Effort Test Performance in Non-litigating Brain Injury Populations

Natalie Elizabeth Hampson

Submitted in accordance with the requirements for the degree of Doctor of Clinical Psychology (D. Clin. Psychol.)

The University of Leeds
Academic Unit of Psychiatry and Behavioural Sciences
School of Medicine

July 2009

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

Dedicated to the memory of my mum, Jess Hampson, who always believed in me.

Special thanks to:

- My amazing supervisors, Dr. Steven Kemp, Dr. Chris Moulin and Dr. Tony Coughlan for all of their support, advice and encouragement throughout the completion of this thesis.
- Ms. Julianne Kinch, Dr. Jaswinder Harrison, and the staff on Ward 1 for their guidance and help with recruitment.
- All of the people who kindly volunteered to participate in the research.
- My family and friends for their never ending support; particularly my Dad, my beautiful niece Aleisha, and Katherine and Lisa for listening to me moan and never complaining.
- All of my friends on the Clinical Psychology course, particularly my housemate Clare, whose commitment to tea breaks almost rivals my own.

Abstract

Over recent years there has been increasing interest in personal gain as a threat to the validity of neuropsychological testing. As a result, clinicians have begun to use specialist measures in an attempt to identify how much 'effort' a person is putting in and therefore whether they may be feigning or exaggerating their difficulties. Such effort tests have become commonplace within a medico-legal context and their use is also increasing within clinical settings. However, thorough investigation of the performance and classification accuracy of such measures is limited and questions have been raised regarding the rates of false positives in clinical populations.

Assessing how people with genuine injuries perform on effort tests is critical for the valid interpretation of test scores, as clinicians have a duty of care not to label people as having a brain injury if they are malingering, or diagnosing someone as malingering when they have a genuine brain injury or are legitimately unwell. Therefore, the current thesis investigated the base rates of failure on a number of effort tests in a genuinely brain-injured population with no identifiable incentives to feign in order to provide further evaluation of the measures. The main focus was on the Word Memory Test, as the author claims that this measure is "virtually insensitive to all but the most extreme forms of impairment of learning and memory" (Green, Lees-Haley & Allen, 2002, p. 99).

A total of 47 participants were recruited to the study, including 20 people in residential community rehabilitation services, 16 outpatients with intractable epilepsy, and 11 people in post-acute inpatient rehabilitation. Each participant was administered a battery of tests, including measures of effort, pre-morbid IQ, memory, speed of processing, and mood.

Analyses of pass/fail rates across effort tests indicated that the rates of false positives within a genuine clinical sample with no incentive to feign were much higher than those proposed within the validation research of the tests. In addition, further statistical analyses identified a number of factors that contributed to scores on tests in addition to effort. Relationships with these factors varied depending on the particular effort test being assessed, with significant associations being identified with memory, depression, processing speed, age and participant subgroup.

These findings are consistent with recent research that suggests people with genuine brain injuries can fail effort tests for reasons related to ability rather than effort. The implications for clinical assessment and intervention are discussed, and potential future research is suggested.

Table of Contents

List of Tables

List of Figures

Abbreviations

AMI	Autobiographical Memory Index
BPS	British Psychological Society
BTEC	Business And Technology Education Council
CARB	Computerised Assessment of Response Bias
C-I-H	Coin-In-Hand Test
CMTF	Camden Memory Test for Faces
CT	Computed Tomography
DS-C	Digit Symbol-Coding Test
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
EEG	Electroencephalogram
GCS	Glasgow Coma Scale
GCSE	General Certificate of Secondary Education
HADS	Hospital Anxiety and Depression Scale
IQR	Inter Quartile Range
M	Mean
MC	Mental Control Test
MND	Malingered Neurocognitive Dysfunction
NAN	National Academy of Neuropsychology
NHS	National Health Service
NVQ	National Vocational Qualification
PTA	Post-Traumatic Amnesia
S.D	Standard Deviation
SoIP	Speed of Information Processing Test
TBI	Traumatic Brain Injury
TOMM	Test of Memory Malingering
WMT	Word Memory Test
WMT-CR	Word Memory Test Consistency Rating
WMT-DR	Word Memory Test Delayed Recognition Trial
WMT-FR	Word Memory Test Free Recall
WMT-IR	Word Memory Test Immediate Recognition Trial
WMT-MC	Word Memory Test Multiple Choice
WMT-PA	Word Memory Test Paired Associates
WTAR	Wechsler Test of Adult Reading

INTRODUCTION

"We must look for consistency. Where there is a want of it we must suspect deception."

Sherlock Holmes in Sir Arthur Conan Doyle's (1927)
"The Problem of Thor Bridge"

1.1 Overview

It is well established within the neuropsychological literature that there are many possible threats to the validity of neuropsychological testing in addition to trauma and disease, such as pain, medication, age, and educational level (e.g. Heaton, Grant & Mathews, 1986; Lishman, 1997). This means that a neuropsychologist's evaluation of any one individual might not give a wholly accurate profile of cognitive functioning, and it may be difficult to disentangle current difficulties from pre-morbid dispositions. However, more importantly, over the past 15 years there has been increasing interest in the potential for personal gain as a threat to validity and the impact that effort can have on test performance (e.g. Bender & Rodgers, 2004; Bianchini, Mathias, & Greve, 2001; Lynch, 2004). This means that there may be an advantage to 'underperforming' on tests for personal gain such as financial reimbursement, and research now indicates rates of up to 40% for malingering of symptoms in mild traumatic brain injury litigation (Larrabee, 2003).

For the interpretation of neuropsychological tests to be valid the examinee has to put in their best effort and give an accurate portrayal of symptoms, otherwise clinicians risk making a Type I error and concluding that someone has been legitimately injured when this is not the case. To make a determination about the validity of neuropsychological testing and to be confident in the accuracy of results and interpretations, clinicians are increasingly including specialist measures of 'effort' within their battery of tests.

Significant advances have been made in the field of assessment and identification of people who are potentially feigning their injuries, with the measures used to specifically assess how much 'effort' a person is putting in supposedly being insensitive to cognitive ability (Green, 2003). Such tests are designed to appear complex, but are actually so simple that even people with severe impairments can pass them. Therefore, if people fail it suggests that they are not putting in their best effort rather than failure being the result of legitimate injury.

The use of tests to assess effort (commonly known as symptom validity or effort tests) has now become commonplace within a medico-legal framework, and some researchers have even suggested that effort should be assessed in all neuropsychological research (Green et al., 2002).

However, thorough investigation of the performance and classification accuracy of such measures is still limited, and there are ongoing questions regarding the procedures and measures used.

To date, the majority of the literature has understandably focussed on populations in which malingering is thought to be most prevalent, such as medico-legal settings. However, the importance of understanding the effort test performance of clinical populations with no incentives to feign is being recognised as vital for the valid interpretation of test scores, and research has begun to consider whether people with genuine injuries fail these tests at higher rates than previously thought for reasons other than effort. The limited research conducted with clinical populations has shown that false positive results are a real difficulty that measures need to overcome if they are to be viewed as valid for use (Merten, Bossink & Schmand, 2006).

The main aim of the current thesis is to investigate the base rates of failure on a number of effort tests in a clinical population with genuine injuries and no identifiable incentives to feign in order to provide further evaluation of the measures. The main focus will be rates of failure on the Word Memory Test (WMT; Green, 2003), as the author of the WMT claims that this measure is "virtually insensitive to all but the most extreme forms of impairment of learning and memory" (Green et al., 2002, p. 99).

The following section first provides information regarding the definitions, diagnostic criteria and prevalence rates of malingering, then goes on to consider the various ways that the feigning and exaggerating of symptoms has been researched. Following this, effort testing is evaluated as the main way of identifying malingering, and an overview of how such effort tests are designed and evaluated is provided. A critique of the current research is then presented, with a specific focus on the WMT, followed by the justification of the current research and methodology. Finally, the aims of the research are outlined.

1.2 Definition and Diagnosis of Malingering

When being assessed by professionals people may display particular response styles depending on the circumstances; both internal and external influences affect the reporting of information, and all individuals vary regarding how much and the type of information they disclose to others (Rogers, 2008a). In an attempt to operationalise the basic concepts and definitions of response styles Rogers (2008a) has proposed the four main categories of non-specific terms, overstated pathology, simulated adjustment and other styles. These terms will be used throughout the current research. Overstated pathology is the most relevant to the current research and will therefore be discussed initially, with the alternative response styles being outlined in section 1.2.3.

Rogers (2008a) recommends the use of three terms for discussing what he describes as overstated pathology, including 'Malingering', 'Factitious Presentations' and 'Feigning'. The DSM-IV defines malingering as "the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (American Psychiatric Association, 2000, p739). Therefore, isolated symptoms or minor exaggeration are not enough to attract such a diagnosis. The definition encompasses the faking or exaggeration of symptoms from a wide range of disorders; including cognitive disorders such as amnesia, physical disorders such as pain, and psychiatric conditions such as dissociative disorders. Under this definition, malingering could involve someone feigning a memory impairment to gain compensation, dissociation to avoid a prison sentence, or pain to obtain drugs. Currently, the most common fabricated deficits appear to be with cognitive disorders (Cercy, Schrellen & Brandt, 1997).

However, other conditions are often associated with malingering, and can be differentiated by the underlying motivation and conscious awareness of symptom production. For a diagnosis of malingering to be given, factitious presentations need to be ruled out. For example, the DSM-IV (APA, 2000) definition of factitious disorder states that although the motivation is intentional, it is differentiated by the nature of the secondary gain, which is the psychological need to assume a sick role. However, researchers have also argued that these definitions are not clear. Rogers, Jackson and Kaminski (2004) point out that being a patient usually results in a change to people's responsibilities in relation to family and work, and exaggeration has also been proposed as a 'cry for help' to ensure recognition of distress (Berry et al., 1996).

The concept of 'Feigning' has also been suggested as appropriate to describe the deliberate exaggeration or fabrication of symptoms where no assumptions are made regarding the goals. This term was introduced because of the lack of ability for standardised tests to assess the underlying motivations of performance at assessment. Therefore, according to Rogers (2008a), psychological tests cannot be used to establish malingering, but they can be used to establish feigning.

As a result of the difficulties with acceptable definitions some researchers have concluded that "malingering behaviour should not be thought of as a dichotomy, but rather as a continuum that ranges from purposeful conscious deception to involuntary unconscious psychogenic deficits" (Haines & Norris, 1995, p127). The National Academy of Neuropsychology (NAN) has recently produced a position paper that supports this assertion, stating that "Invalid responding or performance is not a dichotomous phenomenon. Examinees may vary their performance along a continuum from complete effort and honesty to a complete lack thereof." (Bush, Ruff, Troster, Barth, Koffler, Pliskin, Reynolds, & Silver, 2005, p422). In line with this, Rogers (1997; 2008a) proposes moving away from the criminological model of explaining the underlying motivations of malingering used in the DSM-IV criteria to one based on adaptation. Such a predicted utility model is based upon how useful the individual believes the response style to be and a consideration of the potential risks and benefits in any given situation.

Slick, Sherman and Iverson (1999) have proposed diagnostic criteria for malingering, where the presence of intent is inferred rather than specifically assessed. The Slick et al. (1999) guidelines for the diagnosis of 'malingered neurocognitive dysfunction' (MND) involve three distinct categories; the first is 'definite' (characterised by the identification of an external incentive and definite negative response bias, such as below-chance performance on two or more forced-choice effort tests); the second is 'probable' (including the identification of external incentives and two or more failures on effort tests or one effort test failure and one self-reported noncredible symptom or test data discrepancy, excluding below chance performance on forced choice tests); and the third is 'possible' (which includes an identified external incentive plus self-report information that is discrepant with findings). Slick et al. (1999) also state that where psychiatric, developmental or neurologic factors 'fully account' for the findings then a diagnosis of malingering should not be given. However, professionals have questioned how to decide whether such factors 'fully account' for performance on tests, and how to determine whether the person's behaviour is truly volitional (Boone, 2007). Research indicates that brain injury, learning disability and psychological conditions such as depression cannot account for poor performance on some effort tests (e.g. Boone, Lu & Herzberg, 2002; Goldberg, Back-Madruga & Boone, 2007; Tombaugh, 1996), although this data varies depending on the test in question and the clinical samples used in the research.

The National Academy of Neuropsychology position paper concludes that "When the potential for secondary gain increases the incentive for symptom exaggeration or fabrication and/or when neuropsychologists become suspicious of insufficient effort or inaccurate or incomplete reporting, neuropsychologists can, and must, utilize symptom validity tests and procedures to assist in the determination of the validity of the information and test data obtained" (Bush et al. 2005, pp425-426). As a result, if questionable response styles are observed then interpretation of performance on other tests as valid would need to be justified. The NAN guidelines also state that tests of effort with the best psychometric properties should be used, and that data from tests should be given substantially more weight than subjective indicators because of their objectivity. However, although the NAN position paper has provided some guidelines there is still no gold-standard to assess the construct of effort (Halligan, Bass & Oakley, 2003), and there is continuing debate over the terms and definitions that should be used (Boone, 2007), which has made it difficult to establish how best to evaluate this area.

Therefore, the purpose of the current thesis is to consider the applicability and generalisability of effort tests which aim to identify feigning in brain injury, and to evaluate their performance in non-litigating clinical populations. If effort tests are fit for purpose then there should be very low base rates of failure in people with no incentive to feign.

1.2.1 Prevalence rates

Problems with the definition of malingering, the difficulties identifying true malingerers and the lack of appropriate diagnostic criteria have all affected prevalence estimates. However, a number of studies have attempted to provide base rates for a variety of conditions and settings, including forensic, psychiatric, health, and cognitive assessment. Following this, it is now widely accepted that a proportion of people do exaggerate symptoms during assessment, particularly within medico-legal settings.

Prevalence estimates have often been based on clinical judgement and opinion. For example Mittenberg, Patton, Canyock, and Condit (2002) conducted a survey of the American Board of Clinical Neuropsychology, and identified probable malingering and symptom exaggeration in 29% of personal injury claimants, 30% of disability claimants, 19% of criminal cases, and 8% of general medical cases. The highest base rate was identified in mild traumatic brain injury (TBI) personal injury litigants, with reports ranging from 38.5%-41.2%, and an 8% base rate for clinical cases with no apparent external incentives. Other research also supports these findings (Slick, Tan, Strauss, & Hultch, 2004; Sharland & Gfeller, 2007). For example, Sharland and Gfeller's (2007) survey of 188 neuropsychologists found that the median estimate of deliberate exaggeration or feigning in civil litigation was 20%, which was four times higher than those not involved in litigation, although the range of estimates varied widely, ranging from 0-90% for both contexts.

In relation to population studies, a review conducted by Larrabee (2003) concluded that there was an overall base rate of 40% exaggerating or feigning symptoms in a sample of 1363 mild TBI litigants. Whilst a meta-analysis of factors moderating outcome in mild TBI also identified that people not in litigation recovered within three months of injury, whereas those in litigation continued to report symptoms or got worse over time (Belanger, Curtis, & Demery, 2005). In forensic criminal settings, Ardolf, Denney and Houston (2007) investigated base rates of malingering with 105 defendants. Based on the Slick et al. (1999) criteria, 19% were identified as putting in a valid performance, 26.7% were identified as 'possible', 32.4% were 'probable', and 21.9% were 'definite' malingers. The combined rates of probable and definite malingerers were therefore 54.3% within this forensic setting. Richman, Green and Gervais (2006) also found that 42% of 106 potential disability claimants undergoing medical examination failed effort testing. Whilst social security claims studies have also identified a combined probable/definite rate of 72.4% (Chafetz & Abrahams, 2005).

1.2.2 Illustrative case of malingering

As an example of malingering within the context of traumatic brain injury, Mittenberg and Morgan (2009) report the case of Ms. C, a 50 year old woman who sustained head trauma five years prior to assessment by a neuropsychologist. No loss of consciousness or hospitalisation was noted, Post Traumatic Amnesia was less than 24 hours, MRI brain scan was normal, and she was able to walk home after the accident. All of which indicate a mild head injury, in which symptoms do not usually persist beyond three months (Binder, Rohling & Larrabee, 1997; Schretlen & Shapiro, 2003)

Psychological testing at six months and two years post-accident was at the level of severe cognitive impairment in intelligence and memory, and Ms. C reported inability to return to work due to cognitive difficulties, very poor concentration, poor sense of direction and loss of ability to read or play music. However, all of the findings were inconsistent with the sustained level of head trauma, and even patients with profound amnesia do not forget how to play music.

The neuropsychological report six months post injury recommended that Ms. C attend cognitive rehabilitation for post-concussion syndrome and post traumatic stress disorder, but this was not taken up. A 25-point discrepancy was identified between General Memory and Attention/Concentration Indexes, but since it is not possible to remember information that is not paid attention to initially then this corresponds to a 90% or above chance that this was a malingered memory impairment (Mittenberg, Puentes, Patton, Canyock & Heilbronner, 2002).

Two years after injury Ms. C was assessed again as part of a disability benefits claim. This assessment was performed by a nurse with no qualifications in psychology. Ms. C again scored in the severe range on general IQ tests, problem solving and motor functions, which were again highly inconsistent with what would be expected in mild head injury. In addition, she scores 6/30 on the Seashore Rhythm test, in which patients must indicate if each of 30 pairs of tonal sequences are the same or different. A person who was unable to perform this test would score 50% just by guessing, so the application of the Binomial Theorem indicates a 99.9% chance that Ms. C's score of 20% correct is reflective of the fact that she knew the right answers, but deliberately chose the wrong answers. However, the nurse reported the findings as indicating global cognitive decline and recommended 12 months of rehabilitation and restriction of activities.

Additional information that questioned the veracity of Ms. C's claims was also identified in discrepancies in reported educational level. Ms. C had reported she had an Undergraduate Social Work degree in one examination and a Postgraduate Degree in Law in the other examination, neither of which were found to be the case. A review of medical records also indicated that she had filed a compensation claim ten years earlier in which the results of an orthopaedic examination suggested that Ms. C may be exaggerating her symptoms.

Further medicolegal assessment of Ms. C by a neuropsychologist indicated that she was alert and oriented to person, place and time, affect was normal, behaviour was appropriate, and she was able to provide detailed accounts of past events, all of which are inconsistent with severe head trauma and memory dysfunction. Her scores on the Word Memory Test (Green, 2003) were significantly lower than people with severe brain injury (WMT-Immediate Recognition Trial = 42.5, Delayed Recognition Trial = 45). Scores on the Portland Digit Recognition Test (PDRT; Binder, 1993) and the Test of Memory Malingering (TOMM; Tombaugh, 1996), which are similar to the Word Memory Test in construction, again showed intentional production of incorrect answers below the level expected for guessing. Her scores on the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997a) also indicated that she was lower than 96% of the population, which again was inconsistent with presentation and observation. Scores on the Wechsler Memory Scale-III (WMS-III; Wechsler, 1997b) also showed profound impairment, and people who have scores at this level require constant supervision and someone to manage their daily affairs. Ms. C again showed the intentional production of wrong answers on the Faces subtest of the WMS-III.

Overall, Ms. C presented with symptoms that were significantly disproportionate to the injury she had sustained, her symptoms did not match a known disorder, prior medical history indicated symptom exaggeration, inconsistency was noted in her autobiographical details, and her performance on tests suggested deliberate fabrication of symptoms. Ms. C's ability to recall was better than her ability to recognise, she intentionally gave the wrong answers, and fabrication was indicated across multiple independent measures. Ms. C. therefore satisfied the Slick et al. (1999) criteria for 'definite' malingering.

The above case illustrates the importance of objective measurement of effort for identifying people who may be malingering their deficits as a result of external incentives. Ms. C had received numerous assessments, taking up the limited time and resources of a number of doctors, and had received significant financial assistance through the benefits system. If this woman had not been identified as malingering then she would have been awarded a substantial amount of compensation and continued to receive help and care that she was not entitled to and which would benefit someone who was genuinely in need.

The literature as a whole suggests that malingering does occur with relative frequency among litigants and criminal defendants, and the example presented above shows that clinicians need to take steps to ensure they are attempting to detect malingering where present and rule it out when it is not (Sweet, 1999; Sweet, Condit & Nelson, 2008). However, there is very limited research regarding prevalence rates in clinical populations with no obvious incentives. Ruling out feigning when the person has a genuine injury is dependent upon the clinician knowing what the performance of genuine patients looks like on tests of effort and whether there are any alternative reasons which might explain the results. If a judgment of feigning was made when the person actually had a legitimate injury then this could have significant consequences for the individual in relation to their rehabilitation and future quality of life.

1.2.3 Additional response styles

Rogers (2008a) provided more definitions of response styles in addition to those concerned with overstated pathology. Specifically, Rogers (2008a) suggests several non-specific terms for use by practitioners to avoid the over-specification of response styles when evidence is limited. Where the clinical data is conflicting or inconclusive, Rogers (2008a) suggests the use of several terms. 'Unreliablity' is suggested where there are questions about the accuracy of information reported, but no indication about intent or the reasons underlying it. 'Nondisclosure' is used to describe a situation where the person omits to report information, again with no assumption about underlying motives, and 'Self-disclosure' is where the person chooses what information to disclose, and therefore a lack of self-disclosure does not reflect dishonesty. 'Deception' includes any consequential attempts to deceive through distortion or misrepresentation, and is often accompanied by nondisclosure, but is not evidence of malingering. Finally, 'Dissimulation' can be used to refer to anyone who deliberately attempts to misrepresent or distort psychological symptoms, but are otherwise difficult to classify as a response style.

Rogers (2008a) presents three terms used to describe response styles linked to Simulated Adjustment; including 'Defensiveness', 'Social Desirability' and 'Impression Management'. 'Defensiveness' refers to the deliberate masking and minimisation of symptoms, and can be viewed as the opposite of malingering. 'Social desirability' involves the presentation of the self in a favourable manner through the denial of negative characteristics and attribution of positive characteristics (Carsky, Selzer, Terkelson & Hurt, 1992), and 'Impression Management' refers to deliberate attempts to control others' perceptions, and the reasons can range from the projection of a desired identity to the maximisation of social outcomes (Leary & Kowalski, 1990).

Additional response styles have also been defined, but they are not as well understood. 'Irrelevant Responding' is where the individual is not engaged in the assessment and provides responses that are not necessarily linked to the questions being asked. Linked to this is 'Random Responding' where the individual responds purely on chance factors, such as ticking the third choice on every item on a long questionnaire. For 'Role Assumption' the individual may assume the role of a character in order to respond, and 'Hybrid Responding' is used to describe the use of a variety of response styles when being assessed.

Although the above definitions are useful to consider when discussing response styles during neuropsychological testing, the definitions that are most relevant to the current research involve the feigning or exaggeration of symptoms. Much research has gone into the design and evaluation of tests with the purpose of identifying feigning. However, problems have been identified with the designs of such studies, and therefore the results and conclusions that can be drawn from them. The main research designs commonly used are discussed below.

1.2.4 Research designs used to identify malingering

Four main research designs have been used to investigate malingering performance on effort tests. These include simulation designs, known-groups comparisons, differential prevalence, and bootstrapping.

Simulation designs. Simulation studies are the most common approach. They typically involve asking non-injured participants to feign cognitive deficits in order to receive a reward, and then performance is compared to compensation seeking or non-litigating patients. However, such designs have been criticised for the performance of simulators not being generalisable to real-world malingerers (Henry, 2005), that the strategies they use are not the same (Gunstad & Suhr, 2001), litigants are more informed about cognitive deficits than simulators (Strauss, Slick, Levy-Bencheton, Hunter, MacDonald, & Hultsch, 2002) and the incentives are not as large (Bauer & McCaffrey, 2006). However, they do have the advantages of internal validity due to experimental control (Inman & Berry, 2002).

Known-groups designs. Known-groups approaches compare a non-litigating clinical group with a group independently identified as probable malingerers. The internal validity of this design is weak due to lack of experimental control. However, external validity is increased because the participants and the situations they are in are usually reflective of real-world experiences. The Slick et al. (1999) criteria have been used in many recent studies, although this design is still underused in current research and there are questions about the inaccurate assignment of participants to groups (Rogers, 2008a). For example, the assignment to known groups based on the Slick criteria is dependent upon the sensitivity and specificity of the various measures used to assess dissimulation. If the base rates for passing and failing the tests are derived from limited data then assignment to categories could be spurious. In addition, due to the difficulties with previously identifying adequate groups in known-groups designs, alternative differential prevalence and bootstrapping designs have also been used in research.

