98.

O'Donnell, M. L., Creamer, M. M., McFarlane, A. C., Silove, D. D., & Bryant, R. A. (2010). Should A2 be a diagnostic requirement for posttraumatic stress disorder in DSM-V? *Psychiatry Research, 176*(2-3), 257-260.

O'Shanick, G. J., & O'Shanick, A. M. (2005). Personality Disorders. In J. M. Silver, T. W. McAllister & S. C. Yudofsky (Eds.), *Textbook of traumatic brain injury* (pp. 245-258). Arlington, VA: American psychiatric publishing, Inc.

Ohberg, A., Penttila, A., & Lonnqvist, J. (1997). Driver suicides. *The Royal College of Psychiatrists, 171*(11), 468 - 472.

Ostacher, M. J. (2007). Comorbid alcohol and substance abuse dependence in depression: impact on the outcome of antidepressant treatment. *The Psychiatric Clinics of North Smerica, 30*(1), 69 - 76.

Palav, A., Ortega, A., & McCaffrel, R. J. (2001). Incremental validity of the MMPI-2 content scales: A preliminary study with brain-injured patients. *The Journal of Head Trauma Rehabilitation, 16*(3), 275 - 283

Pollack, I. W. (2005). Psychotherapy. In J. M. Silver, T. W. McAllister & S. C. Yudofsky (Eds.), *Textbook of traumatic brain injury* (pp. 641-655). Arlington, VA: American psychiatric publishing, Inc.

Pollard, C., & Kennedy, P. (2007). A longitudinal analysis of emotional impact, coping strategies and post-traumatic psychological growth following spinal cord injury: A 10-year review. *British Journal of Health Psychology, 12*, 347-362.

Pottie, C. G., & Ingram, K. M. (2008). Daily stress, coping, and well-being in parents of children with autism: A multilevel modeling approach. *Journal of Family Psychology, 22*(6), 855-864.

Prigatano, G. P. (1999). Impaired awareness, finger tapping, and rehabilitation outcome after brain injury. *Rehabilitation Psychology, 44*(2), 145-159.

Prigatano, G. P., Altman, J., & O'Brien, K. (1990). Behavioural limitations that traumatic-brain injured patients tend to underestimate. *Clinical Neuropsychologist, 4*, 163-176.

Quereshi, M., & Kleman, R. (1996). Factor analysis of MMPI-2 basic scales among college students. *Current Psychology, 15*(2), 167 - 178.

Rapoport, M. J., Kiss, A., & Feinstein, A. (2006). The impact of major depression on outcome following mild-to-moderate traumatic brain injury in older adults. *Journal of Affective Disorders, 92*, 273 - 276.

Rapoport, M. J., McCauley, S. R., Levin, H. S., Song, J., & Feinstein, A. (2002). The role of injury severity in neurobehavioural outcome. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 15*(2), 123-132.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35*, 299-331.

Rayls, K. R., Mittenberg, W., William, J., & Theroux, S. (1997). Longitudinal analysis of the MMPI-2 neurocorrection factor in mild head trauma. *Archives of Clinical Neuropsychology, 12*(4), 390 - 391.

Reitan, R. M. & Wolfson, D. (1997) Emotional disturbances and their interaction with neuropsychological deficits. *Neuropsychology Review, 7*(1), 3–19.

Rihmer, Z. (2007). Suicide risk in mood disorders. *Current Opinions in Psychiatry, 20*, 17-22.

Rogers, R., Sewell, K. S., Harrison, K. S., & Jordan, M. J. (2006). The MMPI–2 restructured clinical scales: A paradigmatic shift in scale development. *Journal of Personality Assessment, 87*(2), 139-147.

Rosenthal, M., Christensen, B. K., & Ross, T. P. (1998). Depression following traumatic brain injury. *Archives of Physical Medicine Rehabilitation, 79*, 90 - 103.

Rouse, S. V., Greene, R. L., Butcher, J. N., Nichols, D. S., & Williams, C. L. (2008). What do the MMPI-2 restructured clincial scales reliably measure? Answers from multiple research settings. *Journal of Personality Assessment, 90*(5), 435-442.

Routley, V., Staines, C., Brennan, C., Hawoth, N., & Ozanne-Smith, J. (2003). *Suicide and natural deaths in road traffic - review*. Monash University Accident Research Centre.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management, 18*, 210-222.

Scholte, W., Tiemens, B., Verheul, R., Meerman, A., & Egger, J. (2012). The RC scales predict psychotherapy outcomes: The predictive validity of the MMPI-2's restructured clinical scales for psychotherapeutic outcomes. *Personality and Mental Health, 6*(4), 292-302. doi:http://dx.doi.org/10.1002/pmh.1190

Seel, R. T., Kreutzer, J. S., Rosenthal, M., Hammond, F. M., Corrigan, J. D., & Black, K. (2003). Depression after traumatic brain injury: A national institute on disability and rehabilitation research model systems multicenter investigation. *Archives of Physical Medicine and Rehabilitation, 84*, 177-184.