Differential prevalence designs. Differential prevalence studies involve a comparison of scores between two groups, where one sample is assumed to engage in the response style more than another. For example, it is assumed that differences in the type of referral can be used to identify groups, with those in litigation being more likely to feign. However, a major problem with this type of research is the researcher does not know which patients are actually feigning, and therefore cannot be sure if the findings are false positives (Rogers, 2008a). Differential prevalence designs frequently confound litigation status with malingering, and although the majority of research based on this design does show differential rates of failing, Ross, Putnam and Adams (2006) did not find this pattern, instead showing that litigation status was not related to incomplete effort. The percentage of people scoring below cutoff for effort who were in litigation (20%) being no higher than people not in litigation (18%; $\chi^2 = .26$, $p = .61$).

Bootstrapping designs. Finally, bootstrapping studies involve using previously well validated measures to classify groups into feigners and genuine injuries. Multiple measures are used in order to maximise specificity and ensure that no genuine patients are identified as malingerers, whilst at the same time maintaining reasonable sensitivity. However, problems with this strategy include generalisability from tests based on memory to the feigning of other difficulties, and conclusions are again dependent on the nature of the groups used in the validation studies.

In summary, research into malingering is complicated by diagnostic and methodological issues. Differential prevalence designs are commonplace in effort test research, and form the basis of many validation studies, but have significant limitations that can affect the conclusions drawn. Of crucial importance when conducting any research or assessment of malingering are the identified base rates of failure in clinical samples with no incentive to feign in order to avoid the misidentification of people with legitimate injuries as malingering, rather than reliance on the differential prevalence rates of those in litigation.

1.3 Effort Testing

Although the base rates reported above vary depending on the context and sample, they provide an indication that many people may be feigning or exaggerating symptoms, suggesting the economic burden of such false claims could be substantial. Within the UK, the potential impact on limited National Health Service resources warrants continuing research into this area. In addition to the economic implications of feigning, there are also clinical considerations that must be taken into account. If feigning cannot be accurately identified then this may prevent or delay treatment for those genuinely in need. A label of malingering is also highly pejorative, and wrongful accusation or diagnosis can have a wide ranging impact on the individual and their life, including the withdrawal of needed treatment. In addition, clinicians may be liable for litigation proceedings if they get the diagnosis wrong, and may fear retribution from patients and families as well as damage to their professional reputation. All of these considerations highlight the need for accurate methods of detection and the importance of such accuracy for patients, clinicians, and society as a whole.

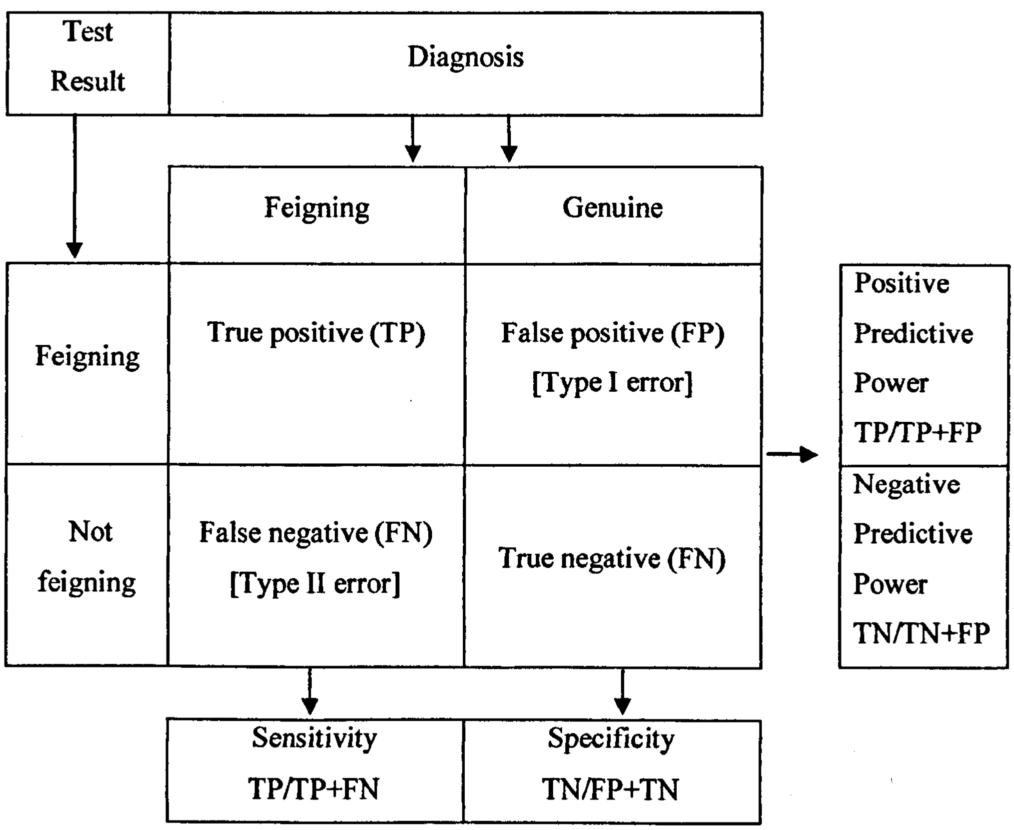
1.3.1 Evaluating effort measures

The difficulty differentiating between genuine and feigned illness has been highlighted in numerous studies. In the classic study by Rosenhan (1973) eight people feigning psychopathology were admitted to mental institutions and were not identified by medical staff as healthy. In addition, empirical research into the subjective judgements of clinicians regarding the validity of examinee symptoms has consistently found such judgements to be inaccurate (Faust, Guilmette, Hart, Artes, Fishburne & Davey, 1988) and unreliable for identifying malingering (van Gorp et al., 1999).

The ability of a test to differentiate between people with and without a particular condition is known as diagnostic validity (Smith, Cerhan, & Ivnik, 2003), and is dependent upon classification accuracy statistics, including sensitivity, specificity, and predictive power (Glaros & Kline, 1988; Grimes & Schulz, 2005). Such statistics are vital to the development of any measures designed to assess effort.

Sensitivity is the probability of a person with a particular impairment/response style being correctly identified, such as a positive test result in those who are malingering, and Specificity is the probability of a negative test result in those who are not malingering (i.e. those with genuine brain injury and no incentive to feign). Distributions of scores overlap for those with and without the condition, therefore there is no one cut-off point which perfectly separates them and a range of sensitivity and specificity values exist for various cut-offs. A review by Vickery, Berry, Inman, Harris, and Orey (2001) of the diagnostic research on malingering revealed that the average sensitivity was .56 and specificity was .96, with an average hit rate (overall diagnostic power) of .77. This is characteristic of most effort tests, with specificity set high to minimise false-positives and avoiding misidentification of people who are not malingering, but this impacts on sensitivity and increases the rate of false-negatives. Baker, Donders and Thompson (2000) suggest that false positive rates should be at or above 90% to reach acceptable specificity levels for clinical use. Figure 1 below provides a basic example of how such calculations are derived.

Figure 1: The relationship of classification accuracy statistics to rates of accurate identification across effort tests



Cut-off scores on tests are usually derived from equal samples of malingerers and TBI patients. However, this sets the base rate at 50%, which may not reflect the actual rate (Baldessarini, Finkelstein, & Arana,, 1983), so Predictive Power is needed to take this into account. Positive Predictive Power (PPP) is the probability of the presence of a condition given a positive test finding (e.g. the probability of malingering), whilst Negative Predictive Power (NPP) is the probability of the absence of a condition given a negative test finding (e.g. the probability of not malingering).

As noted above, accuracy is critical for effort tests, as the consequences of inaccurate reporting can have serious implications. All neuropsychological tests have imprecise measurement capabilities, therefore it is vital to know the accuracy of cut-off scores if appropriate decisions are to be made based on such scores. Greve and Bianchini (2004) argue that the key to improving diagnostic accuracy is improving PPP, but this requires tests to be evaluated in a variety of samples with no incentives to exaggerate or feign. This would give the clinician greater confidence that the test is not affected by other factors such as genuine neuropsychological difficulties or other clinical conditions such as depression.

Berry and Schipper (2007) suggest that there should be a reasonable convergence of sensitivity and specificity rates at the cut-off scores provided for the measure across different reports based on comparable samples. They also support the view that scores from a variety of base rate samples should be established to support the clinical application of tests. Therefore, tests require extensive validation across a wide range of clinical populations, but data of this type is currently limited within the literature, and more research is required (Larrabee & Berry, 2007).

1.3.2 Methods for detecting effort

Assessing the impact of effort is multi-faceted, and there has been a substantial increase in the variety of procedures used, with results indicating varying classification accuracy. A multitude of tests have now been developed to assess the amount of effort an individual is putting forth during neuropsychological examination. Such measures have been designed to identify fabrication or exaggeration of cognitive deficits and although the tests often appear demanding they are designed to be very simple to complete, even for people who have substantial cognitive impairments.

Using tests to investigate poor effort involves looking for atypical performance patterns when compared to people with a genuine deficit. Such atypical patterns fall into the categories of 'unlikely presentations' and 'excessive impairment' (Rogers, 2008b) and can be identified in both freestanding measures and standard neuropsychological tests.

1.3.3 Standard neuropsychological tests

Identification of unlikely presentations and excessive impairments on standard neuropsychological tests tends to include performance curves, serial position effects, floor effects, magnitude of error, atypical performance patterns, and violations of learning principles such as recognition versus recall patterns. For example, people believed to be malingering often perform much slower than people with a genuine brain injury on tasks involving response times (Cercy, Schretlen, & Brandt, 1997).

Such presentations have been identified in a wide range of standard tests (Larrabee, 2007), and their use is supported on the basis of efficiency, as neuropsychologists do not have to administer extra measures, and the tests serve other functions in addition to identifying problems with effort and so may be less recognisable to individuals as assessments of feigning. However, the effectiveness of such tests has been called into question (Rose, Hall, Szalda-Petree, & Bach, 1998) and some researchers have concluded that although such performances may indicate particular response styles, they should not be used to provide a diagnosis (Bianchini, Mathias, & Greve, 2001).

Studies have shown that failure on three or more specific standard neuropsychological tests had a very low false-positive rate (Larabee, 2003; Vickery et al., 2004). However, significant neurological damage and/or institutionalisation is reported as a modifying factor that must be taken into account when evaluating how rare multiple test failure is (Meyers and Volbrecht, 2003). Rogers (2008b) also suggests that clinicians should combine adaptations of standard tests with freestanding measures that have been specifically developed in order to thoroughly investigate feigned cognitive impairment. Some examples of standard neuropsychological tests that have been used as effort tests are included in Table 1.

1.3.4 Freestanding tests

Typically, freestanding tests used to identify excessive impairment involve two-alternative, forced-choice recognition memory tasks that require the examinee to learn and recall previously encountered items (e.g. words, digits, pictures) over variable time periods in the presence of a distractor. The Word Memory Test (WMT; Green, 2003) utilises a forced-choice word recognition paradigm. There are also tests based on other formats and that assess other areas in addition to memory, but again these measures are designed so that even people with severe impairments can pass the tests. In a meta-analysis by Vickery, Berry, Inman, Harris & Orey (2001) forced-choice tests yielded larger effect sizes than more traditional and alternative effort measures and as such are considered sensitive and are widely used. Some of the common freestanding effort measures can be found in Table 1.

Table 1: Common effort measures used to identify feigning or exaggerating of symptoms

Effort Test	Administration Format
Standard Neuropsychological Tests	
Digit Span subtest from the WAIS (Iverson & Franzen, 1994)	Digit recall
Mental Control subtest from the WMS (Kelly, Baker, van Den Broek, Jackson & Humphries, 2005)	Recall of over-learned information
Warrington Recognition Memory Test (Warrington, 1984)	Words/Face recognition
Freestanding Effort Tests: Forced Choice	
Word Memory Test (Green, 2003)	Word recognition
Test of Memory Malingering (Tombaugh, 1996)	Picture recognition
Computerised Assessment of Response Bias (Conder, Allen & Cox, 1992)	Digit recognition
Victoria Symptom Validity Test (Slick et al., 1997)	Digit recognition
Portland Digit Recognition Test (Binder, 1993)	Digit recognition
Coin-In-Hand Test (Kapur, 1994)	Location recognition
Freestanding Effort Tests: Non-forced Choice	
Rey 15-Item Test (Rey, 1964)	Letter/Number/Symbol recall and recognition
B Test (Boone, et al., 2000)	Letter recognition and discrimination
Dot Counting Test (Rey, 1941)	Visual processing
Autobiographical Memory Index (Wiggins & Brandt, 1988)	Autobiographical recall

In the Slick et al. (1999) diagnostic criteria for definite, probable and possible MND, one of the requirements for a 'definite' diagnosis is the person performs below chance on forced-choice recognition effort tests, but researchers have suggested that before such definitive guidelines can be adopted more research is needed. This is specifically due to the problems defining 'malingering', the use of forced-choice measures as the definitive assessment, and the assumption that poor effort test performance is only observed in malingerers (Boone, 2007). The NAN guidelines (Bush et al., 2005) also suggest that when there is potential for secondary gain and the clinician has concerns about effort/reporting, then they must use forced-choice tests and procedures to help determine the validity of testing. As a result, if failure on effort tests is observed then interpretation of performances on other tests as valid would need to be justified. The guidelines also state that tests with the best psychometric properties should be used, and that data from tests should be given substantially more weight than subjective indicators because of their objectivity. Therefore, forced-choice measures are increasingly being viewed as central to understanding participant responses in neuropsychological assessment. As such, an evaluation of the performance of the effort tests themselves in patients typically seen by neuropsychologists is vital for the accurate interpretation of scores in such populations.

There is an underlying assumption that people not trying their best will perform differently than those who are putting in their best effort. When presented with forced-choice tests even chance performance should be 50%. Therefore someone must actively choose the wrong answer to perform at below chance levels, indicating that other factors such as depression or extremely impaired functioning are ruled out and intentionally poor performance is the only alternative (Trueblood & Binder, 1997). However, studies have shown that although people often score poorly, they do not often perform below chance levels (e.g. Binder, 2002). In response to such findings researchers have attempted to improve the sensitivity of these tests through the development of cut-off scores based on the performance of well-documented patient groups. The scores from patients with severe brain injuries and memory disorders who are not in litigation have been used to define what the lowest credible performance could be. Consequently, developing cut-off scores in this way should ensure that people who are genuinely impaired score above the given cut-off, and so only 'below expected' performance is required as opposed to 'below chance' performance when making judgements about effort.

However, cut-off scores are criticised for increasing the probability that a person with genuine impairment will be classified as a malingerer (e.g. Haines & Norris, 1995; Suhr & Gunstad, 2000). It is critical that performance of genuine patients with no incentive to feign is evaluated in order to provide a normative standard to compare performance and draw conclusions about veracity. The purpose of the following research is to consider the applicability and generalisability of effort tests in more detail, and whether the standardised cut-off scores are applicable in samples of non-litigating participants.

It is now widely accepted that standardised tests are needed to measure effort (Iverson & Binder, 2000) and significant emphasis has been placed on well-validated freestanding forced-choice measures (Bush et al., 2005). Research utilising freestanding measures such as the Word Memory Test (Green, 2003) has shown that the amount of effort someone puts in can significantly affect neuropsychological test scores. For example, studies have identified that more people with mild brain injury fail the WMT (Green, Iverson, & Allen, 1999), and score lower on other neuropsychological tests than those with genuine brain injury who put in good effort (Green, Rohling, Lees-Haley, & Allen, 2001).

Green et al. (1999) examined 298 consecutive patients referred for assessment, 64 of which were found to have moderate to severe TBI, with documented Post Traumatic Amnesia of at least 1 day duration. However, 234 were identified as mild head injury. In a differential-prevalence design the severe TBI group performed better on the WMT than the mild TBI group.

In a study with 904 litigants, Green et al. (2001) also concluded that 50% of the variance in a neuropsychological battery was explained by effort, motivation and cooperation, whereas severity of brain injury accounted for less than 5%. In other words, effort accounted for 4.5 times more variance in symptom complaints and neuropsychological test scores than the severity of the head injury. These findings have been identified in other populations, as Gorissen, Sanz de la Torre and Schmand (2005) found that 72% of a sample of patients with schizophrenia failed the WMT and effort scores accounted for a significant proportion of the variance on neuropsychological tests. Such results are of significant importance to the field of Neuropsychology, as it implies that previous studies that have investigated brain injury and relied upon mean group scores will have been erroneous because of the lack of consideration for poor effort.

Studies utilising other effort measures have also identified similar results. For example, Constantinou, Bauer, Ashendorf, Fisher and McCaffrey (2005) found the TOMM (Tombaugh, 1996) explained 47% variance in a summary battery score with mild TBI litigators. However, some additional studies have identified superior functioning of the WMT. For example, Gervais, Rohling, Green and Ford (2004) compared WMT sensitivity rates with that of the TOMM and the CARB in 519 litigants and identified that the WMT had a failure rate of 32% compared to 17% on the CARB and 11% on the TOMM. The authors of the WMT provide this as evidence for the WMT being the most sensitive to feigned cognitive problems. However, Sweet, Condit and Nelson (2008) point out that this design most closely resembled a differential prevalence design, which is not conclusive in terms of the effectiveness of detection strategies. In addition, because the test has good sensitivity, some authors speculate that the specificity (i.e. false-positive rates) of the test could be reduced (Grote & Hook, 2007).

The authors of the WMT believe that people in litigation are often feigning their deficits. For example, recent research published by Flaro, Green and Robertson (2007) claims that failure on the WMT was 23 times higher in 774 adults with mild brain injury engaged in litigation (failure rate of 40%), compared to 118 parents seeking child custody (failure rate of 1.7%) despite similar IQ levels. Therefore, the authors state WMT failure in the brain injury sample is due to feigning and not cognitive ability, and on the basis of such results some researchers have concluded that the WMT is one of the best measures of this type (e.g. Hartman, 2002).

In a UK study, McCarter, Walton, Brooks and Powell (2009) also surveyed members of the British Neuropsychological Society, and of the 130 replies received from practicing neuropsychologists 59% reported using tests of effort in their legal work, and 15% used them in their clinical assessment. 24% of the respondents used the WMT across all assessments, with 34% specifying that they used the WMT in medicolegal settings. The McCarter et al. (2009) survey emphasises the increasing popularity of stand alone forced-choice instruments such as the WMT as measures of effort, particularly in medicolegal settings but increasingly in clinical settings. This again emphasises the importance of the accurate performance of such measures within a variety of clinical settings and populations.

1.4 Critique of Research

The WMT is one of the most popular and well investigated measures of effort currently available, with the authors stating that differences in WMT scores cannot be explained on the basis of actual cognitive deficits (Green, 2003; Green et al., 2001). Green (2003) states that "If any of the primary effort scores are 82.5% or below, there is a very high probability that other test scores from the same person significantly underestimate the person's actual abilities, owing to poor effort" (p. 9). The authors of the WMT have provided several studies that support the use of the measure (for review see Green et al. 2002), and claim that it is unique among tests of effort because of its "extensive validation *in clinical forensic settings*, rather than relying on simulation research with healthy volunteers" (Green et al., 2002, p. 97, emphasis in the original). However, whilst some researchers have supported the author's claims regarding the WMT (Hartman, 2002), it is not clear if the findings from studies such as those conducted by Green et al. (1999, 2001) are applicable in all settings and in all populations, and others have questioned the conclusions drawn from the studies.

For example, Bowden, Shores & Mathias (2006) found no difference in WMT failure rates between mild and severe brain injury in a sample of 100 children and adults in litigation, and concluded that the results "may simply indicate that all of the scores from the WMT measure memory or some component of general cognitive ability" (p. 868), and they also found high false-positive rates in those with severe brain injuries, with failure rates of 40% on WMT-IR and 18% on WMT-DR. However, Flaro et al. (2007) argue this is due to small sample size and small numbers of participants in the severe category. In addition, they also point out that if the WMT did assess memory then the expected findings should indicate that the severe group scored lower than the milder group, but no differences were found, therefore they could all potentially be putting in the same levels of effort. Despite the conclusions drawn by Flaro et al. (2007) the contrary findings of studies such as Bowden et al. (2006) indicate that further research is needed to establish the veracity of findings on the WMT.

A recent study utilising functional neuroimaging with four healthy participants (Allen, Bigler, Larsen, Goodrich-Hunsaker & Hopkins, 2007) has shown that areas known to be involved in cognitive effort, such as the dorsolateral prefrontal cortex, superior parietal lobes and anterior cingulate, are activated during WMT performance. The authors conclude that the WMT does require significant cognitive effort and is therefore susceptible to damage to any part of the neural network that subserves cognitive effort. Although all participants were 100% accurate on the delayed recognition trial of the WMT, and the research is limited by the small number of participants and lack of control task, such findings do raise questions about how people with brain injury may experience difficulties on the WMT due to cognitive effort rather than motivation when frontotemporal and limbic regions are damaged. Clinically, Batt, Shores and Chekaluk (2008) have also shown that a sample of non-litigating participants with severe brain injury struggled to perform well on the WMT when an auditory distraction task was employed during learning, with false positive rates on the WMT significantly higher than those on the TOMM under such conditions. The authors conclude that the effort components of the WMT therefore require more cognitive capacity that was originally thought, despite the claims of Green (2003) that the cut-offs are "very low and very conservative", being "selected to err on the side of caution. Using the 82.5% cut-off, there will be very few false positives indeed, even in cases of severe traumatic brain injury and neurological disease" (p. 16). In addition, Dean, Victor, Boone and Arnold (2008) found a relationship between IQ and nine different measures of effort based on standard neuropsychological tests in a sample of 189 outpatients with no external incentives. All patients with IQ<70 failed at least one effort test, with increasing failure rates as participant IQ decreased, and the specificity rates were unacceptable for most measures (i.e., <90%) when IQ dropped below 80. However, no well established freestanding forced-choice measures were used in the research.

Other researchers have also criticised the WMT on the basis of limited reliability and failure to establish the effectiveness of cut-off scores (Wynkoop & Denny, 2001). Although the specificity of effort tests has been studied with non-litigating samples, this has often been in relation to other conditions such as Schizophrenia (Gorissen, et al., 2005), or utilising other effort measures such as the Victoria Symptom Validity Test (e.g. Macciocchi, Seel, Alderson & Godsall, 2006) and malingering indexes on standard neuropsychological tests (e.g. Iverson, Slick & Franzen, 2000). Relatively few studies have provided classification accuracy data for the WMT, and what evidence exists indicates variability. Without an established gold standard effort measure it is difficult to assess validity adequately. However, classification accuracy statistics are essential for the thorough evaluation of tests intended for diagnostic use (Meehl & Rosen, 1955; Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000). As the validity of neuropsychological test data is critical for accurate interpretation of results it is important to consider how well tests of effort perform and the accuracy of the cut-off scores used to define potential feigning. In the McCarter et al. (2009) survey some of the respondents raised concerns about the reliability of effort measures. Therefore, these authors also reiterate that independent replication of validation studies is needed in order to further verify the test author's claims regarding psychometric properties and evaluate the questions raised regarding the high false positive results in clinical groups, particularly those with severe injuries (Merten, et al., 2007; Batt, et al., 2008). A paper from the British Psychological Society (BPS) professional practice board that is currently out for consultation also emphasises the requirement for "Further evidence on UK base rates of cognitive impairment and failure on effort tests in a range of clinical presentations and service settings" and that "A better understanding of the sensitivity and specificity of tests of effort is needed using UK populations" (p. 15; BPS, unpublished consultation document).

Although Green (2003) asserts that "because the cut-off scores are so much lower than the mean scores for normal controls, people with brain injuries and young children with major psychiatric illness, we would expect the false positive rate to be very low" (p. 40). Green et al. (2002) acknowledge that although their research shows examples of many people failing the WMT for reasons of insufficient effort, more formal studies are needed to support claims that certain people are unable to pass the WMT for other reasons. Very little research has been conducted to identify the base rates of failure within clinical populations of participants deemed as genuine. One such study is the validation research reported in the WMT manual (Green, 2003) with 57 moderate-severe brain injured patients (Glasgow Coma Scale score = 9.6). In this study the authors found a WMT Immediate Recognition Trial (WMT-IR) pass rate of 95.5% (S.D = 5.1), WMT Delayed Recognition (WMT-DR) pass rate of 96.1% (S.D = 3.9), and a WMT Consistency Rating (WMT-CR) of 92.8% (S.D = 6.4). 20 patients with neurological diseases and impaired memory were also assessed, and their mean WMT-IR pass rate was 96.1% (S.D = 3.3), WMT-DR = 94.9% (S.D = 5.5), and WMT-CR = 93% (S.D = 5.9), whilst 20 neurological patients with normal memory were also found to pass the WMT at similar rates; WMT-IR = 96.2% (S.D = 3.6), WMT-DR = 96.2% (S.D = 4.9), WMT-CR = 93.8% (S.D = 5.5). However, Green (2003) does not make it entirely clear whether the data from the 57 moderate to severe cases cited in the manual were actually the same patients reported in the Green and Allen (1999) study, or indeed the Green et al. (1999) study. If this is the case then these participants were also in litigation, and only assumed to be putting in good effort because their brain scans showed actual damage, and their claims for disability benefits had been accepted. A further study by Gorissen, Sanz de la Torre, and Schmand (2003) cited in the WMT manual also identified a group of neurological patients again performing similarly on the WMT-IR = 95.4% (S.D = 4.9), WMT-DR = 96.3% (S.D = 3.6), and WMT-CR = 94.4% (S.D = 4.6), although a later article apparently based on the same data does report a 10% failure rate, with 2 of the 20 participants failing (Gorissen, et al., 2005).

In two other studies with non-litigating brain injury participants conducted by researchers not directly linked to the WMT, the results are very different. For example, Merten et al. (2006) found a base rate failure of 50% (IR and CR) and 58% (DR) was identified with people with 'clinically obvious symptoms' (N = 24). Such findings indicate that people with more severe brain injuries fail the WMT at a much higher rate than those with milder brain injuries. Significant correlations were also identified between WMT scores and measures of memory and learning, a finding which has also been observed by Donders and Boonstra (2007) using the WMT and the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 2000); WMT-IR = .37, WMT-DR = .33, WMT-CR = .30, $p \leq .01$.

Merten et al. (2006) conclude that failure on effort tests does not always provide information about insufficient effort in people with clinically obvious symptoms and may just in fact reflect false positive results. Therefore, it is important to consider the clinical context for interpretation, as these participants could potentially be identified as malingering. Such research highlights the importance of thoroughly assessing effort tests and accurately identifying failure rates in a variety of populations. However, the results of the Merten et al. (2006) study require further support because the 'clinically obvious symptoms' classification was based on clinical judgement, which makes it difficult to replicate. The Merten et al. (2006) study was also conducted utilising the German version of the oral WMT and only the WMT-IR and WMT-DR subtests were used due to lack of German and Dutch norms. Similar research has yet to be conducted with a UK population.

Batt, et al. (2008) support the view that there is very limited independent research currently available regarding the sensitivity and specificity of the WMT, only citing from a sample of 64 litigating participants by Bauer, O'Bryant, Lynch, McCaffrey and Fisher (2007) and 52 student participants by Tan, Slick, Strauss and Hultsch (2002). When using the cut-offs provided in the manual, the Bauer et al. (2007) study found a sensitivity of .69 and specificity of 1.00, whilst the Tan et al. (2002) study found sensitivity and specificity rates of 1.00. High specificity was also reported by Brockhaus and Merten (2004) for the German version of the WMT (r = .97), but only participants who were judged to be 'testable' (less than half of available participants) were included (R. Brockhaus, personal communication, May 18, 2006 – reported in Dean et al., 2008), which raises questions regarding the representativeness of the sample.

In a study of 60 non-litigating participants with severe brain injuries, Batt, et al. (2008) found that 44% of participants in the standard WMT administration group failed the WMT, with specificity at 56%. 75% of participants in the distraction group failed the WMT (where learning on the WMT was accompanied by a distraction task). Those that failed were also found to have significantly lower estimated pre-morbid intelligence than those that passed (*M*s = 100, 110; *SD*

= 9, 11 respectively), $F[1,23] = 6.425$; $p=.019$. With a 30% base rate the positive predictive value was 50% for the WMT, and if this base rate was lower then the predictive values would accordingly fall even lower (Straus, Richardson, Glasziou and Haynes, 2005). Therefore, the results of Batt et al. (2008) suggest that the WMT may not be appropriate to use with some individuals, particularly where their injuries are severe, and they suggest that further research should be conducted to assess whether this extends to other people with legitimate brain injuries.

A retrospective study of the WMT-IR trial based on 132 non-litigating participants has also been performed by Bunnage, Eichinger, Pearce, Duckworth and Newson (2008). In the study Bunnage et al. (2008) attempted to identify failure rates on the WMT-IR, plus any possible external incentives that may have contributed to the rates of failure. Anything that could reasonably be considered an incentive was included, such as social security disability payments, prolonged work absence, and compensation claims. Bunnage et al. (2008) identified a failure rate of 25.8%; 9% of the sample had identifiable potential external financial incentives, and 50% of these participants failed the WMT-IR. 23% of those without any external incentive also failed the WMT-IR. Therefore, the authors identified that the conjunction of external incentive and failure on the WMT was five times less frequent than failure on the WMT alone. This indicates a potentially considerable false positive rate regarding the poor effort interpretation of WMT failure in this population, and suggesting that other variables in addition to effort/malingering may be contributing to the scores. However, this study did not investigate any relationships with other variables in addition to effort, and the retrospective nature of the study meant that not all participants were administered all elements of the WMT effort measures. Such significant failure rates in a normal clinical sample clearly warrant further evaluation and indicate that the conclusion that the WMT is "insensitive to all but the most extreme forms of cognitive impairment" (Green, 2003; p. 1) may be erroneous.

Two studies have also been conducted to investigate WMT performance in samples of epilepsy patients. Drane et al. (2006) found that 8.1% of epilepsy patients failed the WMT. However, this study excluded participants with evidence of recent seizures, those who could not live independently, and those with obvious cognitive deficits. Dodrill (2008) used a broader definition that encompassed more severe epilepsy participants, and identified a 25% failure rate on the WMT. Those who performed poorly on the WMT also did worse on the other neuropsychological tests administered, indicating that the WMT does not have a unique role in assessing test effort in this study. However, in addition to their differential failure rates, both the Drane et al. (2005) and Dodrill (2008) studies used the oral version of the WMT, and did not specifically employ other tests of effort. Therefore, these studies require further investigation to confirm failure rates on effort measures within an epilepsy sample.

In summary, effort tests have been used with increasing frequency over the past few years, but most of the research has been conducted in populations going through the litigation process (see Larrabee, 2003) or with analogue populations asked to simulate deficits (e.g. Iverson & Franzen, 1996). This has helped to establish the sensitivity of effort measures, but less attention has been paid to the specificity. The clinical application of effort tests is dependent on research conducted to establish cut-off scores, but the samples used in developing these interpretative guidelines are often heterogeneous and relatively small (Macciocchi et al., 2006). Most effort tests have not been extensively validated with non-litigating populations, severely brain injured populations, or with alternative populations such as epilepsy. Establishing the base rates of failure in populations with genuine cognitive problems and no incentive to feign can provide evidence for the robustness of such measures, as specificity may be lower than anticipated in these populations, which would increase the risk of false-positives.

1.5 Current Thesis and Justification of Methodology

As stated, currently there is an interest within the literature on forced-choice measures, particularly those that have been developed as freestanding tools. Therefore, the main focus of this research will be on the WMT. However, several other measures that have been used to assess feigning will also be included. This will provide a more rounded picture of the performance of the WMT and a clearer indication of scores that can be expected on effort tests in the clinical population being assessed, along with adding to the literature regarding the properties of the additional measures chosen. A straightforward method that specifically investigates rates of failure on effort tests in three clinical populations with no incentive to feign has been chosen. This method is novel for a number of reasons:

- Cognitive functioning is frequently affected following brain injury, but it is difficult to assess if the person genuinely has deficits, especially if there is potential for secondary gain. The use of effort testing has substantially increased, and forced-choice measures are becoming accepted as the 'gold standard' way of assessing suboptimal effort. However, recent debate has arisen regarding sensitivity and specificity (e.g. Bianchini et al., 2001). Several authors have reported that that prevalence estimates for effort test failure in clinical evaluations are extremely limited, despite the importance of such rates for the accurate interpretation of tests (Larrabee, Greiffenstein, Greve & Bianchini, 2007; Rogers 2008b). Therefore, they recommend further studies that include a wide range of non-litigating participants to investigate specificity, in order to develop a reasonable convergence of sensitivity and specificity rates at the provided cut-off scores across comparable samples.

- It remains necessary to determine whether measures of effort, particularly the WMT, accurately classify particular clinical groups, as clinicians have a duty of care not to label people as having a brain injury or psychiatric disorder if they are malingering, or diagnosing someone as malingering when they have a genuine brain injury or are legitimately unwell. Such misdiagnosis can have serious implications, including delayed treatment, the provision of inappropriate treatment, or lack of treatment altogether. Therefore, base rate information can help with diagnostic accuracy, as the more accurate the base rate estimates the more unlikely an error in classification.

- All standardised measures need independent evaluation of their psychometric properties. However, many measures of effort are lacking in this area. With regard to the WMT, most of the research has been conducted by the authors and their colleagues (e.g. Green, 2003; Green et al., 1999; Green et al., 2001). Therefore, the current study aims to provide information about the performance of effort measures in order to add to their psychometric validation.

- There is a lack of published information relating to the reasons why people may fail effort tests and the frequency of failing for genuine reasons is not known, particularly as most of the research conducted with the WMT has been with people in litigation. Such effort measures need research to independently assess performance in clinical samples of participants not in litigation in order to establish whether any people may fail effort tests for genuine reasons. No studies have looked specifically at whether there are differential failure rates in UK samples of inpatient, community-based, and epilepsy-related brain injury populations who have no incentive to feign when using effort measures such as the WMT. The current research is designed to investigate effort test performance in these distinct groups and compare their performance relative to the standard cut-offs. Although rates of sensitivity (i.e. correct classification as not malingering) can only be assumed, this research uses the strategy of assessing people with genuine brain injuries with no external incentives as a strategy for evaluating specificity rates. This cross-sectional design allows comparisons across several distinct groups, although it cannot establish whether rates may vary over time.

- In addition to identifying base rates within groups, the method allows for groups to be combined and analysed as a non-litigation group. This provides a way of investigating particular variables in addition to effort that may be contributing to effort test scores. In-depth analyses on individual data for people who fail the tests can also be conducted to develop a better understanding of why people with genuine impairments and no incentive to feign may perform poorly.

- Studies involving genuine malingerers are extremely difficult to undertake due to inherent problems involved in the identification of such people. Therefore, alternative ways of investigating performance on measures purported to assess effort are required. The current research provides a method of investigating the validity of effort tests in populations that are more easily accessible to clinicians.

- The current methodology is important for establishing meaningful effort test interpretations, and maximising confidence in the results of neuropsychological testing, diagnoses, and recommendations made on the basis of such results. Data from people with genuine brain injuries can therefore provide information about the appropriateness of current interpretive guidelines.

1.6 Research Questions

Based on the above review, the current research will investigate three central research questions. The main question concerns the identification of the base rate incidence of failure on effort tests in a genuine brain injury population with no external incentive to feign, as there is a lack of research into the performance of clinical populations on measures of effort. Two subsidiary research questions will also be investigated. Specifically, whether any factors in addition to effort contribute to effort test scores and whether different genuine brain injury populations (inpatients, community or epilepsy) fail effort tests at different rates, as it may be that some effort tests are influenced by factors other than effort at certain stages of injury, and the base rates across varying samples of participants with legitimate brain injuries have not been thoroughly researched.

METHOD

Overview

This section includes the methods employed in the current study to investigate the research questions outlined above. Specifically, this part of the thesis describes ethics, participants, inclusion and exclusion criteria, recruitment, study design, measures used, and administration procedures.

2.1 Ethics

The research was given approval by Leeds Central Ethics Committee. Managerial approval was also granted by the Research and Development Directorate of Leeds Teaching Hospitals NHS Trust. Further ethical approval was also obtained from Charity A (to protect anonymity). The Charity offers community supported housing and home support services to people at the final stage of rehabilitation following acquired brain injury. All required ethical approvals were obtained by July 2008, prior to the commencement of data collection.

2.2 Participants

2.2.1 Inclusion/exclusion criteria

All participants were required to be over 18 and have a diagnosed brain injury or epilepsy (identified through brain scans and/or brain EEG). Participants were also required to have a good grasp of the English language, and provide written informed consent.

Participants were excluded if they were in ongoing litigation or using any substances that might influence cognitive test scores (e.g. psychotropic medication or current substance misuse). The presence of comorbidity judged by the clinical team to be a significant influence (e.g. serious somatic or psychiatric disease), or visual or motor dysfunction that precluded administration of computerised tasks (e.g. Hemiparesis), were also used as the basis for exclusion from the study. In addition, potential participants were excluded if the clinical team considered that they did not have the capacity to consent, or the team judged them as not capable of participating in the study.

2.2.2 Recruitment

The recruitment of participants took place between July 2008 and May 2009. The total data collection period was 10 months.

All potential participants were identified by clinical staff involved in their care, at the neuropsychological service they were engaged with. Recruitment for all participants took place through a number of methods:

a) Information regarding the research was provided to all three recruiting centres. Therefore, some participants were referred directly by consultants and clinical team members.
b) Regular visits to a hospital ward were made to recruit suitable inpatients with a brain injury.
c) Attendance at rehabilitation team meetings established the suitability of potential participants with a brain injury currently living in the community.
d) Neurology clinics were attended regularly to recruit participants with epilepsy.

During ward and clinic visits, initial contact with potential participants was made by a member of the clinical team, who would obtain agreement for the investigator to approach them.

Inclusion and exclusion criteria were consulted, and if the potential participant was suitable they were provided with a Participant Information Sheet (see appendix 5.1). Those who were interested in participating were then contacted by the investigator, and in some cases the investigator was on hand to speak directly to those interested. A diagram to show the initial contact procedure is shown in appendix 5.3.

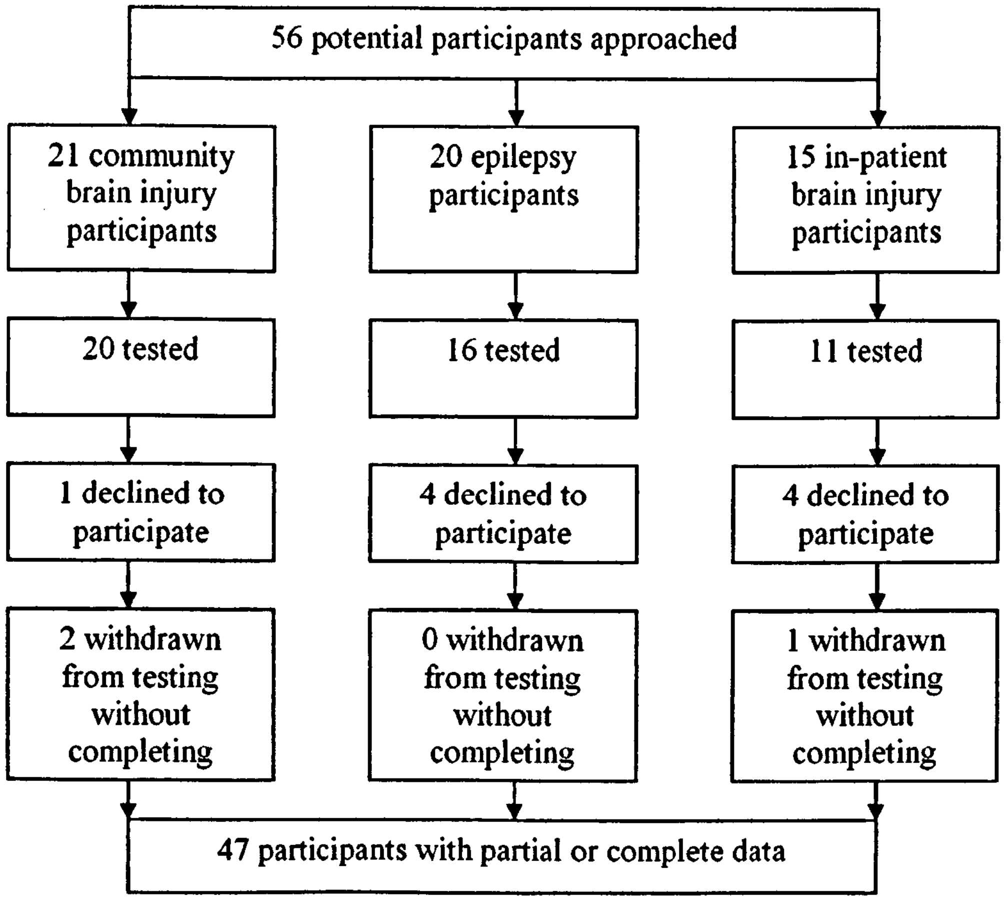
If the potential participant was willing to take part they were then contacted by the investigator, who answered any questions the participant had. Testing was then arranged at a time and place suitable to the participant. The investigator did not have contact details for potential participants. However, the participant could choose to be assessed either at the hospital or at their home. If the participant chose to be seen at home, then they, or their carers, supplied the contact details directly. No external incentive was provided to any of the participants, although reimbursement of travel expenses was offered.

Although the brain injury groups represent participants who all have a cognitive impairment, this was at a level where they were still able to provide valid consent for themselves. However, due to the nature of the participant's difficulties, the recruitment and consent process was carefully considered. Some participants needed to have the information presented to them in a user friendly way and allowed time for repeated exposure and opportunity to ask questions, so that they could gain a thorough and comprehensive understanding of what participating in the research would mean for them before they made their decision.

Consent took the form of both verbal and written consent that was checked throughout the process (see appendix 5.2 for Consent Form). Participants were provided with written information about the study that they could keep, and given the opportunity to meet directly with the researcher to discuss any issues they had. It was also made explicit that the study was for research purposes and not an intervention, and that whatever they decided about participating would not affect any services they received.

Due to the variety of methods used for recruitment, it is not possible to accurately identify everyone who was considered for participation in the study but excluded by the clinical team based on inclusion/exclusion criteria before being approached. However, a total of 56 people were directly approached to participate; eight declined to participate, and three were withdrawn before testing was completed. Figure 2 shows the number of people who were approached and participated.

Figure 2: Breakdown of the number of people participating in the study



All of the 20 community participants were assessed in their own homes, and all of the 11 inpatient participants were assessed on the rehabilitation ward they were currently staying on. Eleven of the participants with epilepsy were also assessed in their own home, whilst the other five participants with epilepsy were assessed at the hospital they attended for outpatient appointments.

2.2.3 Demographic Information

Demographic data were obtained from medical notes provided by the treating consultant, and from the participant. Table 2 provides a list of the specific demographic information collected for all participant groups.

Table 2: Demographic information collected from all three brain-injury populations

Inpatient and Community Participants	Epilepsy Participants
Date of birth	Date of birth
Gender	Gender
Education level	Education level
Handedness	Handedness
Benefits	Benefits
Date of brain injury	Age at seizure onset
Nature of brain injury	Seizure type
Lesion Location	Lesion Location

Data were collected from 47 people in total and a variety of statistical procedures were used to analyse the data. As will be shown in the results section, any strategy for analysis produced similar findings.

All participants were aged between 19 and 73 years old, with a mean age of 44.6. The participants from the Epilepsy (M = 36.1; SD: 10.2) sample were also observed to be younger than the Community (M = 46.01, SD: 12.1) or Inpatient (M = 54.1; SD: 13.5) samples. This difference was significant ($F[2, 44]=7.746$, $p=.001$), with the Epilepsy participants being significantly younger than the Community (Hochberg's GT2, $p=.048$) or Inpatient groups (Hochberg's GT2, $p=.001$).

The average number of years in education was 13.0; ranging from no qualifications to Masters Degree level. 32 men (68%) and 15 women (32%) participated in the study, 37 of which were right handed (79%) and 10 left handed (21%).

All participants in the Community and Epilepsy groups were found to be in receipt of at least some form of benefit as a result of their condition; including free prescriptions, incapacity benefits, or disability living allowance. Data were not available regarding the benefit status of the Inpatient group. Despite this, it is unlikely that these participants were receiving any benefits as a result of their current injury due to how recent their injury was.

Data were available from medical notes regarding the nature of brain injury, location of injury, and time since injury/diagnosis. As expected, the Community and Inpatient groups had more varied brain injuries than the Epilepsy group. Table 3 contains a breakdown of the nature of injuries across the three groups of participants.

Table 3: Nature and location of brain injury within the Community and Inpatient groups

Nature of Brain Injury	Community	Epilepsy	Inpatient	Total
Hypoxic	3 (15%)	-	-	3 (6.4%)
Traumatic Brain Injury	11 (55%)	-	3 (27.3%)	14 (29.8%)
Cerebrovascular accident	4 (20%)	-	6 (54.5%)	10 (21.3%)
Infection	2 (10%)	-	-	2 (4.3%)
Tumour	-	4 (25%)	2 (18.2%)	6 (12.8%)
Seizures	-	12 (75%)	-	12 (25.5%)

Location of injury varied depending on the group membership. People in the Epilepsy sample had focal lesions (mainly in the left or right temporal brain regions), whereas participants from the Community and Inpatient groups had more diffuse and anterior damage. In addition, differences were identified between the time since injury/diagnosis of the participant subgroups, with the Community participants having a mean time since injury of 11.9 years (range 1 - 30 years), the Epilepsy sample having a mean time since diagnosis of 22.9 years (range 2 - 43 years) and the Inpatient sample having a mean time since injury of 1.1 months (range 0 - 2 months). This difference was found to be significant ($F[2, 22.4]=37.812$, $p<.001$), with the Inpatient group having more recent injuries compared to the Community group (Hochberg's GT2, $p=.035$) and the Epilepsy group (Hochberg's GT2, $p<.001$).

The information from the Epilepsy sample also indicates that a number of the participants had received an epilepsy diagnosis in early childhood, which had continued throughout their adult lives.

It was anticipated that there would be some overlap in epilepsy diagnosis between participant groups, as epilepsy can be a consequence of a brain injury. Therefore, those with a secondary diagnosis of epilepsy as a result of a brain injury were included in the Community/Inpatient groups, and those with epilepsy as their primary presenting problem, or due to an in situ tumour resulting in monitoring at an epilepsy clinic, were included in the Epilepsy sample.

The information collated regarding location of brain injury was divided into three groups (Left, Right and Other). However, information regarding location varied from individual to individual, with some participants having much more detailed notes available than others. Therefore, when information within notes indicated several areas of damage, the participant was either allocated to the 'Left' or 'Right' group depending on where the majority of damage was reported to be, or to the 'Other' group if no main site of damage was specified or the main site was reported as another area (e.g. 'frontal').

It was also hoped that Post Traumatic Amnesia (PTA) scores and Glasgow Coma Scale (GCS) scores would be identified for all participants within the Community and Inpatient samples. However, PTA and GCS data could not be located for the Community group, PTA scores were not available for the Inpatient group, and GCS scores were only available for a limited number of participants from the Inpatient group (mainly those people who had experienced a traumatic brain injury).

Information regarding time since injury/diagnosis was also recorded in the current study. However, it is acknowledged that this variable is problematic due to the inherent differences between the participant groups. It is not possible to identify a 'time since injury' for the Epilepsy sample, only 'time since diagnosis', and the nature of Epilepsy and improvements in treatments means that the severity of their symptoms is likely to have varied over time.

2.2.4 Psychological/Neuropsychological data

Data was available regarding Pre-morbid IQ, Speed of Information Processing, Anxiety, Depression and Memory. The mean, standard deviation, median, inter-quartile range (IQR), and absolute range of these variables can be found in Table 4.