Sellbom, M., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., & Graham, J. R. (2006). Elevation differences between MMPI-2 clinical and restructured clinical (RC) scales: frequency, origins, and interpretative implications. *Assessment, 16*(4), 430-441.

Senior, G., & Douglas, L. (2001). Misconceptions and misuse of the MMPI-2 in assessing personal injury claimants. *NeuroRehabilitation, 16*, 203-213.

Silver, J. M., Kramer, R., Greenwald, S., & Weissman, M. (2001). The association between head injuries and psychiatric disorders: Findings from the New Haven NIMH Epidemiologic Catchment Area Study. *Brain Injury, 15*, 935-945.

Silver, J. M., McAllister, T. W., & Yudofsky, S. C. (Eds.). (2005). *Textbook of traumatic brain injury*. Arlington, VA: American psychiatric publishing, Inc.

Simms, L. J., Cassillas, A., Clark, L. A., Watson, D., & Doebbeling, B. N. (2005). Psychometric evaluation of the restructured clinical scales of the MMPI-2. *Psychological Assessment, 17*(3), 345-358.

Smith, L., & Godfrey, H. P. D. (1995). *Family support programs and rehabilitation: a cognitive-behavioural approach to traumatic brain injury*. New York, NY: Plenum Press.

Smith, S. R., Gorske, T. T., Wiggins, C. L., & Little, J. A. (2010). Personality assessment use by clinical neuropsychologists. *International Journal of Testing, 10*(1), 6-20.

Spitzer, R. L., & Frances, A. (2011). Psychological warfare: Robert L. Spitzer and Allen Frances criticize the DSM-5 process. *Psicoterapia e Scienze Umane, 45*(2), 247-262.

Steinberg, L., & Thissen, D. (1995). Item response theory in personality research. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: a festschrift honoring Donald W. Fiske* (pp. 161-181). Hilldale, NJ: Erlbaum.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5*(1), 1-25.

Strom, T. Q., & Kosciulek, J. (2007). Stress, appraisal and coping following mild traumatic brain injury. *Brain Injury, 21*(11), 1137-1145.

Tarescavage, A, M., Wygant, D. B., Gervais, R. O., & Ben-Porath, Y. S. (2013) Association between the MMPI-2 Restructured Form (MMPI-2-RF) and malingered neurocognitive dysfunction among non-head injury disability claimants. *The Clinical Neuropsychologist, 27*(2), 313-335.

Taylor, G. P. (1970). Moderator-variable effects on personality-test-item endorsements of physically disabled patient. *Journal of Consulting and Clinical Psychology, 35*, 183-188.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). The MMPI-2 restructured clinical scales: Development, validation, and interpretation. Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y., & Sellbom, M. (2009). Construct validity of the MMPI-2 restructured clinical (RC) scales: Reply to rouse, greene, butcher, nichols, and williams. *Journal of Personality Assessment, 91*(3), 211.

Tellegen, A., Ben-Porath, Y. S., Sellbom, M., Arbisi, P. A., McNulty, J. L., & Graham, J. R. (2006). Further evidence on the validity of the MMPI–2 restructured clinical (RC) scales: Addressing questions raised by Rogers, Sewell, Harrison, and Jordan and Nichols. *Journal of Personality Assessment, 87*(2), 148-171.

Thomas, M. L., & Youngjohn, J. R. (2009) Let's not get hysterical: Comparing the MMPI-2 validity, clinical, and RC Scales in TBI litigants tested for effort. *The Clinical Neuropsychologist, 23*(6), 1067-1084, DOI: 10.1080/13854040902795000

Thomson, B. (2004). Internal replicability analyses. In *Exploratory and confirmatory factor analysis: understanding concepts and applications.* (pp. 99 - 108). Washington, DC: American Psychological Association.

Ursano, R. J., Fullerton, C. S., Epstein, R. S., Crowley, B., Kao, T., Vance, K., et al. (1999). Acute and chronic posttraumatic stress disorder in motor vehicle accident victims. *American Journal of Psychiatry, 156*, 589-595.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4 - 69.

Vosvick, M., Martin, L. A., Smith, N. G., & Jenkins, S. R. (2010). Gender differences in HIV-related coping and depression. *AIDS behaviour, 14*, 390–400.

Walsh, K., Fortier, M. A., & DiLillo, D. (2010). Adult coping with childhood sexual abuse: A theoretical and empirical review. *Aggression and Violent Behaviour, 15*, 1-13.

Whitnall, L. M., Mcmillan, T. M., Murray, G. D., & Teasdale, G. M. (2006). Disability in young people and adults after head injury: 5-7 year follow up of a prospective cohort study. *Journal of Neurology, Neurosurgery, and Psychiatry, 77*, 640-645.

WHO. (2006). *Neurological disorders public health challenges*. Switzerland: World Health Organisation.

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

Williamson, D. J. G., Scott, J. G., and Adams, R. L. (1996). Traumatic brain injury. In Adams, R. L., Parsons, O. A., Culbertson, J. L., & Nixon, S. J. (Eds.). *Neuropsychology for clinical practice. Etiology, assessment, and treatment of common neurological disorders.* (pp. 9-64) Washington, DC: American Psychological Association.

Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A monte carlo study. *Structural Equation Modeling*, *14*(3), 435-463.

Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head Injury and the MMPI-2: Paradoxical severity effects and the influence of litigation. *Psychological Assessment.*, *9*(3), 177 - 184

Youngjohn, J. R., Wershba, R., Stevenson, M., Sturgeon, J., & Thomas, M. L. (2011). Independent validation of the MMPI-2-RF somatic/cognitive and validity scales in TBI litigants tested for effort. *The Clinical Neuropsychologist*, *25*(3), 463-476.

Zasler, N. R., Martelli, M. F., & Nicholson, K. (2005). Chronic Pain. In J. M. Silver, T. W. McAllister & S. C. Yudofsky (Eds.), *Textbook of traumatic brain injury* (pp. 419-436). Arlington, VA: American psychiatric publishing, Inc.

Appendix 1 – Mplus syntax for selected Hs1 candidate factor models

Norm A

```
F1 by 141,10,45,47,143,152,164,173,175,179,208,224,249;
F2 BY 111,2,18,28,59;
F3 BY 101,3,8,20,39,57,97,176;
F4 BY 53,91,117,149,247,255;

28 with 111;
28 with 59;
45 with 141;
173 with 249;
```

Norm B

```
F1 by 3,2,8,39;
F2 BY 141,10,45,47,152,175,224,249;
F3 BY 111,20,28,59;
F4 BY 53,18,91,117,143,149,164,173,179,208,247,255;
F5 BY 101,57,97,176;

45 with 141;
173 with 249;
```

TBI

```
F1 by 59,18,28,111;
F2 BY 101,57,97,149,176,224;
F3 BY 3,2,10,39,141,152,173,249;
F4 BY 175,8,20,45,47,53,91,117,143,164,179,208,247,255;

39 WITH I3;
```

Appendix 2 – Mplus syntax for selected Hy3 candidate factor models

Norm A

```
F1 by 95,9,65,125,148;
F2 BY 81,7,26,58,76,110,124,129,151,213,241;
F3 BY 40,11,31,44,116,135,172,193,218,253,263;
F4 BY 167,161,185,230,243,265;
F5 BY 157,14,29,98,115,159,166;

161 with 185;
81 with 110;
44 with 159;
7 with 14;
29 with 213;
167 with 265;
```

Norm B

```
F1 by 65,9,11,31,95,125,148,159,166,230;
F2 BY 81,7,26,44,58,76,110,124,135,172,193,241,253,263;
F3 BY 116,14,29,40,98,115,129,151,157,213,218;
F4 BY 185,161,167,243,265;

161 with 185;
81 with 110;
7 with 14;
29 with 213;
167 with 265;
58 with 81;
58 with 110;
243 with 167;
243 with 185;
```

TBI

```
F1 by 95,9,65,125,148,157,263;
F2 BY 135,14,26,58,76,81,98,110,124,151,193,241,253;
F3 BY 172,11,31,40,44,159,166;
F4 BY 213,7,29,115,116,218,230;
F5 BY 167,129,161,185,243,265;

161 with 185;
81 with 110;
7 with 14;
```

Appendix 3 – Mplus syntax for selected Sc8 candidate factor models

Norm A

```
F1 by 138,17,23,42,48,92,145,170,177,190,221,
   234,268,274,278,292,303,307,311,316,319,320,
   322,333;
F2 BY 6,90,192,276;
F3 BY 85,16,21,32,34,35,242,256,287,323,332;
F4 BY 281,12,22,46,180,210,252,277,280,291,
   329,343;
F5 BY 168,106,182,229,295,296,298,355;
F6 BY 325,38,147,165,233,273,279,290,299;

  168 with 229;
  299 with 325;
  6 with 90;
  221 with 287;
  32 with 316;
  16 with 316;
```

Norm B

```
F1 by 145,23,38,42,48,138,168,170,177,180,182,
   210,229,234,268,274,292,295,296,298,299,307,
   311,319,322,329,333,355;
F2 BY 276,6,90,192,343;
F3 BY 316,16,32,34,35,85,106,221,242,287,290,
   320,323,332;
F4 BY 277,12,17,21,22,46,92,190,256,273,278,
   279,280,281,291,303;
F5 BY 325,147,165,233,252;

  299 with 325;
  6 with 90;
  221 with 287;
  32 with 316;
  233 with 38;
  276 with 192;
```

TBI

```
F1 by 22,12,16,17,21,23,32,35,38,42,46,48,85,92,
   106,145,147,165,168,170,177,180,182,190,221,229,
   233,234,252,256,268,273,274,277,278,279,280,281,
   290,291,292,295,296,298,299,303,307,311,316,319,
   320,322,323,325,329,332,333,355;

F2 BY 90,6,34,138,192,210,242,276,287,343;

   299 with 325;
   6 with 90;
   221 with 287;
   16 with 316;
   276 with 192;
```