Table 4: Descriptive statistics for psychological and neuropsychological data

Variable	Mean	S.D.	Median	IQR	Absolute Range
Pre-morbid IQ	96.8	10.7	98.0	86.0 - 104.8	81.0 - 115.0
Processing Speed	50.8	18.4	47.0	36.0 - 64.0	19.0 - 100.0
Anxiety	6.4	4.2	6.0	3.0 - 9.0	0.0 - 20.0
Depression	4.9	4.0	4.0	2.0 - 7.0	0.0 - 20.0
WMT-MC	66.3	17.4	65.0	53.8 - 76.3	35.0 - 100.0
WMT-PA	54	21.9	52.5	38.8 - 71.3	15.0 - 90.0
WMT-FR	33.5	18.4	32.5	16.9 - 45.6	5.0 - 72.5

As Table 4 shows, from the scores that could be accurately calculated no participants were identified as within the learning disabilities range (IQ < 70) prior to injury. A large range of scores were identified on all the other variables. Processing speed ranged from below the 2nd percentile to the 82nd percentile (compared to standardised norms from the normal population).

Memory scores were also widely distributed, with descriptive statistics reflecting the increasing difficulty of the three memory tests from the WMT. This was also noted to differ across participant groups (see Table 5). However, no scores are reported for the Inpatient group, as these measures were not administered to this population.

Table 5: Differences between the Community and Epilepsy subgroups based on the WMT memory measures

Variable	Community Mean (SD)	Epilepsy Mean (SD)
WMT-MC	61.4 (18.8)	71.9 (14.2)
WMT-PA	46.1 (21.8)	63.1 (18,7)
WMT-FR	25.1 (15.3)	42.8 (17.5)

Significant differences were identified between the Epilepsy and Community subgroups on the WMT-PA (F[1,32]=5.890, p=.021) and WMT-FR (F[1,32]=9.869, p=.004), and scores approached significance for WMT-MC (F[1,32]=3.299, p=.079), revealing that the Community participants had significantly poorer memory scores than the Epilepsy participants.

Age was related to memory scores on the WMT-PA (F[1, 32]=4.500, p=.042) and WMT-FA (F[1, 32]=7.080, p=.012) in the expected direction, with those in the higher age group having lower memory scores. Time since injury/diagnosis was also linked to WMT-MC (F[1,32]=6.720, p=.014), with lower scores in the less time since injury/diagnosis group. All WMT memory measures were also related to each other at the p<.05 level.

Although the overall mean and median scores for Anxiety and Depression are below cut-off levels for caseness, some participants did score above this level (see Table 6). 23.4% of the total sample scored at cut-off or above for significant levels of anxiety, and 36.2% of the overall sample scored at cut-off or above on measures of depression. A statistically significant difference was found between people reaching the clinical cut-off for anxiety across the three subgroups (Likelihood ratio: p=.030), with more people in the Epilepsy subgroup falling above the clinical cut-off. Although Depression levels are noted to be higher in the Epilepsy subgroup, this difference was not found to be significant (Likelihood ratio: p=.558). When analysed using ANOVA the expected relationship between Anxiety (F[1,45/46]=18.696, p<.001) and

Depression was also found, with scaled Anxiety scores being higher in the depressed group, and vice versa ($F[1, 45]=13.336$, $p=.001$).

Table 6: Anxiety and Depression Scores across all participant groups

Brain Injury Group	Anxiety		Depression	
	Anxious	Not Anxious	Depressed	Not Depressed
Community	4 - 20.0%	16 - 80.0%	3 - 15.0%	17 - 85.0%
Inpatient	3 - 27.3%	8 - 72.7%	3 - 27.3%	8 - 72.7%
Epilepsy	10 - 62.5%	6 - 37.5%	5 - 31.2%	11 - 68.8%
Total	11 - 23.4%	36 - 76.6%	17 - 36.2%	30 - 63.8%

2.3 Design

This study used an exploratory design to investigate measures of effort in relation to other variables and within a range of participants. Participants across three distinct groups of non-litigating brain injury populations were each administered a number of neuropsychological, psychological, and demographic measures in addition to measures of effort. The dependent variable throughout was measures of effort, and group comparisons were made based on effort test pass and failure rates.

2.3.1 Power calculation

As mentioned in the literature review, the only studies providing information regarding the performance of genuine brain injured participants with no incentives to feign on the WMT are those conducted by Green (2003) and Merten et al. (2006), whilst only Drane et al. (2006) and Dodrill (2008) have studied epilepsy patients and reported conflicting findings.

As a result of the limited amount of research conducted using the WMT in genuine samples of brain injured and epilepsy participants, it is not appropriate to calculate statistical power. However, it is believed that 47 participants is enough to find an observed statistical difference because the WMT authors state that it is meant to be "virtually insensitive to all but the most extreme forms of impairment of learning and memory" (Green et al. 2002, p. 99), with a very narrow range of genuine scores. It is anticipated that at least some of the participants in the current study will fail even though they do not have 'the most extreme forms of learning and memory impairments'. In addition, the participant sample from Merten et al. (2006) was 48. Therefore, 47 participants were felt to represent an adequate number of participants in the absence of a calculable power analysis.

2.4 Measures

There were a number of individual tests within the main battery, but most were designed to be short, which helped to ensure participants engaged, and frustration with any particular test was minimised because they moved on quickly. The tests have also been designed specifically for, or validated for use with, people who have a brain injury, ensuring that the tests were appropriate for the participant group.

2.4.1 Effort test measures

A number of independent measures of effort were employed in the research. These measures were selected based on evidence of effectiveness, availability, and ease of administration. Table 7 provides a breakdown of the effort test cut-off scores and the validation research used to identify these cut-off levels.

Table 7: Summary of Effort measures and cut off scores used in the current study

Effort Measure	Cut-off Score	
WMT Immediate Recognition	≤ 82.5%	(Green, 2003)
WMT Delayed Recognition	≤ 82.5%	(Green, 2003)
WMT Consistency Rating	≤ 82.5%	(Green, 2003)
Coin-In-Hand Test	< 8.5	(Kelly et al. 2005)
Autobiographical Memory Index	< 10.5	(Kelly et al. 2005)
Mental Control Subtest	< 13.5	(Kelly et al. 2005)
Digit Symbol-Coding Subtest	< 5	(Inman & Berry, 2002)
Short Recognition Memory Test for Faces	≤ 17	(Kemp et al. 2008)

The Word Memory Test. (Green, 2003)

The WMT has been chosen as the main outcome measure because of its widespread use and its limited evaluation by researchers who are not directly involved with the WMT. As can be seen in the literature review, the authors claim a unique role for the WMT, and several validation studies support its use.

The WMT involves a computerised administration procedure, in which a list of 20 word pairs is presented (e.g. rat-tail). After the word pairs have been viewed there is an Immediate Recognition (WMT-IR) trial, where the participant must choose the word from the original list from 40 new word pairs (e.g. rat, from rat-shoe). After a delay of 30 minutes there is then a Delayed Recognition (WMT-DR) trial, which is similar to the WMT-IR, but includes different foil words (e.g. rat-sock). A consistency rating (WMT-CR) is then calculated on the basis of these two results. When the participant chooses the correct word it is highlighted in green and a sharp tone is heard, when the incorrect word is chosen it is highlighted in red and a flat tone is heard.

Participants are then administered other measures designed to assess memory ability; a Multiple Choice Task (WMT-MC; where participants are shown the first word from each pair and asked to choose from eight options), a Paired Associates Task (WMT-PA; where the examiner provides the first word and the participant must provide the second word), and a Free Recall Task (WMT-FR; where participants are asked to say as many of the words as possible that they remember). A Long-Delayed Free Recall task can also be administered 20 minutes after the end on the WMT-FR administration, though this was not utilised in the current research due to time constraints.

The computerised output for the WMT states whether the participant has passed or failed the WMT-IR, WMT-DR and WMT-CR effort measures based on the established cut-off score of 82.5%, which Green (2003) states is "very conservative indeed" and "there are very few "false positives" with the exception of people with dementia" (p. 9, quote in original), as scores at this cut-off represent 5.5 standard deviations (S.D) below the normal adult control group mean for WMT-IR, and 6.7 S.D below then normal mean for WMT-DR. It also represents scores at the second percentile for patients with moderate to severe brain injuries as assessed by Green and Allen (1999). One or more scores below cut-off on any of these three WMT effort measures is taken as a failure on the test.

The WMT output also provides a 'caution' if the participant scores between 82.5% and 90.0% on these effort measures, as Green et al. (2001) found that these scores were not very different from those cases who failed, and still represent two S.D below the normal adult mean and one S.D below the mean for people with severe brain injuries. Therefore, such scores are "doubtful in most cases" (Green, 2003; p. 16).

In addition, 'warnings' are also provided if scores are 70% or below for the WMT-MC, and 50% or below for the WMT-PA, and Green (2003) states that "This is to draw attention to the fact that, *if the person has dementia and is in need of supervision*, or has a similar very profound impairment, these results *might* be valid. However, such scores "are of questionable validity in most cases" (Green, 2003; p. 7).

Although Green (2003) recommends the WMT-CR as a measure of effort, it was decided that this measure would not be used in the current research due to the reliance of this calculation on the scores obtained for the WMT-IR and WMT-DR.

The Coin-In-Hand Test. (C-I-H; Kapur, 1994)

The C-I-H test involves asking participants to remember in which hand a British ten pence coin is placed. The experimenter shows the participant a coin in one hand for two seconds, and then closes both hands. Participants are then asked to count backwards from ten to one out loud with their eyes closed, then open their eyes and point to which hand the coin is in. The experimenter then opens their hands to reveal the coin. The C-I-H test is repeated for ten trials. In the current research the participants were informed that it was not a magic trick, and the coin would always be in the hand in which they had seen it. The order of administration for all participants was as follows: right, left, left, right, right, left, right, left, right, left.

This measure is proposed to be a very quick test of effort. Kapur (1994) reported that two suspected malingerers performed at chance levels on the C-I-H test, whereas five patients with neurological conditions were correct on all trials. Cochrane, Baker and Meudell (1998) found that the C-I-H identified 95% of simulating malingerers, and Kelly, Baker, van den Broek, Jackson and Humphries (2005) also report good sensitivity and specificity data when comparing simulators versus 40 patients with an acquired brain injury (sensitivity = 92.5; specificity = 87.5), and identified a cut off score of <8.5 in the detection of feigning.

Autobiographical Memory Index. (AMI; Wiggins & Brandt, 1988)

The AMI used in the current research is a slightly modified version (Kelly et al. 2005), which includes 11 questions concerning general information about the participant. The modified version omits the question 'what is your social security number', because it is not relevant to a UK population. Answers on the AMI are scored as either plausible or incorrect, with incorrect answers being clearly wrong or the participant having no knowledge of the answer.

The AMI was designed specifically to investigate malingering (Wiggins and Brandt, 1988), and Cochrane et al. (1998) found it was able to differentiate between simulators and non-simulating participants. Kelly et al. (2005) identified a cut-off of <10.5 (sensitivity 77.5; specificity 72.5) to identify feigning.

Mental Control Subtest. (MC; Wechsler Memory Scale-III; Wechsler, 1997b)

MC is an optional test from the WMS-III, and includes eight tasks involving the retrieval and manipulation of over-learned information. For example, participants are asked to count backwards from 20 to 1, and say the letters of the alphabet. A total score for all eight tests is computed.

Although not widely used as a test of effort, its utility has been investigated by Kelly et al. (2005), who propose that feigning could be identified using MC when participants take a long time to complete each subtask, feign response accuracy, or use a combination of both strategies. These researchers showed that MC has good sensitivity (80.5) and specificity (82.5) for detecting simulating participants, and when used in combination with C-I-H it is an effective screening tool for obvious malingering. Cut off for failure is taken from Kelly et al. (2005) as <13.5.

Digit Symbol-Coding Subtest. (DS-C; Wechsler Adult Intelligence Scale-III; Wechsler, 1997a)

DS-C is a subtest of the WAIS-III, in which participants must copy symbols paired with numbers. In the standardisation sample for the WAIS-III it was found that most normal individuals perform well on this task (Wechsler, 1997a), so any low scores provide an indication of neuropsychological deficits in these areas. However, a DS-C scaled score of <5 has been proposed to identify malingerers (Trueblood, 1994). This cut-off was also used by Inman and Berry (2002), resulting in specificity levels of 100% and sensitivity levels of 2%.

The Short Recognition Memory Test for Faces. (from the Camden Memory Test, CMTF; Warrington, 1996)

The CMTF measures a distinct aspect of recognition memory and is suitable for use with people with language and verbal memory difficulties. The CMTF involves the presentation of 30 pictures of faces at 3 second intervals. After the presentation of 25 faces, the participant is then asked to identify which face they have seen from a choice of two.

The Camden Memory Test has been proposed as a measure of effort due to its forced-choice recognition format and ceiling effects with people who have mild TBI (Warrington, 1996). The probability of a score occurring by chance can be calculated using Binomial Probability analysis. Cut-off for failure on the CMTF task was taken as <5th age-related centile (i.e. scores below which only 5% of sample scored). This is in line with the research of Kemp, Coughlan, Rowbottom, Wilkinson, Teggart and Baker (2008) who used the Short Pictorial Recognition Memory Test from the Camden battery as a measure of effort. The 71-89 age group was selected to be conservative, with 5% of this sample scoring 17/25 or below.

2.4.2 Supplementary measures

Wechsler Test of Adult Reading. (WTAR; Wechsler, 2001)

During administration of the WTAR, participants are asked to correctly pronounce 50 phonetically regular and irregular words. The WTAR is included to assess pre-morbid intellectual comparability across three groups of participants. The WTAR has also been co-normed with the WAIS-III and WMS-III, and therefore is the most appropriate measure of prior intelligence levels to use because it shares a common standardisation sample with some of the other measures in the study.

Speed of Information Processing Subtest. (SoIP; from the BIRT Memory and Information Processing Battery; Coughlan, Oddy, & Crawford, 2007)

During the SoIP task, participants are presented with a sheet of paper containing rows of five numbers between 10 and 99, and then asked to mark the second highest number in each row for a total duration of four minutes. After this, a test of hand speed is also administered, in which the participant must cross out numbers as quickly as possible. This test is administered in order to factor in any motor difficulties that may impact on the speed of processing findings.

Hospital Anxiety and Depression Scale. (HADS; Zigmond & Snaith, 1983)

The HADS is a self-report questionnaire measure of emotional wellbeing consisting of 14 items (seven-item depression subscale and seven-item anxiety subscale) used to screen for depression and anxiety over the past week. Questions are answered on a 0-3 scale. The measure was specifically designed for use with people who have medical conditions; as a result it reduces the likelihood of false-positives by excluding items related to somatic symptoms. Research has also indicated that the HADS performs similarly to the Beck Depression Inventory (BDI; Beck, Steer & Brown, 1996) and the General Health Questionnaire (GHQ; Bjelland, Dahl, Haug & Neckelmann, 2002). Scores of eight or above for both anxiety and depression has been deemed the most appropriate cut-off for classifying people as anxious or depressed (Bjelland et al., 2002).

Brief Effort Questionnaire.

Finally, to assess the participant's own perception of their effort, a number of reflective questions about their experience of the test session were asked. The first question was:

- As a percentage, how much effort do you think you put into these tasks?

Depending on the answer given to the first question, additional scaling questions were asked based on a motivational interviewing format. If they indicated that they put in 100% effort, then they were asked the following questions:

- Did anything ever stop you from putting in 100%
- Would anything have helped you do any better?

If the participants indicated that they did not put in 100% then they were asked the following questions:

- What stopped you from being able to put in 100%?
- What would need to happen for you to put in 100%?

Prompts were provided as necessary to allow the participants time to grasp what the questions meant. However, despite thorough explanation of the question in relation to concepts such as 'trying your best' and 'doing as well as you could', and diagrams of a Visual Analogue Scale, some of the participants found these questions difficult to understand and provided answers in relation to their brain injury/poor memory as an obstacle to performing well rather than whether they tried their best. Therefore, calculations based on these answers were not conducted as part of the results section.

2.4.3 Time taken to complete tests

A breakdown of the approximate testing time can be seen in Table 8.

Table 8: Approximate testing time for community and epilepsy participants

Measure	Time Taken to Complete (in minutes)
WTAR	2-5
HADS	5-10
WMT	15-20
C-I-H	2-4
AMI	1-2
SoIP	5
DS-C	3
MC	4-5
CMTF	5-6
Total	42-60

The acute brain injury participants were expected to experience significantly more difficulties with fatigue and attention, therefore a shorter battery was employed. Table 9 contains a breakdown of the approximate testing time for the Inpatient participants.

Table 9: Approximate testing time for inpatient participants

Measure	Time Taken to Complete (in minutes)
WTAR	2-5
HADS	5-10
WMT (IR & DR)	10-15
Total	17-30 (with a break between WMT administrations)

2.5 Procedure

At assessment the investigator went through the Information Sheet again with the participant, ensuring that they were aware of confidentiality, that their participation was voluntary, and that they were free to withdraw at any time. Written informed consent was obtained from all participants prior to their inclusion in the study, and this was administered by the investigator to ensure that the participant fully comprehended the implications of participation. All participants were assigned an identification number to protect anonymity.

Standardised administration instructions were followed and all testing was conducted individually. Participants were again instructed to put in their best effort, although they were not provided with explicit information regarding the effort testing component of the study. Each participant was then administered the battery of tests as outlined in Figures 4 and 5.

Prior to any memory tests being administered, two other tasks not involving memory were introduced to alleviate anxiety. The WMT is structured so that a half hour gap in testing is required between Immediate and Delayed conditions, where no verbal tests of memory can be administered. Therefore, the non-verbal tests and other tests that are not related to memory were presented at this point.

Figure 3: Community and Epilepsy participant test administration

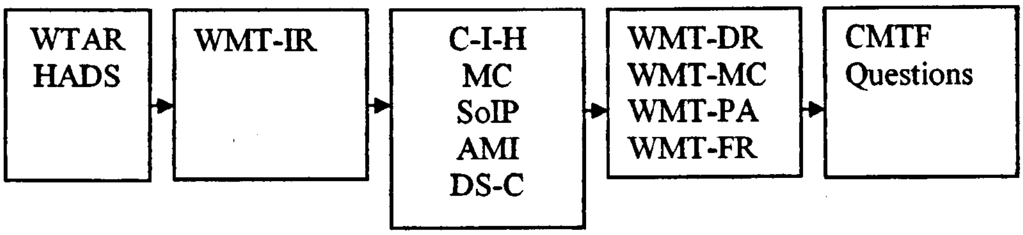
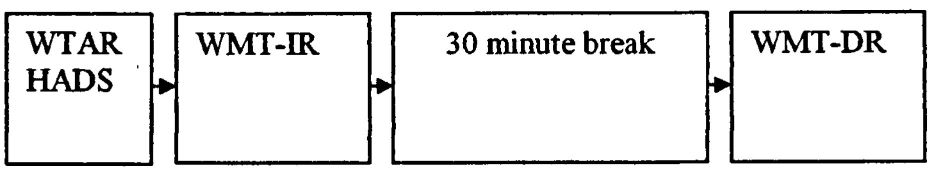


Figure 4: Inpatient participant test administration



Although a variety of procedures were employed to ensure that participants were able to complete the test battery in good time and with minimal distress, the researcher was also mindful of the difficulties in recruiting and engaging the particular participant groups involved in the study. Therefore, more time was often spent with the participant in order to build up rapport and help the participant feel at ease.

It was anticipated prior to testing that some participants would become distressed during the research for a variety of reasons, such as fatigue or frustration. Therefore, a standardised protocol for halting assessment was designed prior to the commencement of the research and this protocol was followed based on the clinical judgement of the researcher. The protocol was followed on three occasions, with two participants requiring the test administration to be stopped by the researcher, and one participant choosing for the assessment to be stopped for reasons unrelated to distress (see appendix 5.4 for protocol).

At the end of testing, all of the participants were offered the opportunity to debrief and ask any questions they had relating to the research. Where requested or required, the participants were also provided with support independent of the interviewer. Such support included informing an appropriate member of staff (with the participant's consent), or providing written information regarding local brain injury services.

RESULTS

3.1 Overview

The main aim of the current research is the investigation of the base rates of failure on a number of effort tests in three genuine populations with no external incentives to feign (inpatient, community, and epilepsy-related brain injury). This is in order to provide further evaluation of the measures and answer questions in relation to the base rate incidence of failure on effort tests in a non-litigating brain injury population, whether different non-litigating brain injury populations (inpatients, community or epilepsy) fail effort tests at different rates, and whether any other factors in addition to effort contribute to effort test scores. Therefore, the main outcome measure is the proportion of people who failed effort tests. Other explanatory variables were also administered and analysed to investigate potential reasons why people may fail the WMT.

This results section is subdivided into six main areas in order to answer the research questions described above; data preparation, description of statistical analyses, overall failure rates, group differences, relationships between tests, and individual cases.

First, a description of the data is provided, with consideration given to missing data and outliers. Second, a description of the statistical procedures used to analyse the data is presented. Third, data demonstrating overall failure rates is presented, in which all non-litigating brain-injury participants were combined into a single group to increase the statistical power of the analyses and provide a good source of diversity to the overall sample. Statistical tests were performed to identify whether any variables were related to scores on effort tests within an overall non-litigating brain injury population (Fisher's exact/Likelihood ratio tests based on categorical groupings of variables, ANOVAs based on categorical groupings and scaled scores, and Correlations between scaled scores). Fourth, differences in effort test failure rates between the three participant groups were also analysed where appropriate. Fifth, the relationships between the various tests of effort will be discussed, and finally, theoretically everyone apart from people with the most extreme impairments of memory and learning should pass the WMT based on the information put forward by the test authors. Therefore, any person that fails the WMT is illuminating and an in-depth case by case analysis is carried out to examine individuals scoring below the effort test cut-offs.

3.2. Data screening and preparation

Prior to analysis, data were screened for outliers and missing values. Demographic data was available for all participants. However, in addition to the selective battery administered to the Inpatient group, a number of participants were identified as having partial effort test, injury or neuropsychological data. For example, two participants were unable to complete particular effort measures (one person could not complete the Digit-Symbol Coding test due to manual dexterity issues, and one person could not complete the Mental Control test due to a stutter). Two participants were also withdrawn part way through assessment based on clinical judgement, and one participant voluntarily withdrew before completion of all the tasks. In addition, some WTAR scores were not computed because the discrepancy between observed and predicted scores was too great for scores to be reliable. Table 10 provides information regarding missing data and outliers.

Table 10: Total amount of participant data available per variable, and the number of outliers identified within each variable

Variable	Total number of cases available	Number of outliers identified
WMT-IR	47/47	1
WMT-DR	45/47	1
WMT-CR	45/47	1
MC	34/36	0
DS-C	33/36	1
AMI	34/36	0
CMTF	34/36	1
C-I-H	36/36	4
Effort Questions	34/36	2
Pre-morbid IQ	40/47	0
Processing Speed	34/36	0
Anxiety	47/47	0
Depression	47/47	0
WMT-MC	34/36	0
WMT-PA	34/36	2
WMT-FR	34/36	0
GCS	8/31	0
Education	47/47	3
Age	47/47	0
Time since injury/diagnosis	47/47	0

Note: Table contains information regarding the number of cases of missing data per variable, along with the number of cases identified as outliers across the demographic, injury, psychological, neuropsychological and effort variables.

Outlying scores were retained because such scores are of interest to the present research and are important to consider within the analyses. Removing outliers of this nature would reduce the external validity of the analyses, as they represent a specific part of the population distribution.

All outliers were identified using box plots, therefore, any scores that were 1.5 times or greater than the interquartile range were considered to be outlying.

3.3 Description of Statistical Analyses

Multiple approaches to data analyses were employed to investigate the relationships between effort test scores and explanatory demographic, injury, and psychological variables. Due to the design of the effort tests and the nature of the brain injury sample, it was predicted that some variables would not be normally distributed. Tests for skewness and kurtosis were applied and accepted if they fell between the standard of +/- 1 (Miller, 1984). The Kolmogrov-Smirnov test of normality was also applied if any of the skewness and kurtosis tests were marginal. The variables of years of education, months since injury, WMT-IR, C-I-H, and AMI were not normally distributed.

Some continuous variables were divided into discrete categories based on the median value of the variables (including IQ, Age, SoIP scores and WMT memory scores). Other continuous variables were divided into categories based on clinical appropriateness (e.g. Anxiety and Depression scores were split based on those who scored below and above the cut-off for caseness, and education was divided based on those who obtained GCSEs at grade C and above, and those who scored below this or did not receive any qualifications).

Effort test pass/fail rates were cross-tabulated with binomial demographic, injury, psychological and neuropsychological test scores to identify any differences between those who passed and failed effort tests. To use chi-square tests on categorical data the expected frequencies in each cell must be greater than five. However, in the current data this was not observed due to the variance in pass and fail rates across effort tests. Therefore, Fisher's exact tests or Likelihood ratios were employed. The majority of reported scores are two-sided Fisher's exact tests. However, Likelihood ratio has been reported where the brain injury group variable has been employed (for three groups).

A series of one-way ANOVAs were also performed using the scale scores from the effort variables and dichotomous demographic, psychological and neuropsychological measures, in order to provide additional information about the relationships between effort scores and the independent variables. Where Levene's test of homogeneity of variance was $p<0.05$ Welch's F statistic is reported. Post-hoc tests were performed using Hochberg's GT2 calculations. All ANOVA findings based on non-normally distributed variables were also supported by the Kruskal Wallis and Mann Whitney U test calculations.

Two-tailed Pearson's r correlations were calculated where variables were normally distributed, and Spearman's rho correlations are reported when the variable was not normally distributed. A point-biserial correlation was performed for the dichotomous variables of gender and handedness. Correlations were not performed on variables that contained small numbers of ordinal categories, which includes the participant sampling groups, and this must also be taken into account when interpreting the C-I-H and AMI variables due to the limited number of different scores. However, Jaccard and Wan (1996) highlight that, "for many statistical tests, rather severe departures (from intervalness) do not seem to affect Type I and Type II errors dramatically" (p. 4). Therefore, where judgements about the ordinal or interval level of measurement was unclear, data were assumed to be interval.

Due to the number of analyses carried out, there was also an increased risk of Type 1 error. Therefore, Bonferroni corrections were applied as appropriate. However, the findings of the current research are exploratory and preliminary in nature; therefore uncorrected significance levels are also given consideration.

All analyses were performed using SPSS computer package for Windows, Version 16.

3.3.1 Effort test data

This thesis aimed to examine the base rate incidence of effort test failure in a non-litigating brain injury population. To this end, the examination of the distribution of scores on the various effort tests and the rate at which people failed these tests forms the critical analysis for this thesis. To examine these issues, the descriptive statistics for each of the effort tests is considered (see Table 11) and the failure rates (according to test manuals and validation research) across the effort tests is described (see Table 12). To foreshadow these results, a wide range of performances were displayed by the participants; with failure rates of between 2.8% to 38.9% depending on the test.

Table 11: Descriptive statistics for all of the effort tests used in the current study

Effort Test	Mean	S.D	Median	IQR	Absolute Range
WMT-IR	88.5	12.2	90.0	85.0 - 97.5	42.5 - 100.0
WMT-DR	90.0	9.1	92.5	85.0 - 97.5	65.0 - 100.0
C-I-H	9.7	1.0	10.0	10.0 - 10.0	5.0 - 10.0
MC	19.1	7.3	19.0	12.8 - 25.5	8.0 - 36.0
DS-C	5.8	2.0	6.0	4.5 - 7.0	2.0 - 12.0
AMI	10.1	1.1	11.0	9.0 - 11.0	8.0 - 11.0
CMTF	21.2	3.7	22.0	18.8 - 24.0	12.0 - 25.0

Table 11 shows that the mean and median for all effort tests falls above the established cut-offs for failure. However, the standard deviation and range indicate large variation between these scores, particularly within the Immediate Recognition trial; where the two participants whose test administrations were terminated early obtained the two lowest scores. In addition, although above the cut-off scores provided in the manual (Green, 2003), the mean and S.D of the WMT-IR and WMT-DR are different to those reported by Green and Allen (1999) as part of the WMT validation using people with moderate to severe brain injuries. Green and Allen (1999) report that the WMT-IR mean was 95.5% (S.D = 5.1) and the WMT-DR mean was 96.1% (S.D = 3.9) in their population.

When broken down into pass and fail rates across all participants Table 12 shows that failures have occurred across all of the effort tests. Based on the assumption that all participants in the current research were genuine clinical patients with no incentive to feign their impairments, then the number of false positive results in this sample produces unacceptable specificity rates for all but the C-I-H test (i.e. <90%; Baker et al., 2000).

Table 12: Pass and fail rates of all participants on tests of effort used in the current research

Effort test	Pass	Fail	Missing
WMT-IR	37 (78.7%)	10 (21.3%)	0 (0.0%)
WMT-DR	37 (78.7%)	8 (17.0%)	2 (4.3%)
C-I-H	34 (94.4%)	1 (2.8%)	1 (2.8%)
MC	25 (69.4%)	9 (25.0%)	2 (5.6%)
DS-C	25 (69.4%)	8 (22.2%)	3 (8.3%)
AMI	20 (55.5%)	14 (38.9%)	2 (5.6%)
CMTF	28 (77.8%)	6 (16.7%)	2 (5.6%)

Although failure on the C-I-H test was uncommon (2.8% of the overall sample), failures on other tests of effort were more marked, with rates ranging from 16.7% to 38.9%. Therefore, the base rate incidence of failure within the current non-litigating clinical sample of people with brain injuries is higher than that suggested by the authors of the measures, particularly the claims put forward regarding the performance of the WMT. For example, 12 of the 47 participants failed at least one WMT effort measure (25.5%), while the WMT-IR failure rate of 21.3% and WMT-DR failure rate of 17.0% is much higher than that suggested by Green (2003), who proposes that "there are very few "false positives", with the exception of people with dementia" (p. 9, quote in original). However, as described below, this finding is also likely to be a conservative estimate of the numbers who are usually considered to be failing on the WMT.

3.3.2 WMT Consistency, Caution and Warning Data

As outlined in the Method section, in addition to providing pass and fail rates, the WMT output also provides a 'caution' rating for the Immediate Recognition and Delayed Recognition trials, as well as using the WMT Consistency Rating as a measure of effort. The WMT-MC and WMT-PA scores are also provided with a 'pass' or 'warning' label. The authors of the WMT state that they have included such labels in an attempt to draw the assessor's attention to scores which could potentially indicate feigning. A breakdown of these additional scores within the current research can be found in Table 13.

Table 13: Total numbers of 'cautions' and 'warnings' across the WMT measures split by participant subgroup

		Participant Subgroup			
		Community	**Inpatient**	**Epilepsy**	**Total**
WMT-IR					
	Fail	6	3	0	9/47
	Caution	4	2	3	9/47
	Total	10	5	3	18/47
WMT-DR					
	Fail	4	2	1	7/45
	Caution	6	1	3	10/45
	Total	10	3	4	17/45
WMT-CR					
	Fail	8	3	3	14/45
	Caution	4	4	3	11/45
	Total	12	7	6	25/45
WMT-MC					
	Warning	14	-	8	22/34
WMT-PA					
	Warning	12	-	5	17/34

Twelve of 47 participants failed either the WMT-IR or the WMT-DR (25.5%), but if the Consistency Rating was also included as a measure of effort then 19 of the 47 participants included in the current research would have failed at least one WMT effort measure. This equates to a 40.4% failure rate on the WMT across a non-litigating brain injury population.

When including the 'cautions' within the current calculations then 23 of the 47 participants obtained a caution or fail on either the WMT-IR or WMT-DR effort measures (48.9%), and if the Consistency Ratings are also added to these scores then the number of participants who have scores supposedly indicative of poor effort rises to 29 of the 47 participants. This gives a very large potential poor effort rate of 61.7% within the current research.

Within the memory measures, 14 participants score in the 'warning' range on the WMT-MC, and 12 score in the 'warning' range on the WMT-PA. According to Green (2003) scores in this range "would be very suspicious, except in someone with dementia or profound amnesia" (p. 11). However, the sample in the current research did not have dementia, and although many of the sample have memory problems it is unlikely that many would attract a classification of 'profound' as it would be considered in line with dementia.

According to Green (2003), a caution on the WMT-IR and WMT-DR is still two S.D below the normal adult mean and one S.D below the mean for people with severe brain injuries, and a warning on the memory measures is in most cases suggestive of poor effort. As these scores were treated as passes throughout the current analyses it is likely that the findings provide a conservative estimate of the performance of the WMT across the current sample of participants and of the relationships between the WMT and the other variables included in the research.

Another feature of note within the WMT data is that some participants also scored lower on the Immediate Recognition trial of the WMT than on the Delayed Recognition trial, indicating counterintuitive scores. Four participants in the Inpatient group, five participants in the Community group, and four participants in the Epilepsy group had higher scores on the Delayed trial compared to the Immediate trial.

3.3.3 Pass/fail rates on effort tests across participant subgroups

Given that a substantial amount of people in the current research failed effort tests, where the authors of such tests suggest that failure rates should be low or zero, it was of interest to see whether the groups differed in regard to base rate failure. To analyse this, the sample was divided into its separate diagnostic groups.

As can be seen in Table 14 the Community participants failed the AMI at much higher rates than the Epilepsy subgroup; none of the Epilepsy participants failed this test, whereas a substantial proportion of the Community participants did fail. This difference was statistically significant (Fishers; $p<.001$). When scaled scores were analysed this finding remained significant ($F[1, 19.02]=37.152$, $p<.001$), with Community participants significantly more likely to have lower scores compared to the Epilepsy participants. This pattern of results was also found in relation to the Camden Memory Test for Faces ($F[1, 32]=9.433$, $p=.004$), with Community participants again having significantly lower scores than the Epilepsy participants.

Although not significant, some trends in differential failure rates on effort tests were observed. Within the WMT-IR, failure rates in the Community and Inpatient samples were 30% (6/20 people) and 27.3% (3/11 people) respectively, whereas only 6.2% (1/16 people) failed in the Epilepsy group (p=.223). When analysed based on scaled scores, this non-significant trend was also observed (F[2, 44]=2.684, p=.079), with post hoc Hochberg's GT2 tests identifying the main differences as between the Community and Epilepsy samples (p=.079).

Within the WMT-DR condition, approximately 20% of people failed in the Community (4/18 people) and Inpatient (2/10 people) subgroups, compared to 12.5% (2/16) in the Epilepsy subgroup (p=.786), but scaled score calculations were not found to be significant (F[2, 44]=0.980, p=.374).

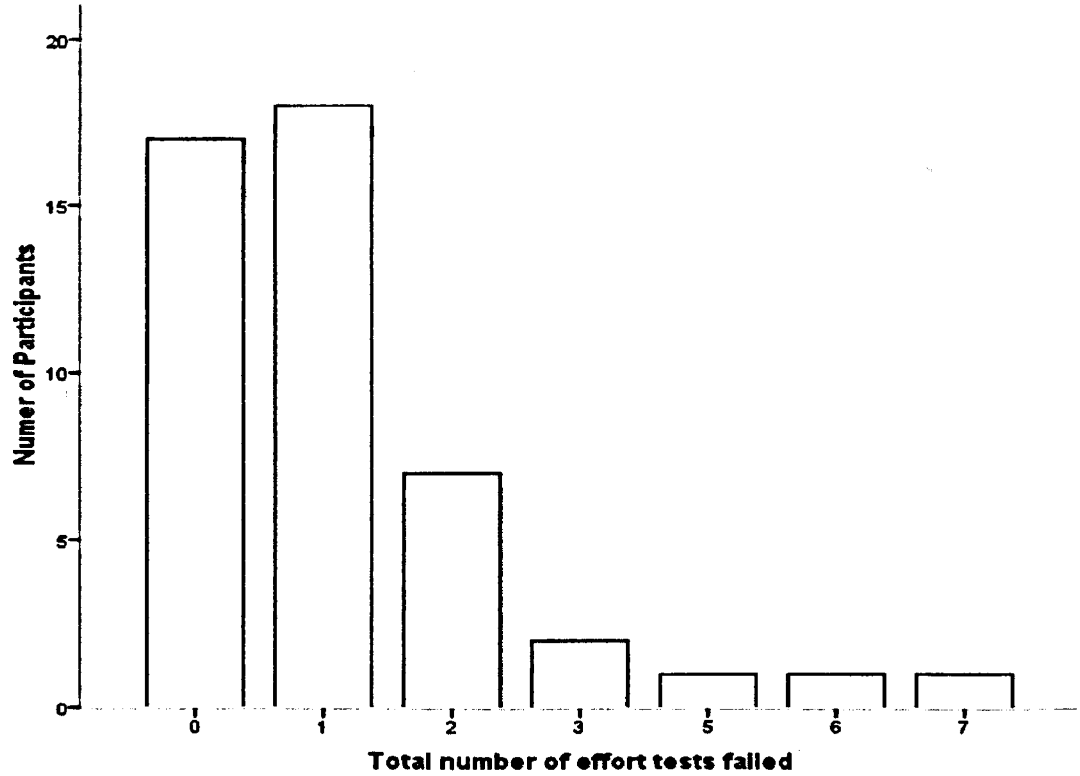
Overall failure rates on the WMT were 35% in the Community subgroup, 27.3% in the Inpatient subgroup, and 18.8% in the Epilepsy subgroup.

Table 14: Failure rates on effort tests across all participant groups

Effort Test	Community		Epilepsy		Inpatient	
	Pass	Fail	Pass	Fail	Pass	Fail
WMT-IR	14 (70%)	6 (30%)	15 (93.8%)	1 (6.2%)	8 (72.7%)	3 (27.3%)
WMT-DR	15 (78.5%)	4 (21.1%)	14 (87.5%)	2 (12.5%)	8 (72.7%)	2 (18.2%)
C-I-H	19 (95%)	1 (5%)	15 (93.8%)	1 (6.2%)	-	
MC	14 (73.7%)	5 (26.3%)	11 (73.3%)	4 (26.7%)	-	
DS-C	13 (65%)	4 (20%)	12 (75%)	4 (25%)	-	
AMI	4 (22.2%)	14 (77.8%)	16 (100%)	0 (0%)	-	
CMTF	13 (72.2%)	5 (27.8%)	15 (93.8%)	1 (6.2%)	-	

Figure 6 shows the total number of effort test failures across all participants, again highlighting that people with legitimate brain injuries and no incentive to feign do fail effort tests at rates used to identify malingering. It can be seen from Figure 6 that 30/47 participants failed at least one effort test, and five participants failed three or more of the tests, despite the fact that only two effort tests were administered to the Inpatient group and some participants did not complete all of their testing.

Figure 5: The total number of effort test fails across all participants in the current research



Based on the Slick et al. (1999) criteria for 'probable' malingering, part of identifying people in this category involves two or more types of evidence from neuropsychological tests, including where "well-validated psychometric tests or indices designed to measure exaggeration or fabrication of cognitive deficits is consistent with feigning" (p. 553), and where the test data are discrepant with known patterns of neuropsychological functioning. Therefore, at least 12 of the current participants would potentially have met the criteria for probable malingering if they had an identifiable external incentive, and doubts would be raised regarding 63.8% of the total sample with regard to effort.

3.3.4 Relationships between effort tests and explanatory variables

As identified in Table 15, although calculations were not performed using C-I-H scores due to the low rates of failure, statistical analyses did reveal several significant results between categorical demographic and injury variables and pass/fail rates on the effort tests at the $p<0.05$ level.

Table 15: Fisher's exact test and Likelihood ratio significance levels for effort test scores and demographic/injury variables

	WMT IR	WMT DR	MC	DS-C	AMI	CMTF
Age	.480	.269	1.00	.031*	1.00	.638
Hand	.424	1.00	.348	1.00	1.00	1.00
Gender	.704	.689	.692	.673	1.00	1.00
Education	.227	1.00	1.00	1.00	1.00	.653
Time	.724	.135	1.00	1.00	.098	.641
SoIP	.335	1.00	.118	.001*	.688	.175
Anxiety	1.00	1.00	1.00	.416	.689	.364
Depression	.679	.382	.017*	.643	1.00	.609
IQ	.480	.410	.354	.055	.375	.268
Group	.223	.786	1.00	1.00	<.001*	.180
WMT-MC	.046*	.340	.699	1.00	.233	.180
WMT-PA	.044*	.335	.688	1.00	.225	.656
WMT-FR	.053	.355	.242	.416	.053	.672

* significant at the $p < .05$ level

Failure rates on the Mental Control test were found to be related to depression scores; with 62.5% of those who reached the clinical cut-off for depression failing this test compared to 15.4% of those who scored below the cut-off (Fisher's; $p=.017$). When based on scaled scores, the relationship between scores on the Mental Control test and depression was again confirmed ($F[1, 32]=6.248$, $p=.018$), with people who were above the cut-off for depression having significantly lower scores on the Mental Control task. A further relationship was also identified between the Mental Control test and processing speed, with those in the slower processing speed group having lower scores on Mental Control ($F[1, 31]=7.914$, $p=.008$).

The findings for Mental Control were also confirmed through correlations, with lower scores related to higher depression levels ($r=-.455$, $p=.007$) and lower processing speed ($r=.462$, $p=.007$). In addition several other variables were correlated with the scaled scores on this effort test. Specifically, more years of education ($rho=.411$, $p=.016$), lower pre-morbid IQ ($r=-.496$, $P=.006$), and higher anxiety ($r=.462$, $p=.007$), although education was not significant once a Bonferroni correction was applied.

The Digit Symbol-Coding test was also found to be related to age (Fisher's; $p=.031$) and speed of processing ($p=.001$), with all of the failures occurring in those who were in the younger group and the slower processing speed group. The relationship between Digit Symbol-Coding and processing speed was confirmed via ANOVA ($F[1,31]=24.267$, $p<.001$) and correlations ($r=.569$, $p<.001$), with significantly lower Digit Symbol-Coding scores observed in the group with slower processing speed. However, the identified relationship between Digit Symbol-

Coding and age was not supported when related to scaled scores rather than pass and fail rates, although the findings did approach significance ($F[1, 31]=3.061$, $p=.090$).

Although not significant when applying categorical scores or ANOVAs, higher scores on the AMI effort test were found to correlate with longer time since injury/diagnosis (rho=.459, p=.006) and lower processing speed (rho=.354, p=.040), although the relationship with processing speed was no longer significant once the Bonferroni correction was applied. Age was also found to be correlated with scores on the WMT-IR (rho=-.304, p=.038), with people who were younger scoring lower on this effort test. Scores on the WMT-DR were also related to lower processing speed (r=.377, p=.020). However, these findings also did not remain significant once the Bonferroni correction was applied.

A relationship between failure rates on the WMT-Immediate Recognition trial and low memory scores on the WMT-Multiple Choice test (Fisher's p=.046) and WMT-Paired Associates test (Fisher's p=.044) was identified, and the relationship with the WMT-Free Recall test approached significance (Fisher's p=.053). Overall, those who failed the Immediate Recognition trial scored in the lower group on the memory measures of the WMT, and this trend was also observed for the Delayed Recognition trial, but did not reach significance.

When analyses were based on scaled scores statistically significant relationships were also identified between the WMT memory scores and several of the effort tests. As Table 16 shows, participants in the WMT-Multiple Choice low score group had significantly lower scores on the Immediate and Delayed Recognition trials of the WMT, along with lower scores on the Camden Memory Test for Faces. These relationships were also supported by correlations (WMT-IR; rho=.809, p<.001, WMT-DR; r=.714, p<.001, CMTF; r=.555, p=.001). Further, the WMT-Multiple Choice low score group was found to be related to significantly lower scores on the Autobiographical Memory Index, and this was highlighted in the correlations (rho=.367, p=.033) although when the Bonferroni correction was applied this became non-significant.

Table 16: ANOVA analyses for WMT memory measures and effort test scores

Effort Test	WMT-MC			WMT-PA			WMT-FR		
	F	Df	Sig.	F	df	Sig.	F	Df	Sig.
WMT-IR	19.649	1,32	<.001*	46.059	1,24	<.001*	12.397	1,32	.001*
WMT-DR	23.542	1,32	<.001*	19.440	1,32	<.001*	4.010	1,32	.054
C-I-H	0.579	1,32	.452	0.030	1,32	.864	0.422	1,32	.520
MC	0.710	1,31	.406	0.663	1,31	.422	2.131	1,31	.154
DS-C	0.092	1,31	.763	0.521	1,31	.476	0.026	1,31	.936
AMI	6.442	1,27	.017*	2.907	1,28	.100	2.245	1,32	.144
CMTF	6.107	1,28	.020*	8.519	1,32	.006*	2.538	1,28	.121

* significant at the p <.05 level

Table 16 also shows that participants in the WMT-Paired Associate low score group had significantly lower scores on the Camden memory Test for Faces and the Immediate and Delayed Recognition trials on the WMT. These relationships were supported by correlations between the scaled scores on these variables (WMT-IR; rho=.852 p<.001, WMT-DR; r=.683, p<.001, CMTF; r=.571, p=.001). Additional relationships were also identified via correlations between the Autobiographical Memory Index and the WMT-Paired Associates test (rho=.450, p=.008) and WMT-Free Recall test (rho=.373, p=.030), but again this last relationship was not significant once it had been Bonferroni corrected.

Further, the WMT-Free Recall low score group was related to significantly lower scores on the WMT-Immediate Recognition trial (see Table 16), and this was confirmed via correlations (rho=.686, p<.001). Additional significant relationships were also identified when WMT-Free Recall test scores were correlated with Mental Control (r=.357, p=.041) and the Delayed Recognition trial of the WMT (r=.466, p=.006), but the relationship with Mental Control did not remain significant once the Bonferroni correction had been applied.

3.3.5 Relationships between all effort tests

As part of the validation process for psychometric measures the validity of the measure must be assessed. A common way of assessing whether a particular test is measuring the construct it is supposed to measure (in this case 'effort') is to see how well it relates to other tests purportedly measuring the same construct. Therefore, the effort tests used in the current research were analysed to identify any relationships between them.

A number of the effort tests did display statistically significant relationships, with failures on some of the tests relating to failure rates on other tests. However, although this trend was observed across all tests of effort, some of the results were non-significant and the inconsistency of the identified relationships across the tests also indicates differential rates of specificity. Table 17 includes the Fisher's exact test significance levels across all effort tests.

Table 17: Fisher's exact test significance levels for all effort tests

	WMT IR	WMT DR	C-I-H	MC	DS-C	AMI
WMT-DR	.002*					
C-I-H	.200	.176				
MC	.031*	.031*	.273			
DS-C	.078	.078	.250	.047*		
AMI	.007*	.570	.250	.366	.366	
CMTF	.029*	.205	.242	.020*	.020*	.126

*significant at the p <.05 level

As expected, correlations between the scaled effort test scores also supported several of the identified relationships between categorical pass/fail rates. Table 17 shows the interrelationships between all effort test scaled scores, and highlights that higher scores on effort tests relate to higher scores on other effort tests. There was a 26.49% chance of finding one or more significant differences within the six tests that were performed for each effort variable. When Bonferroni corrections were employed the alpha level was reduced to $p<0.009$ for a significant result. Significant correlations between five of the eight cases remained once this correction had been applied.

Again, it can be seen that there is differential strength in the identifiable relationships between the tests and not all of the measures are significantly related to one another. Some of the correlations also contradict the significant results identified based on the categorical pass/fail data, with the previously identified relationships between the Camden Memory Test for Faces and the Digit Symbol-Coding and Mental Control tests failing to reach significance, and additional relationships being identified between the Camden Memory Test for Faces and the Autobiographical Memory Index.

Table 18: Correlations and significance levels between all effort test scores

	WMT IR	WMT DR	C-I-H	MC	DS-C	AMI
WMT-DR	.757					
	<.001**					
C-I-H	.110	.210				
	.522	.226				
MC	.354	.411	.221			
	.040*	.016*	.210			
DS-C	.054	.216	.340	.580		
	.765	.227	.053	<.001**		
AMI	.344	288	.266	.234	.325	
	.047*	.098	.128	.190	.065	
CMTF	.473	.505	.319	.194	.325	.560
	.005**	.002**	.066	.279	.065	.001**

*significant at the p<.05 level
**significant at the Bonferroni corrected level.

These analyses provide an indication of the construct validity of the effort measures, with not all measures correlating with each other. The Immediate and Delayed Recognition trials of the WMT also display differing relationships regarding the other effort tests.

3.4 Individual Data Inspection

The following section provides a breakdown of individual findings from a selection of participants who failed effort tests. Brief background information on the participant is described, along with their pattern of scores on the tests and qualitative information recorded by the researcher regarding their performance across the testing session.

The variables of WMT-IR, C-I-H and AMI are not normally distributed, therefore the z scores for these variables are likely to be skewed. However, the z-score is ultimately a descriptive measure for identifying how the person performs relative to everyone else in the sample. It is fairly robust and the SPSS calculations include the results from the individual when calculating the scores, which should alter the z-scores closer to 0 and actually provide a more conservative estimation.

Anxiety and Depression z-scores have been reversed in order to make interpretation easier, as high scores on both of these measures indicate more difficulty in these areas, whereas higher scores on the other measures indicate better performance.

3.4.1 Community Participants

Community Participant 1

This 36 year old female was educated to NVQ level and obtained a brain injury in 2002 as a result of a hypoglycaemic coma and possible alcohol abuse. The damage to her brain was recorded in her notes as widespread atrophy. During testing her processing speed was slow, as was her overall reaction to questions. The extent of this bradyphrenia meant that she often required that questions be repeated before she understood the nature of the tasks, and in comparison to other participants it took a relatively long time to complete all tests. The staff working with her also felt she was on a lot of medication and it was suggested that these drugs could be interacting or at a level high enough to induce a sedative effect. Table 19 provides a breakdown of this participant's scores across the whole range of administered tests.

Table 19: Obtained scores, z-scores and percentile ranks for community participant 1

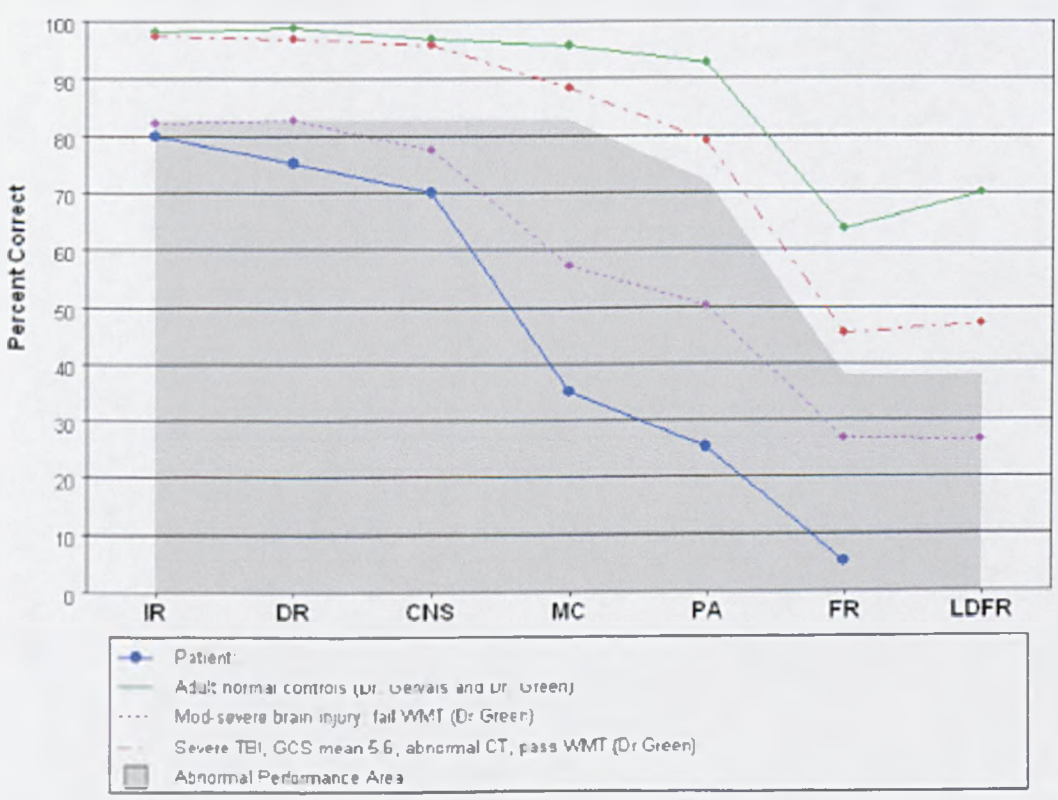
Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	81	-1.47	7
Anxiety	10	-1.34	9
Depression	12	-1.29	10
Processing Speed	19	-1.73	4
WMT-CR	70	-1.26	5
WMT-MC	35	-1.80	4
WMT-PA	25	-1.33	9
WMT-FR	5	-1.54	6
WMT-IR	80 (fail)	-0.32	37
WMT-DR	75 (fail)	-1.64	5
C-I-H	9 (pass)	-0.70	24
MC	10 (fail)	-1.27	10
DS-C	3 (fail)	-1.44	7
AMI	8 (fail)	-1.84	3
CMTF	12 (fail)	-2.62	0

As can be seen in Table 19, scores on all tests were much lower in comparison to the overall sample, with her speed of processing and Camden Memory Test for Faces scores being the lowest across all participants. In addition, although her Autobiographical Memory Index score is low, it was also in line with four of the other Community participants who all got the same answers wrong; specifically their address, telephone number, and what they had for breakfast that morning and dinner the night before. Such memory difficulties are not uncommon in this sample, and struggling to remember such information is also likely to reflect the transient nature of their stay in rehabilitation housing.

As can be seen from Figure 7 below, this participant was well below the normal adult scores and those with moderate to severe brain damage who pass the WMT, falling between 4 and 6 S.D below the mean for these two groups. Although her scores are more in line with people with a moderate to severe brain injury who fail the WMT, Green (2003) automatically views these people as putting in insufficient effort because they fail the task and are in litigation.

This participant's subjective reports of how much effort she was able to put in was also representative of her overall scores. She stated that she was only able to put in 40% effort, and gave the reasons as her 'brain not being as good as it used to be' and that she would have done better if she 'had a clearer head and a better memory'.

Figure 6: Graph of performance on the WMT for community participant 1



<u>*Community Participant 2*</u>

This 44 year old man was educated to GCSE level before being involved in a road traffic accident in 1995. The resulting evidence of episodic dyscontrol was reported as consistent with contusional damage to the right temporal lobe. During testing the participant was distractible and found it difficult to maintain attention, and the background noise in the testing environment (the participant's home) was also likely to impact on the participant's performance.

As can be seen in Table 20, scores on all tests were much lower in comparison to the overall sample, with consistency in his choices across the WMT-IR and WMT-DR being particularly low. Analyses of the WMT data also showed that his scores on the Immediate Recognition trial were six S.D below, and the Delayed Recognition trial was nine S.D below the normal adult mean and those with moderate to severe brain damage who pass the WMT. In addition, all of the WMT memory measures were approximately four S.D below the normal adult mean.

The poor performance on the C-I-H test was potentially a social desirability effect, as a staff member was visible to the participant during administration, and the participant looked towards her and laughed when choosing the wrong answer. However, in contrast to Participant 1, when asked the Brief Effort Questionnaire at the end of testing he stated he was able to put in 100%, and added that nothing would have helped him to do any better.

Table 20: Obtained scores, z-scores and percentile ranks for community participant 2

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	84	-1.19	3
Anxiety	3	0.81	79
Depression	0	1.22	89
Processing Speed	36	-0.81	21
WMT-CR	47.5	-3.17	0
WMT-MC	40	-1.52	7
WMT-PA	25	-1.33	9
WMT-FR	7.5	-1.41	8
WMT-IR	72.5 (fail)	-0.81	21
WMT-DR	65 (fail)	-2.74	0
C-I-H	5 (fail)	-4.88	0
MC	8 (fail)	-1.55	6
DS-C	3 (fail)	-0.93	18
AMI	8 (fail)	-1.84	3
CMTF	16 (fail)	-1.46	7

Community Participant 3

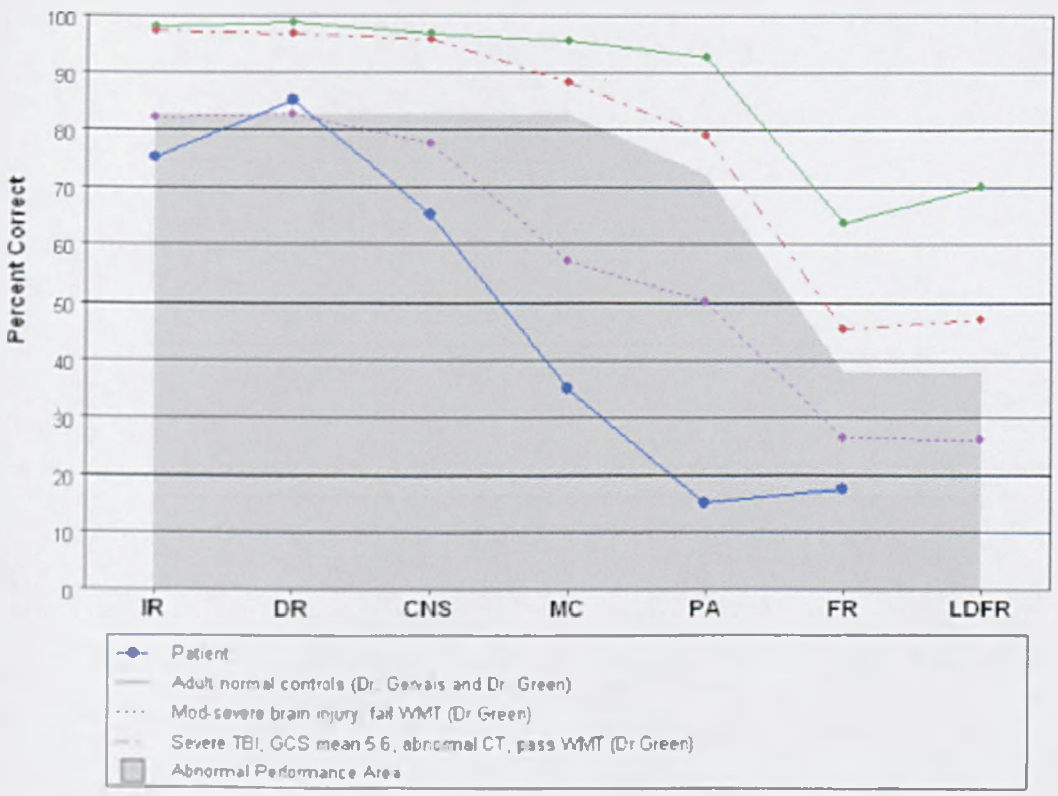
As the result of a fall in 2002, this 44 year old man sustained diffuse brain damage, with focal damage particularly noted in the frontal brain regions. He was educated to GCSE level prior to his brain injury. During testing he struggled to keep his attention on the tasks. He said the words out loud on WMT when viewing the first presentation, and also became distracted during parts of the second viewing of the word list, commenting throughout about how many he thought he could remember. He also struggled to understand some of the instructions during the testing session, particularly the more lengthy instructions on tasks such as Digit Symbol-Coding, unless they were presented several times and in manageable chunks.

Table 21: Obtained scores, z-scores and percentile ranks for community participant 3

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	Not reliable	-	-
Anxiety	7	-0.15	44
Depression	7	-0.54	29
Processing Speed	36	-0.37	36
WMT-CR	65	-1.69	5
WMT-MC	35	-1.80	4
WMT-PA	15	-1.79	4
WMT-FR	17.5	-0.87	19
WMT-IR	75 (fail)	-0.64	26
WMT-DR	85 (caution)	-0.55	29
C-I-H	10 (pass)	0.35	64
MC	11 (fail)	-1.13	13
DS-C	3 (fail)	-1.44	7
AMI	8 (fail)	-1.84	3
CMTF	13 (fail)	-2.33	1

As can be seen from Table 21 and Figure 8, this participant found some of the memory components of the WMT particularly difficult, with performance approximately five S.D below the normal adult mean. Performance on the Camden Memory Test for Faces was also poor. Although scores on the Immediate Recognition trial were below the established cut-off and fell approximately six S.D below the normal adult mean (as well as those with a severe TBI who pass the WMT used in the validation studies), he was able to perform relatively well on the Delayed Recognition trial of the WMT. When asked how much effort he was able to put in he stated 100%, and could not think of anything that would have helped his performance.

Figure 7: Graph of performance on the WMT for community participant 3



Community Participants 4 and 5

Both of these participants had their testing session stopped by the researcher based on the protocol outlined in appendix 5.4, and Table 22 provides a breakdown of the participants' scores across the tests that were completed.

Table 22: Obtained scores, z-scores and percentile ranks for community participants 4 and 5

Measure	Participant 4			Participant 5		
	Obtained Score	z-score	Percentile Rank	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	114	1.61	95	105	0.75	78
Anxiety	0	1.52	6	7	-0.15	44
Depression	0	1.22	11	7	-0.54	29
WMT-CR	n/a	-	-	62.5	1.90	97
WMT-IR	42.5 (fail)	2.76	100	50 (fail)	2.27	99
WMT-DR	n/a	-	-	77.5 (fail)	1.37	91
C-I-H	10 (pass)	0.35	64	9 (pass)	0.69	75
MC	n/a	-	-	9 (fail)	1.41	97

Participant 4 was a 66 year old male who had a brain haemorrhage in 1999 followed by a further haemorrhage in 2002. A CT brain scan conducted in 1999 revealed no specific local abnormality, but generalised brain shrinkage was noted, and the intracerebral haemorrhage in 2002 resulted in damage to the right frontal lobe. The participant was educated to GCSE level, but worked up to a high level in his chosen career, reflecting his relatively high pre-morbid IQ of 114. His scores at floor level on the Anxiety and Depression measures are also likely to reflect Rogers (2008a) 'impression management' response strategy, as he was known to be the type of man who did not like to show weakness or emotional vulnerability.

During testing the participant became extremely agitated by the direct feedback from the WMT when he incorrectly identified an item, stating that 'the test must be wrong', and he failed to focus attention on the next trials as he became more agitated. The scores for most of the WMT-IR therefore likely reflected Rogers (2008a) 'random responding' response style. The researcher halted the administration of the WMT after the WMT-IR, and administered the C-I-H to help the participant regain a sense of control (as this test was judged as simple enough to provide the participant with an opportunity to answer correctly).

Participant 5 was a 45 year old male who acquired a frontal brain injury after a fall in 2000. The participant was educated to GCSE level and had a pre-morbid IQ level within the average range. This participant was extremely slow at filling in the consent information and questionnaire measure, despite understanding what was required. He continually attempted to engage the examiner in conversation, and often went off at a tangent when the examiner tried to bring the participant back to the task. As testing progressed the participant also became increasingly anxious before the onset of each test, stating that he was 'worried about what I will have to do'. He also struggled to maintain his attention as the testing went on, and administration was eventually halted when the participant became distressed by the level of anxiety about what he would have to do next.

As can be seen in Table 22, the z-scores and percentile ranks indicate poor performance for both participants on the majority of the tests they completed, and both participants displayed behaviour typically associated with a frontal brain injury throughout the testing sessions. However, both participants were able to pass the Coin-In-hand test.

3.4.2 Inpatient Participants

Inpatient Participant 6

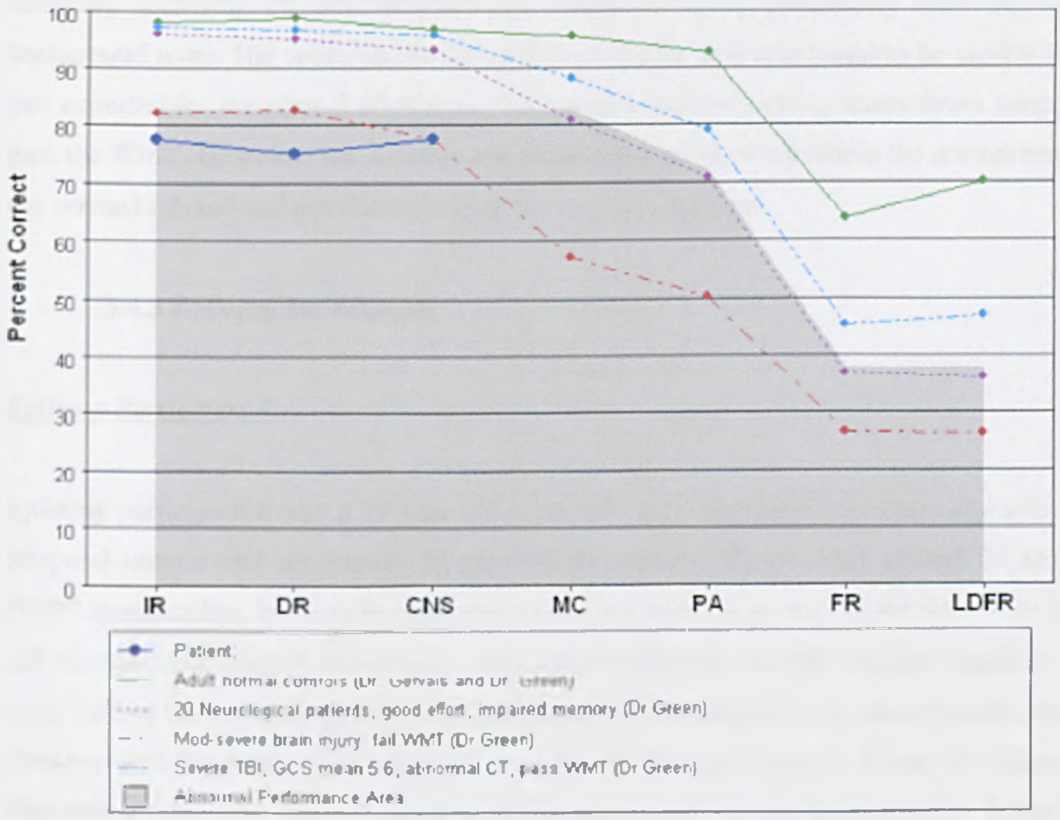
Inpatient Participant 6 was a 57 year old male who left school with no qualifications. He was tested 2 months post injury following a diagnosis of Endocarditis with accompanying multiple cerebral embolisms. His medical notes recorded diffuse damage across both hemispheres, with small areas of acute infarction within multiple vascular territories. During testing the participant was clearly very anxious about his current situation and about his performance on the measures, and as can be seen in Table 23, his very high scores on measures of Anxiety and Depression supported this observation. He also found it difficult to focus on the task and seemed unsure about the answers he was providing, although he repeatedly stated that he was happy to help.

Table 23: Obtained scores, z-scores and percentile ranks for inpatient participant 6

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	82	-1.38	8
Anxiety	20	-3.25	0
Depression	20	-3.81	0
WMT-IR	77.5 (fail)	-0.89	19
WMT-DR	75 (fail)	-1.64	5
WMT-CR	77.5 (fail)	-0.63	26

Figure 9 also provides a breakdown of the participant's scores on the WMT. Only the first three measures are provided because the Inpatient sample were not administered the memory components of the WMT. As can be seen from Figure 8, the WMT-Immediate Recognition score is well below that which would be expected from both a normal population and a brain injured population, falling more than four S.D below the mean, and the WMT-Delayed Recognition trial is more than six S.D below the mean for these groups. The participant is also well below the scores obtained from neurological patients tested by the WMT authors as part of the validation of the WMT. However, during the half hour gap between administrations the participant was engaged in a completely different, but still demanding, task as part of his ongoing rehabilitation on the ward. Therefore, it is possible that the Delayed Recognition results are poorer than what would be expected from this participant, as the other Inpatient participants just had a sedentary break during the half hour gap.

Figure 8: Graph of performance on the WMT for inpatient participant 6



Inpatient Participant 7

This 26 year old woman was tested 2 weeks after a craniotomy and resection of an intraventricular central neurocytoma. Her lowest GCS score was recorded as 10. The participant was educated to degree level, and she was particularly keen to participate in the study and find out about the research. However, she was noticeably distractible, and the background noise from the ward likely made this even more difficult for her. Table 24 provides a breakdown of the participant's scores across the administered measures.

Table 24: Obtained scores, z-scores and percentile ranks for inpatient participant 7

Measure	Score	z-score	Percentile Rank
Pre-morbid IQ	99	0.21	58
Anxiety	7	-0.15	44
Depression	5	-0.04	48
WMT-IR	82.5 (fail)	-0.43	33
WMT-DR	70 (fail)	-2.19	1
WMT-CR	67.5	-1.46	7

As can be noted from Table 24, her scores on the Delayed Recognition component of the WMT were particularly low, which is supportive of the observations regarding her distractibility and difficulty focussing on one task for any length of time, particularly with the observed background noise. Her score on the Delayed Recognition trial was found to be eight S.D below that expected for the normal adult mean and those with moderate to severe brain damage who pass the WMT. However, her Anxiety and Depression scores were within the normal range, and she seemed relaxed and positive regarding her current situation.

3.4.3 Epilepsy Participants

Epilepsy Participant 8

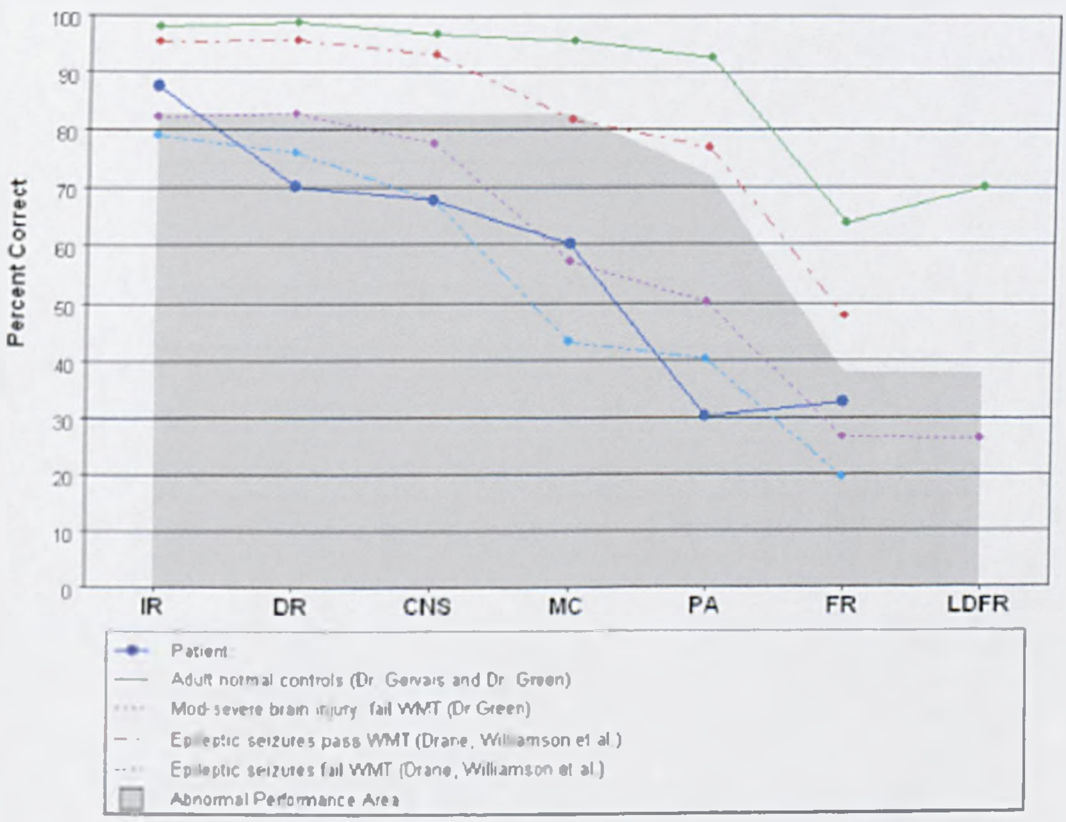
Epilepsy participant 8 was a 19 year old male who was diagnosed two years ago with a right temporal tumour and accompanying generalised seizures. He obtained several GCSEs and a BTEC qualification, but was unemployed and living with his parents at the time of testing. He did not seem particularly interested in the research and checked his mobile telephone several times during the assessment, although when asked by the researcher he stated that he wanted to continue and he remained compliant throughout. His self-reported levels of Anxiety and Depression were low, but a flatness of affect was noted. As can be seen from Table 25, his results on the WMT effort measures reflect these observations of variable performance, with the participant just passing the WMT-IR, but going on to fail the WMT-DR.

Table 25: Obtained scores, z-scores and percentile ranks for epilepsy participant 8

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	Not reliable	-	-
Anxiety	6	0.09	46
Depression	3	0.47	32
Processing Speed	41	-0.53	30
WMT-CR	67.5	-1.48	7
WMT-MC	60	-0.36	36
WMT-PA	30	-1.10	14
WMT-FR	32.5	-0.05	48
WMT-IR	87.5 (caution)	-1.20	11
WMT-DR	70 (fail)	-2.19	1
C-I-H	10 (pass)	0.35	64
MC	17 (pass)	-0.30	38
DS-C	4 (fail)	-0.93	18
AMI	11 (pass)	0.80	79
CMTF	23 (pass)	0.57	72

The performance of epilepsy participant 8 on the WMT test can also be seen in Figure 10, where his scores on the Immediate recognition trial are approximately two S.D below the normal adult mean and the mean for people with Epilepsy who pass the WMT, and his scores on the Delayed Recognition Trial are approximately eight S.D below the means for these two groups. However, his scores are in line with research looking at people with Epileptic seizures who are assumed to be putting in poor effort (Drane et al. 2005).

Figure 9: Graph of performance on the WMT for epilepsy participant 8



The fact that this participant's pre-morbid intellectual functioning could not be estimated is also a relatively frequent finding in men of this age group and socio-economic background, as poor language skills seem to mask overall intellectual ability. Again, observations support the idea that performance could also be reflective of Impression Management or Social Desirability response styles (Rogers, 2008a).

When asked the Brief Effort Questionnaire at the end of testing, the participant stated that he was able to put in 70% effort, and reported that he felt this was because he had only just woken up before the researcher arrived. He also stated that it would have been easier for him to remember if there had been more pictures, and the implied feeling was taken to be that he was embarrassed by his current difficulties, and was potentially hiding his fear of performing badly behind a veneer of poor application to the tasks.

Epilepsy Participant 9

Epilepsy participant 9 was a 31 year old male with no formal qualifications who had attended a special school due to his epilepsy. He was diagnosed with a frontal brain tumour at the age of three, which was removed, but left the participant with seizures as a result. Over time the tumour has developed again, with corresponding increase in symptoms, and the participant had also experienced one episode of status epilepticus in childhood.

Table 26 shows that despite slowed processing speed the participant was able to perform well on the WMT and obtained full marks on the Camden Memory Test for Faces. However, he performed poorly on Mental Control and Digit Symbol-Coding, indicating that scores on these measures could also reflect poor education, memory and processing speed. This participant displayed some nervousness generally during testing, and potentially some Social Desirability effects when filling in the Anxiety and Depression measure. His behaviour during testing was sometimes inappropriate; he made several comments about the attractiveness of the researcher being responsible for his poor performance.

When asked the Brief Effort Questionnaire at the end the participant indicated that he had put in 100% to all of the tests completed.

Table 26: Obtained scores, z-scores and percentile ranks for epilepsy participant 9

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	81	-1.47	7
Anxiety	3	0.81	21
Depression	0	1.22	11
Processing Speed	30	-1.13	13
WMT-CR	90	0.43	67
WMT-MC	70	0.21	58
WMT-PA	55	0.04	52
WMT-FR	12.5	-1.14	13
WMT-IR	95 (pass)	0.21	58
WMT-DR	95 (pass)	0.55	71
C-I-H	10 (pass)	0.35	64
MC	8 (fail)	-1.55	6
DS-C	4 (fail)	-0.98	18
AMI	11 (pass)	0.80	79
CMTF	25 (pass)	1.15	87

Epilepsy Participant 10

This 23 year old male was educated to GCSE level and had been working prior to the onset of partial and secondary generalised seizures five years ago originating from his right temporal lobe. As a result of the seizures the participant had experienced several episodes of post-ictal psychosis.

Table 27 shows that the participant had fairly high levels of Anxiety and Depression, along with a slowed processing speed. Variable performance on the WMT effort measures was also observed, although his WMT memory scores were reasonable for this sample. Low scores were also obtained on the Mental Control task compared to the rest of the participants.

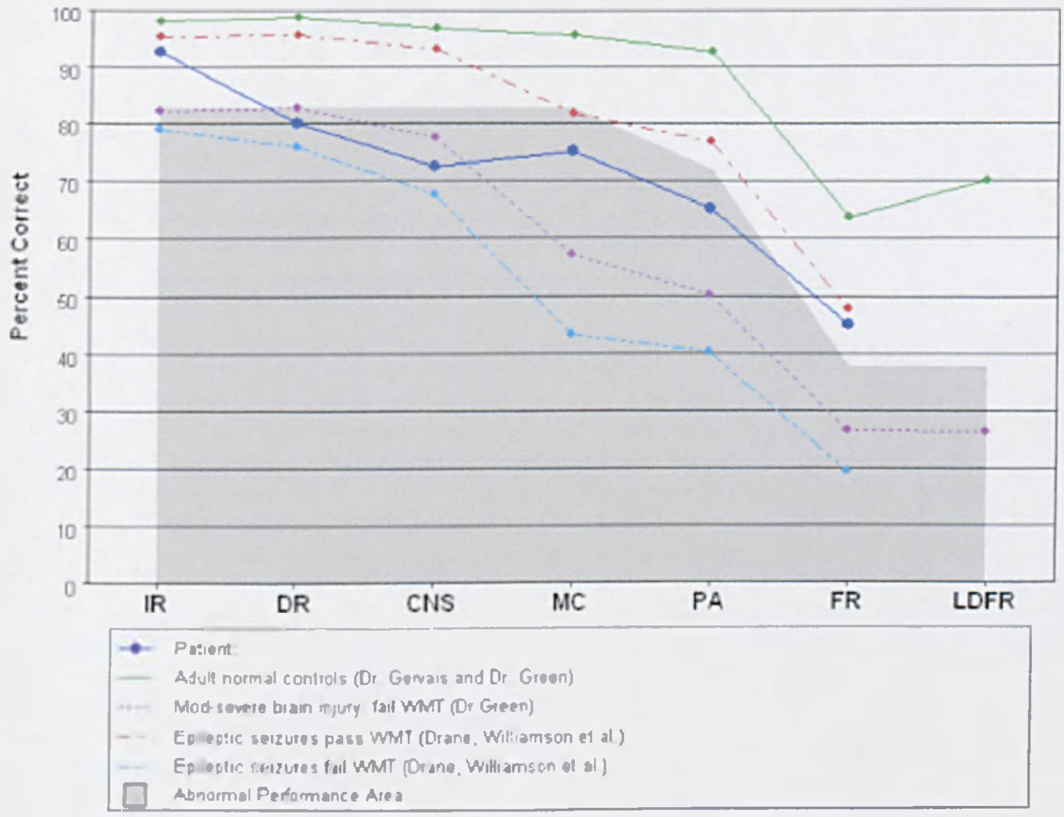
Table 27: Obtained scores, z-scores and percentile ranks for epilepsy participant 10

Measure	Obtained Score	z-score	Percentile Rank
Pre-morbid IQ	86	-1.01	16
Anxiety	11	-1.10	86
Depression	10	-1.29	90
Processing Speed	82	1.70	96
WMT-CR	72.5	-1.05	15
WMT-MC	75	0.50	69
WMT-PA	65	0.50	69
WMT-FR	45	0.63	74
WMT-IR	92.5 (pass)	-0.26	40
WMT-DR	80 (fail)	-1.10	14
C-I-H	10 (pass)	0.35	64
MC	11 (fail)	-1.13	13
DS-C	5 (pass)	-0.42	34
AMI	11 (pass)	0.80	79
CMTF	22 (pass)	0.28	61

Observation at testing indicated a flatness of affect, but he completed all of the tasks. When asked the questions from the Brief Effort Questionnaire about how much effort he felt he had been able to put in to the tasks he stated 60%, and when questioned further he alluded to possible family difficulties and arguments prior to my arrival for testing, and that he was not entirely comfortable with the atmosphere and his father being in the next room.

Figure 11 shows that this participant's scores on the WMT-DR fell more than four S.D below the normal adult mean and the mean expected in participants with Epilepsy who pass the WMT. Although his scores on the memory components of the WMT were relatively in line with what is expected of people with seizures.

Figure 10: Graph of performance on the WMT for epilepsy participant 10



3.4.4 Participants Passing Effort Tests

As well as a number of participants failing effort tests within this sample, a number of participants also passed despite having impairments in a variety of areas of functioning. For example, a 34 year old male Epilepsy participant had temporal lobe seizures originating from both the left and right hemispheres, and has significant memory difficulties related to Accelerated Forgetting; which is a very rare memory disorder in which the person forgets information over a period of days and weeks rather than minutes and hours. This participant's performance across effort tests was good, obtaining 90% for WMT-IR and 87.5% for WMT-DR (constituting a 'caution' rating) and he passed all other effort measures administered as part of the current research.

Another of the participants from the Community sample was a 44 year old male who had been in institutionalised care for most of his life after suffering a brain injury at the age of 15, finally moving to community based services when in his late 30s. Although he had significant frontal brain damage that meant living on his own in the community would be difficult, he was able to perform reasonably well on the WMT and other measures included in the current research, although he still scored below what the manual would predict for the WMT and obtained a 'caution' rating for WMT-DR. Figure 12 shows the results of this participant's WMT test scores and again highlights the necessity to approach scores on effort tests on a case by case basis and consider the complex and multifaceted issues that contribute to such scores.

Figure 11: Example graph from a community participant passing the WMT effort measures

3.5 Individual Analyses Summary

From the individual analyses performed above, participants in the current research were more likely to struggle due to anterior brain damage and personality rather than any focal lesions affecting areas for memory. The Community participants did particularly poorly, and dysexecutive reasons were often implicated, along with potential over-medication and genuine deficits.

In addition, distractibility and attentional factors were important in performing well on the tests used in the current research, particularly for the Community and Inpatient participants. Participants were often focussed on other aspects of the testing situation or environment as well as the measures themselves. The people who failed the effort tests from the Epilepsy sample were also identified as likely to have an Impression Management style of responding rather than a deliberate attempt to feign.

Tests based on established measures from neuropsychological batteries are also likely to have identified actual damage for most of the participants assessed in the current research, particularly difficulties with slowed thinking speed (Digit Symbol-Coding) and memory (Camden Memory Test for Faces), rather than measuring just effort in these populations.

3.6 Overall Summary

In relation to the first research question identified in the introduction, the above results reveal that the base rate incidence of failure on effort tests in a non-litigating brain injury population is substantially higher than that proposed by the authors of effort tests such as the WMT. Significant failure rates were identified across most of the effort tests in the current research, with rates reaching up to a quarter of the whole sample for some of the measures. Such findings are not in line with the proposition that the WMT is "virtually insensitive to all but the most extreme forms of impairment of learning and memory" (Green et al. 2002, p. 99), and indicates that legitimate patients with the injuries observed in these samples could be at risk of being formally identified as possible or probable malingerers based on the Slick et al. (1999) criteria if an external incentive for such participants was also identified.

In relation to the second research question, within the overarching non-litigating brain injury sample, participant subgroup had a significant impact on the scores obtained across most of the injury, psychological, neuropsychological and effort variables. Differential failure rates across effort tests were observed, with the implication being that other factors in addition to effort are contributing to effort test scores. Statistical analyses then went on to confirm that relationships with other explanatory variables were indeed evident in relation to a number of the effort tests administered in the current research.

The third research question considered whether any factors in addition to effort could account for scores on effort tests, and the data analyses did identify relationships between scores on effort measures and the explanatory variables, although such relationships varied widely. Table 28 highlights the statistically significant relationships identified between the variables in the current research.

Table 28: All significant relationships identified through statistical analyses between effort measures and additional explanatory variables in the current research

Explanatory Variables	Effort Measures					
	WMT IR	WMT DR	MC	DS-C	AMI	CMTF
Age	C*			F		
Hand						
Gender						
Education			C*			
Time since injury					C	
Processing Speed		C*	A, C	F, A, C	C*	
Anxiety			C			
Depression			F, A, C			
Pre-morbid IQ			C			
Group					LR, A	A
WMT-MC	F, A, C	A, C			A, C*	A, C
WMT-PA	F, A, C	A, C			C	A, C
WMT-FR	A, C	C	C*		C*	

F = Fisher's Exact Test, LR = Likelihood Ratio, A = ANOVA, C = Correlation, C* = Correlation that fails to retain significance once the Bonferroni correction is applied.

Table 28 shows that failures across six of the seven effort tests were related to at least two of the additional explanatory variables included in the study, and the findings from one type of statistical data analyses were usually supported by findings from other statistical tests (the C-I-H test could not be included in the statistical analyses based on the statistical assumptions required for such calculations). However, the correlations need to be interpreted with caution due to the increased risk of Type I error, and not every type of statistical analyses could be performed on each variable due to the nature of the data. In addition, the finding that failure on the tests of effort used in this study was not just related to a lack of effort on the part of the participant being assessed was also supported by an analysis of individual participant data.

The case study material presented in the results provides more in depth information regarding potential reasons for failure among individual participants, with such factors including nature of brain injury, memory and processing speed deficits, attention, and impression management. Further, as can be seen in Table 28, although all the tests were applied as measures of effort, the relationships with other variables were not consistent. Different effort tests were related to different demographic, injury, psychological and neuropsychological variables, indicating that specific effort tests were measuring other constructs in addition to effort. This finding was also supported by further data analyses relating to the construct validity of the tests. Some of the scores on effort tests showed clear significant relationships with scores on other effort tests, which provides support for the contention that the tests are measuring similar things, but this was not consistent across every effort test.

DISCUSSION

4.1 Review of Aims

For any neuropsychological testing to be valid people have to put in their best effort, otherwise clinicians may conclude that the person has a genuine injury, or is significantly more impaired than they actually are, when in fact this is not the case. In an attempt to identify people who may be feigning or exaggerating numerous tests of 'effort' such as the Word Memory Test have been designed, with the authors concluding that such measures are insensitive to genuine impairment when using the derived cut-offs (e.g. Green, 2003). However, McCarter et al. (2009) has recently found that 35% of clinicians in the UK think that measures of effort are unreliable or misclassify people, which indicates that neuropsychologists are wary of the claims made by standardised tests. Recent literature has also raised doubts regarding the claims about effort tests, and highlights the importance of further research to establish base rates of failure in genuine clinical populations (e.g. Merten et al., 2006). Such studies provide further validation of effort tests and provide clinicians with data regarding specific groups of patients so that they are able to avoid making a Type I error and assuming someone to be feigning or exaggerating their injuries when they are not.

In this straightforward and novel study, the base rates of effort test failure in a genuine brain injury population with no incentive to feign were investigated, with specific focus on the Word Memory Test. The aim was to establish how well such patients performed on the WMT, along with several less well established or researched tests of effort. This is the first time that a study has set out to investigate failure on a variety of effort tests in people from the UK who either have long-standing brain injuries and severe impairments within a community rehabilitation setting, inpatients who have recently obtained a head injury, and people with intractable epilepsy who have identifiable brain lesions/tumours.

The following section reviews the findings of this research in relation to the initial aims of the study and situates them in the context of previous literature regarding effort test performance. Firstly, the findings related to the base rates of effort test failure are discussed. A detailed examination of the statistical procedures used to investigate the data is then discussed, along with the implications of any shortcomings. The strengths and limitations of the study as a whole are also considered, in addition to the clinical implications for practice. Finally, directions for future research extending from the findings of the current study will be considered.

4.2 Results from the Current Research

The purpose of the current research was to add to the psychometric validation of the Word Memory Test and several other measures of effort by providing data from a clinical sample with mixed cognitive impairments. Of particular interest were base rates of failure on effort tests, and in contrast to previous research the current findings identified unacceptable rates of false positives across a wide range of effort tests at the recommended cut-off levels (<90%; Baker et al., 2000).

This study provides data from a relatively large clinical sample (N=47) on a number of different measures of effort. Mean and median scores across all of the effort tests in this sample were above the cut-offs for feigning, indicating that as an overall group this population were able to pass effort tests. However, participant variation within the tests was considerable. The results indicate that caution is required when interpreting results based on established cut-off scores on a significant majority of the tests administered in this research. Specifically, scores below cut-off were obtained by 25.5% of a genuine clinical sample with no incentive to feign on the WMT, despite the fact that such levels of failure are thought to be highly suggestive of poor effort and potential feigning. As Green (2003) states "The recognition subtests were designed to avoid confusing actual impairment with deliberate exaggeration. They are virtually insensitive to all but the most extreme forms of impairment of memory and the range of genuine scores is very narrow indeed." (p. 6). Had the people in the current study been in litigation or had ongoing benefits claims then the obtained scores could have been interpreted as evidence of feigning or exaggeration based on the Slick et al. (1999) criteria, and the veracity of their claims may have been questioned.

Assuming that the effort test failures in the current research are false positive findings, this indicates that the rate of inaccurate identification is unacceptably high, which has a significant impact on the specificity of the test. Such results challenge Green et al.'s (2002) claim that all patients except for 0.02% with "very severe and widespread cognitive impairment" (p. 117) pass the WMT. A central claim of the WMT is that it assesses only effort. Therefore, in order to support the validation of the WMT and back up the claims made regarding how performance should be interpreted, the measure must be demonstrated to be insensitive to genuine impairment in actual patients rather than with normal participants asked to simulate deficits. However, the findings from the current study suggest that scores on the WMT could also be affected by ability rather than just effort.

Table 29 includes a comparison of the results quoted in the WMT manual (Green, 2003) and the current study, showing that the overall mean scores identified in the current research fell much below the scores reported by Green (2003). T-test calculations established that all comparisons between the current research and Green's (2003) validation samples were significant at the $p<.03$ level (see table 30 in appendix 5.5). The scores from the present study actually fall within Green's (2003) 'caution' ratings rather than a clear pass, which Green reports as one S.D below the mean for people with severe brain injuries, and as such are viewed as "doubtful in most cases" (Green, 2003; p. 16).

Table 29: Comparison of Green (2003) scores and the scores obtained in the current research

	Study	WMT-IR	WMT-DR
As reported in	Moderate-severe brain injuries	M = 95.5%	M = 96.1%
Green (2003):	(N=57)	S.D = 5.1	S.D = 3.9
	Neurological - impaired memory	M = 96.1%	M = 94.9%
	(N=20)	S.D = 3.3	S.D = 5.5
	Neurological - normal memory	M = 96.2%	M = 96.2%
	(N=20)	S.D = 3.6	S.D = 4.9
Current	Genuine brain injury sample	M = 88.5%	M = 90.0%
Research:	(N=47)	S.D = 12.2	S.D = 9.1

However, as highlighted in the Introduction, it is unclear how many of the participants included in the validation research fail the WMT when they are from a genuine clinical sample with no incentive to feign. Green (2003) does not make it clear whether the data from the 57 moderate to severe cases who pass the WMT, as cited in the manual, were actually the same patients reported in the Green and Allen (1999) study, or indeed the Green et al. (1999) study. Nor does the manual state where the people with moderate-severe brain injuries who fail the WMT originate from (N=34). If they are all from the same core data set then all of these participants were assessed as part of litigation proceedings, but the ones who passed were reported as having adequate levels of effort because their brain scans showed actual damage and their claims for disability benefits had been accepted (Green et al. 1999). It is possible that the same claims could be made regarding the 34 people who failed the WMT, who had mean scores of 82% (S.D = 11.1) on the WMT-IR and 82.7% (S.D = 14.3) on the WMT-DR. The standard deviations alone indicate wide variation in the scores of this sample of participants.

Further evidence of differential failure rates on the WMT within the literature comes from Gorissen, et al. (2005), who found a 10% failure rate on the WMT in their control sample of neurological patients. Although higher than the rate predicted by Green (2003), this is still substantially below the rate of 25.5% found in the current research. Merten et al. (2006) also found that all people with genuine brain injuries who did not have 'clinically obvious symptoms' passed the oral version of the WMT, but a failure rate of 50% on the WMT-IR and 58% on the WMT-DR was found in 24 participants with clinically obvious symptoms. In addition, Batt et al. (2008) found a 44% failure in their standard administration group of 60 people with severe injuries, and a 75% failure rate in people with severe brain injuries when they were distracted. The findings of the current study are not as large as those found in people with 'obvious' symptoms in the Merten et al. (2006) study or Batt et al. (2008). This indicates that the sample in this research may be less impaired, and Batt et al (2008) state that all of their brain injured participants met the Merten study criteria for obvious symptoms; specifically, bradyphrenia, repetitive speech and word finding difficulties. Whereas some participants across all three of the subgroups used in the current research would not have met these criteria, particularly in the Epilepsy subgroup. However, the failure rates in the current study still indicate a significant increase from that predicted by Green (2003).

In the only other UK study with brain injured participants, the retrospective research by Bunnage et al. (2008) found an overall failure rate of 25.8% on the WMT-IR, 9% of which had a potential external incentive such as benefits, work absence or compensation claims. The rate of 25.8% is in line with the current finding of 21.3%, although it is likely that with such a wide definition of external incentive a significant proportion of the current sample would be identified as such. This is discussed further in the limitations section.

In an epilepsy sample, Drane et al. (2006) found an 8.1% failure rate on the WMT, whereas Dodrill (2008) found a rate of 25%. However, despite the current study identifying failure rates of 18.8% in this subgroup, there are differences in selection criteria between the participants in the current study and that of Dodrill (2008). It is likely that the current research reflects the strict criteria used by Drane et al. (2006), that is, excluding people who could not live independently or those with cognitive deficits, and thereby reducing the potential for failures based on genuine impairment. A choice was made in the current research regarding subgroup membership despite the clear comorbidity of brain damage, tumours and epilepsy across the whole clinical sample. It is likely that a number of the Community participants in the current research could equally be judged as fitting the Epilepsy group when using the more relaxed criteria of Dodrill (2008). This would likely increase the failure rates of people with epilepsy observed in the current research in line with that observed in the Dodrill (2008) study.

This study used other tests in addition to the WMT. Again, the failure rates for these tasks are higher than expected based on the rates suggested in other research, ranging from 2.8% to

38.9%. For example, regarding the Digit Symbol-Coding task, Inman and Berry (2002) found a mean of 11.9 (S.D = 1.45) and no failures in participants with genuine brain injuries when using the cut-off scaled score of <5. However, all of these participants were enrolled at university and are therefore likely to be functioning at much higher levels than the participants in the current research.

The cut-off score of <8.5 for the Coin-In-Hand test identified in the work of Kelly et al. (2005) was derived from research that included 40 brain injured participants. This cut-off performed well when applied to the people participating in the current research. However, Kelly et al. (2005) also established the cut-off of <13.5 for the Mental Control test and <10.5 for the Autobiographical Memory Index based on the same research. Despite the mean for the overall sample in the current research being above the cut-off for the Mental Control test (M = 19.1; S.D = 7.3) nine participants still fell below the cut-off (25%). A large proportion of the Community participants (14/18 who completed) also failed the AMI at the proposed cut-off. Although no participants in the current study scored less than eight on this measure. It is possible that the participants in the Kelly et al. (2005) study were less impaired than those in the current research, as they were described as having attended a head injury rehabilitation service, so they may have performed better overall compared to people currently still in rehabilitation services. Further suggestions for the high failure rates observed on both the Mental Control test and the Autobiographical Memory Index are discussed in section 4.3 below.

Overall, the rates of failure on the WMT identified in the current research are more consistent with the levels of failure found in independent studies using the same measures than the levels of failure identified in research from the authors and associates. Failure rates across most of the additional effort tests are also higher than with other studies. The evidence provided in the current research raises significant questions regarding the suitability of the WMT for use with participants with a brain injury severe enough to require community rehabilitation services, patients with newly acquired injuries currently in a post-acute inpatient rehabilitation ward, and people with intractable epilepsy, even though Green (2003) states that "using the 82.5% cut-off, there will be very few false positives indeed, even in cases of severe traumatic brain injury and neurological disease" (p. 16).

This study also suggests potential differences in the reasons why specific subgroups within an overarching genuine brain injury population and no incentives to feign may perform poorly on effort tests. Although the number of participants recruited to the study is relatively large when considered as a genuine clinical sample, further research is required to establish the reliability of these findings through investigating subgroups. This would increase the statistical power of the research through the acquisition of larger and more homogenous samples. As it is, the findings suggest potential differences between subgroups, but a lack of power prevents the results being more than just exploratory in nature. Additional research could focus on recruiting greater numbers of participants within the individual subgroups, which may help to draw firmer conclusions based on comparisons. It also remains to be identified whether such results can be replicated and extended to people with less severe brain injuries/epilepsy. Failure to continue research into purported measures of effort could result in serious consequences for participants if they are judged to be feigning rather than understood as a false positive. The current findings suggest that despite the high sensitivity rates of measures such as the WMT, the specificity rate is unacceptable in people with genuine brain injuries.

4.3 Factors Contributing to Effort Test Scores

Failure rates on the Mental Control test were found to be related to depression scores; with 62.5% of those who reached the clinical cut-off for depression failing this test compared to only 15.4% of those who scored below the cut-off. Poor Mental Control scores were also significantly associated with lower scores on the processing speed task. Several other variables were correlated with lower scaled scores on the Mental Control test, including fewer years of education, lower pre-morbid IQ, and higher anxiety. None of these are surprising given that the test is based on over-learned information from childhood and the concomitant anxiety that usually accompanies depression.

Slowed mental processing is recognised as a key symptom of depression (American Psychiatric Association, 2000). The timed tests on the WAIS have been shown to be more sensitive to such slowed processing (Pernicano, 1986) and Gorlyn, Keilp, Oquendo, Burke, Sackeim, and Mann (2006) have recently shown that depression is related to the timed tasks targeting processing speed on the WAIS-III. This supports the identified relationship between depression, speed of processing and the Mental Control task found in the current research; the Mental Control task in the current study could be measuring speed of processing at impaired levels rather than effort.

Lower scores on the Digit Symbol-Coding test was also found to be related to lower scores on speed of processing, and again this is an expected finding considering that this test in its usual role mainly identifies difficulties in speed of processing, with a smaller contribution for memory (Joy, Kaplan, & Fein, 2004). Therefore, the results noted in the current research indicate that failures on this test at the cut-offs provided for measuring effort could in fact be measuring processing speed in this sample. Overall, it appears that both the Mental Control and the Digit-Symbol Coding tests are simply too sensitive to genuine disorders in the current population, likely because they were developed specifically to encompass all levels of genuinely testable impairment. For example, Epilepsy Participant 9 within the individual analyses section failed both of these tests and also performed poorly on the processing speed task.

It is not disputed that people with moderate-severe brain injuries would do poorly on the Mental Control and Digit Symbol-Coding tasks, rather it is suggested that people with mild injuries not putting in enough effort would also do poorly for reasons of poor effort rather than genuine impairment. The results from the current study suggest that the Mental Control and Digit-Symbol Coding tasks are very sensitive to genuine impairment. Therefore, passing these tests at the standardised effort cut-offs indicates that the person is not very impaired, but failing the tests is less informative. It is also surprising that failures on the WMT-IR and WMT-DR are in line with failure rates on these other measures, particularly as these other tests were initially designed to identify severe impairments rather than effort.

Failure on the Coin-In-Hand test was uncommon in the current study, supporting the proposition that this is an extremely simple task (Kapur, 1994) with very high specificity rates, although sensitivity is likely to be much reduced as a result. As reported in the individual analyses, the one person who did fail was likely due to a social desirability response style (Rogers, 2008a) rather than genuine impairment. The high failure rate for the Autobiographical Memory Index was a surprise finding because the measure was previously designed as a very simple effort test that is insensitive to injury (Wiggins & Brandt, 1988). It is possible that this could be due to changes in the applicability of the questions. For example, when considering the increase in mobile phone use and the concomitant reduction of people who no longer need to remember their telephone number, this question may now need to be revised. The Community participants were also staying in community residential housing, with some of them only just moving into the homes they were residing in and they often had their evening meal prepared for them. Therefore, failure to provide their full address or information regarding the food they had eaten the day before is also not surprising given the impaired memory of this sample and the instability of their housing situation. Future research could evaluate the performance of the AMI in genuinely brain injured participants when these questions have been removed or substituted. The current research suggests that a revised cut-off score of <8 may need to be applied in this population. The Camden Memory Test for Faces was also found to be related to some of the WMT-memory measures and scores on the AMI, again indicating that the results from this test could also be measuring memory in this population rather than a lack of effort.

No major relationships were identified between most of the demographic, injury, psychological or independent neuropsychological variables for failure rates across the WMT effort measures. A small correlation was found between the WMT-IR and age, with people who were younger scoring lower on this effort test (rho=-.304) and between WMT-DR and lower processing speed (r=.377). However, these findings did not remain significant once the Bonferroni correction was applied, and the main remaining relationships were actually identified within the WMT memory measures themselves.

According to Green (2003), people should score well above the cut-off scores on the WMT effort test measures; "there is a strong ceiling effect on the WMT effort measures" (p. 34) and therefore scores should be relatively consistent across the Immediate and Delayed tasks. The findings in the current research suggest that lower scores on the WMT effort measures are significantly related to the WMT memory measures; indicating that a relationship exists between memory and effort tests in the current sample of people with genuine brain injuries and no incentive to feign. However, these findings are limited by the fact that the memory measures are taken from the WMT, and as such no separate test of memory function independent of the WMT was taken. Future research could attempt to establish the pattern identified in the current findings through the administration of independent tests of memory. Regardless, on the basis of the current research, clinicians do need to be aware of the potential for low scores on the effort components of the WMT if the patient has particularly poor memory.

Significant relationships between failure rates on the WMT-Immediate Recognition trial and low memory scores on the WMT-Multiple Choice test and WMT-Paired Associates were identified, and the relationship with the WMT-Free Recall test approached significance. Overall, those who failed the WMT-IR scored in the lower group on the memory measures of the WMT, and this trend was also observed for the WMT-DR, but did not reach significance. When analyses were based on scaled scores statistically significant relationships were also identified. Participants in the WMT-Multiple Choice low score group had significantly lower scores on the Immediate and Delayed Recognition trials of the WMT, and this was also supported by correlations. Although the significant relationships were not consistent across the WMT-IR or WMT-DR, nor across the different types of WMT memory measures, the fact that significant relationships exist between the scores could indicate that the WMT effort measures are also measuring some element of memory in the population under investigation. The consistency of the WMT profile across all of the subtest scores, with lower scores on one relating to lower scores on the others, therefore provides an indication that the participants in the current research could have failed due to severity of cognitive impairments rather than insufficient effort. The individual analyses provide examples of this, as low memory scores are observed alongside lowered WMT effort scores. For example, Community Participant 3 failed the WMT-IR, received a caution on the WMT-DR, and then went on to obtain scores ranging from 15 to 35 out of 100 on the WMT memory measures.

As previously suggested, if Mental Control and Digit-Symbol Coding are actually measuring processing speed, whilst Camden and AMI are measuring memory in this study, then the finding that the WMT-IR is related to several of these measures (WMT-DR, MC, AMI, CMTF) cannot be interpreted as supportive of *solely* measuring effort. There was also inconsistency within the identified relationships across the tests, which also provides and indication of differential rates of specificity for these tests within the current subgroups. It is also interesting that the Immediate and Delayed Recognition trials of the WMT display differing relationships regarding the other effort tests. Both measures are derived from the same test and very obviously linked with regard to content and structure. In addition, validation research suggests that both tests are measuring effort, with scores at ceiling levels in genuine participants with no incentive to feign. Therefore, the differences in relationships with other tests within the current research could suggest that other factors are affecting the performance of the WMT-IR and the WMT-DR in different ways. It may be that these measures have differential sensitivity to impairment and effort in populations such as those involved in the current study, with the delay in administration of the WMT-DR being an obvious choice for further investigation.

The current research suggests all of the effort measures apart from the Coin-In-Hand test are sensitive to genuine brain impairment, and as such are not suitable for identifying poor effort in these populations. It is not clear why the participants performed so poorly when compared to the participants discussed in the WMT manual. As Green (2003) states that "there are very few "false positives", with the exception of people with dementia" (p. 9). It is surprising to find that a genuine clinical sample of people with brain injuries and no incentive to feign failed the WMT and other effort tests at such significant rates. However, it is possible that the patients assessed in the current research had more severe impairments than those assessed as part of the validation research. For example, the differences may be related to variation in classification, as it is not entirely clear what Green (2003) means by 'profound' impairment, and if a more stringent assessment of such deficits was applied then such failure rates may also have been found in the validation studies. In addition, Green (2003) states that if a patient "needs constant care, has a definite brain disease, has an abnormal brain scan and is not allowed to drive" (p. 42) then the possibility that the scores are the result of a false positive should be considered. It is noted that the vast majority of Inpatient and Community participants were within 24 hour services, although these services were 'rehabilitation' rather than 'care' and most participants did not require 'constant' supervision. Furthermore, Green (2003) does repeatedly point out that even people with 'severe' impairments can pass the WMT, and in relation to false positives "in the first 1,000 cases given the WMT clinically by Green, it is thought that there were only about five cases of this type" (p. 42). However, it is very important to assess such populations with regard to their performance on the WMT, as many of the Community participants had been through the litigation process and it is likely that a proportion of the Inpatient'sample will be expected to go through this process in the future.

4.3.1 Additional explanations for the current findings

One alternative explanation is that the participants identified in the current study were in fact feigning or exaggerating their symptoms. Some studies support the view that malingering is present in people who have genuine injuries (e.g. Bianchini, Greve, & Love, 2003; Trueblood & Schmidt, 1993) even when they are severe (Boone & Lu, 2003). Therefore, one explanation for the observed rates may be that the participants in the current research were deliberately not putting in effort, or that a psychiatric condition was compounding their performance.

However, it seems unlikely that the participants in this study were malingering, as there were no clear incentives for them to do so and no psychiatric conditions were noted that would have precluded administration of the tests on such grounds. Green (2003) also does not report any more information about his 'genuine' brain injured participants other than that they are assumed to be putting in appropriate effort by the very fact that they have observable injuries on brain scans, are 'obviously' impaired, and have already been awarded disability benefits. Therefore, the participants in the current study matched the validation sample on this basis, and could therefore be assumed as putting in adequate effort based on these criteria.

All participants also went through a lengthy consent process with consultants and care workers; participating in this research was entirely voluntary and they could withdraw at any time, so if anyone was genuinely malingering it is likely that they would have refused. However, the refusal rate was low, particularly with the most impaired groups (Community and Inpatient), and these subgroups were also the ones who showed most interest in taking part, reporting that they were often bored during the day and were glad of something to do. In addition, it is difficult to imagine anyone faking a brain injury as a result of wanting community rehabilitation or inpatient admission, or that the severity levels of such injuries could be consistently replicated enough to deceive professionals in such institutions. All of the participants within the Community and Epilepsy groups had also already been through litigation and/or benefits claims. However, it could also be that the participants were not motivated to perform well on tests based on the fact that it was research rather than an assessment that would benefit them.

It would be useful to conduct a study in which participants with genuine injuries similar to those observed in the current research were provided with an incentive to identify whether failure rates reduced under these conditions. Future research could also consider the inclusion of simulators or suspected malingerers so that sensitivity and predictive power could be assessed.

This study highlights that additional research is needed to investigate the psychometric properties of the effort measures, particularly when using clinical samples, to help improve the diagnostic utility of the measures.

When considering magnitude of error, no true malingering was evident (i.e. below chance performance) even though some people did do worse than others. The WMT also counts as a 'floor effect' detection strategy because it is supposed to be a simple test that can be completed by most impaired people. However, as this research has shown, the test may not be as infallible as previously claimed. The WMT also counts as a 'violation of learning principles' detection strategy, but it is limited because it is only applicable to the current learning during the WMT and other factors may be contributing to the results. For example, in the current study a number of participants scored higher on the Delayed Recognition component of the WMT compared to the Immediate Recognition trial, but this could just be because of speed of processing difficulties. Due to the similar rates across the WMT-IR and WMT-DR, it is also likely that the findings of higher scores were an artefact, as the differences between the scores were only modest.

Despite Green (2003) stating that "the recognition subtests were designed to avoid confusing actual impairment with deliberate exaggeration" (p. 6), it is also possible that some participants were unable to perform well on effort tests for other reasons. As Bunnage et al. (2008) point out, clinicians may not know what all the possible secondary gains are in this situation. As outlined by Rogers (2008a), people may fail effort tests due to a variety of different response styles, and such styles vary depending on context. Rogers (2008a) also states that there are both internal and external influences on reporting, but points out that there are no standardised measures to assess these idiosyncratic factors. The examinees in clinical settings may view dissimulation as adaptive and positive, including maintaining autonomy, avoiding difficult circumstances, disengaging from an involuntary process, maintaining a sense of worth, or minimising avoidable pain. In addition, it is possible that the researcher being young and female produced increased levels of Impression Management than a middle aged man would within a population of young brain-injured men.

As highlighted in the individual analyses section, clinical observations from the current research suggest factors such as medication and impression management could play a role (Lishman, 1997; Rogers, 2008a). For example, Community participant 1 showed marked bradyphrenia and was known to be on high doses of medication thought to be affecting her daily functioning, and Epilepsy Participants 8 and 10 were both young men who seemed to be embarrassed by their current situation and problems. This highlights the importance of considering personal internal and external factors when anyone obtains a low score on effort tests, and clinicians must think through the alternatives for poor test performance in addition to malingering. Future research could attempt to assess alternative response styles in more detail in order to establish the extent of their influence on tests purported to measure effort.

Another potential reason for the differences in failure rates identified between the effort tests and across subgroups is the different presentation and testing structures of each test. For example, Batt et al. (2008) suggest differences in failure rates across the WMT and TOMM could be due to differences in the stimulus material, as the WMT uses semantically related words, which could create confusion and more errors due to 'semantic interference' (Wehner, Ahlfors, & Mody, 2007) even if they are applying their full effort. During the WMT people are presented with two words that belong to the same category and strongly related to one another. If a participant cannot remember the exact word then they may remember the general category, but struggle when faced with a choice between two very similar words. (e.g., 'fire' *versus* 'flame'). Such effects of semantic memory are well documented even in healthy populations when using the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). This paradigm involves presenting participants with lists of semantically associated words (e.g. fire, flame, sun, warm and so on) that are all related to a non-presented target theme or related lure (e.g. hot). Healthy participants frequently intrude the related lure on free recall tests and have high levels of false identification of these words on recognition tests.

When assessing people with amnesia based on a modified DRM paradigm, research by Schacter, Verfaellie, Anes and Racine (1998) found that whereas control participants had decreasing levels of false recognition over successive trials the amnesic participants showed no such pattern. Schacter et al. (1998) proposed that this was a result of true and false recognition being dependent upon different types of information, including item-specific recollection based on a previous encounter with an item, and the general gist or meaning gained from a collection of items. Therefore, gist representations are developed which are experienced as recollection or familiarity when a studied item or a related lure is presented in later recognition tests.

Based on research suggesting that frontal lobe damage appears to be associated with higher rates of false recognition (Parkin, Bindschaedler, Harsent, & Metzler, 1996), impaired item-specific memory (Parkin, Ward, Bindschaedler, Squires, & Powell, 1999) and source memory confusion (Janowsky, Shimamura, & Squire, 1989). Budson, Sullivan, Mayer, Daffner, Black and Schacter (2002) applied the DRM paradigm to people with frontal lobe lesions. They found that unlike controls, patients with frontal lesions could not reduce their rates of false recognition across successive trials. Budson et al. (2002) suggest that repeated presentation produces a more robust gist-based semantic representation, leading to high rates of false recognition because of poor item-specific recollection. Anterior prefrontal regions have been identified as involved in post-retrieval monitoring and verification (Rugg, Fletcher, Firth, Frackowiak, & Dolan, 1996), which would be needed to use item-specific information to suppress false recognition. However, if there is damage to source monitoring and verification-inhibition mechanisms located in the frontal lobes it may mean that patients are unable to use increasing explicit recollection to counteract or suppress the developing gist representations.

Batt et al. (2008) also suggest that if the resulting brain injury has affected the ability to monitor performance and learn from feedback then this could also contribute to test results. These suggestions are consistent with the notion that more anterior or widespread damage could be contributing to scores on the WMT in the current research despite no significant differences being identified due to a lack of statistical power. For example, in their study of people with Schizophrenia, Gorissen et al. (2005) point out that people may fail effort tests because of an inability to monitor their performance and correct for waning attention, which is consistent with an executive syndrome. It has been established that people with frontal brain injuries find it more difficult to sustain their attention, inhibit responses, and become frustrated more easily (Wood, 2002). The individual analyses support these views, as it was observed that some participants found the feedback on the WMT distracting, plus the nature of the injuries necessary to warrant community rehabilitation suggests that monitoring performance and learning from feedback is particularly difficult for the Community subgroup. In addition, individual question analyses of the WMT indicated semantic relatedness effects within the incorrect answers. For example, Community participant 2 was able to correctly identify fire from 'fire/flame' on the Immediate Recognition trial of the WMT, but failed to identify fire from 'fire/smoke' on the Delayed Recognition trial; this continued with the Multiple Choice memory measure, where 'smoke' was chosen, and with Paired Associates, where 'alarm' was chosen. It is possible that the participant developed a gist based representation that was able to be overridden during the Immediate Recognition trial, but once specific-item recollection had deteriorated then the gist representation could not be inhibited and false recognition occurred. Future research could establish just how common this finding is on the WMT in people who have frontal lobe lesions.

The results from the Inpatient sample also suggest that being able to maintain attention and concentration is a significant factor in performing well on tests of effort, whereas failures in the Epilepsy subgroup seemed to originate from younger men struggling with their current circumstances, with a flatness of affect that could belie lack of interest or identity management rather than purely anxiety or depression.

Research suggests that divided attention and distraction interferes with encoding and ability to complete tasks (Schmitter-Edgecombe & Nissley, 2000; Watt, Shores, & Kinoshita, 1999), and this is particularly pronounced in people with a brain injury due to an increased inability to inhibit distractions (Knight, Titov & Crawford, 2006). Batt et al. (2008) showed that performance on the WMT is affected by distraction, and conclude that the WMT is thus measuring ability as well as effort. This is not consistent with the view that the WMT requires effort, but little or no 'ability' (Green et al., 2002). Batt et al. (2008) also found a relationship with pre-morbid intelligence, but the current research was unable to establish a difference in pre-morbid IQ between those who passed and failed effort tests. However, several of the participant's pre-morbid IQ scores could not be calculated. Despite the numerous measures included in the current study, a significant drawback was the lack of any thorough neuropsychological data such as current intelligence and memory functioning. It could therefore be that failure rates within the current research are an accurate reflection of those identified by Dean et al. (2008), who found that the cut-offs on some widely used effort tests are inappropriate for people with low IQ levels. Ideally, future research will assess current IQ levels to more thoroughly investigate the suggestion that WMT results are influenced by cognitive ability.

Green et al. (2001) imply that effort has such a large effect that it "literally inverts" (p. 1059) the patterns of scores expected on the basis severity of injury. Therefore, the authors state that mildly injured patients are supposed to do worse than more severely injured patients when secondary gains are present and perform at the same level when they are not. Although no formal severity rating was calculated in the current research the participant subgroups are assumed to differ in the severity of their injuries. With regard to the WMT, participants in the Epilepsy group scored higher than the Community and Inpatient participants on the WMT-IR and WMT-DR, although this did not reach significance. However, statistically significant differences were found with regard to the AMI and the CMTF, with Community participants failing these tests at much higher rates than the Epilepsy participants. Such findings provide an indication that people with more severe or more recent brain injuries may fail some effort tests at higher rates than people with milder injuries when no secondary gains are present.

Demographic and psychological differences across subgroups of participants could also contribute to variability in effort test scores. For example, the participants in the Epilepsy sample were significantly younger, more anxious, and had better memory scores than the participants in the Community group. 23.4% of the overall sample were above the cut-off for anxiety and 36.2% were above the cut-off for depression. However, subgroup analyses revealed that 62.5% of the Epilepsy sample were anxious and 21.1% were depressed, compared to anxiety levels of 20% and depression levels of 15% in the Community and levels of 27.3% for both anxiety and depression in Inpatients. It is unclear why the participants with Epilepsy were so much more anxious than the other two subgroups, although most of the Epilepsy participants were undergoing continued monitoring in anticipation of receiving surgery in the near future. There were also more men than women in the research. However, this is representative of the finding that more men experience brain injuries than women, with 70-88% of all people that sustain a head injury being male (Department of Health, 2001), so it is unlikely that such a finding is contributing to rates of failure on the effort tests.

4.4 Clinical Implications of the Study Findings

The findings from the current research have a number of implications in relation to the clinical assessment and management of people with brain injuries. Specifically, such results suggest that:

- Clinicians need to be wary of interpreting the results of failures on effort tests as evidence of feigning in the populations identified in this study.

- Although the precise contribution of other factors in addition to effort is unclear, the current research suggests that anterior brain lesions could be contributing to failure rates in this population, along with ability to sustain attention, and other response styles suggested by Rogers (2008a) such as impression management. Memory, processing speed and depression were also linked to some of the effort test scores. Therefore, there are certain factors that may warrant further investigation if effort test failure is observed in clinical practice, particularly processing speed, neurobehavioural presentation, and response styles.

- Clinicians need to carefully consider whether failure on effort tests is due to brain damage or poor effort. If the patient fails effort tests due to insufficient effort then scores on other neuropsychological tests would be invalid, but if the patient fails due to brain damage then the scores on other tests would be an accurate reflection of abilities. It appears that when interpreting any test of effort, full consideration should be given to the cognitive abilities that may be involved, such as memory, attention, and language, and to the context of assessment, such as external incentives, consistency of reports, observations and behaviours.

- In line with both the Slick et al. (1999) criteria and the NAN guidelines, this research highlights the importance of clinicians not relying on the results of just one effort test when assessing effort within clinical settings, and consideration must be given to performance of multiple effort tests within populations similar to those assessed in the current research in order to increase specificity.

- The cut-off scores may need to be adjusted for the effort measures used in the current research in order to increase specificity. For example, scores of <8 on the AMI produce a specificity rate of 100% in this population. However, even if the cut-off scores are adjusted to increase specificity and avoid the possibility of false positives, this means that sensitivity would reduce and people who do not have a credible performance would be less likely to be identified.

- Clinicians need to place a greater emphasis on explaining the importance of testing from the client's perspective so that they are more invested in and focussed on the process, as clinicians may be significantly overestimating how interested or motivated clinical patients are during assessment.

- Notwithstanding the indication that brain damage may affect performance on the WMT in a severely brain-injured population, the WMT probably remains a valid measure of effort in people with much milder injuries. Therefore if people with much milder injuries were scoring below the levels obtained in this study then there would be questions raised regarding their performance. As a result, passing or failing such effort tests could also have implications for treatment and rehabilitation, as those who perform poorly on neuropsychological tests despite passing effort tests may require different rehabilitation strategies than those who perform at the same cognitive level, but fail effort tests.

Although effort tests may not be as effective as first thought at evaluating patients with legitimate brain injuries, it is still important for clinicians to recognise the impact that reduced effort can have and must consider ways of assessing it within their practice. McCarter et al (2009) revealed that lots of clinicians within the UK are not including such measures in their assessments. For example, 26% indicated that they did not use tests in clinical samples due to the perceived low base rate, and 11% reported using their own idiosyncratic methods and unpublished tests. Both of these are potentially dangerous situations, as this research has already indicated that even supposedly well validated tests do not have the required evidence available to unequivocally support their use in all populations, and idiosyncratic methods could be even more flawed. In addition, despite the perceived low base rate of reduced effort in clinical samples, knowing whether someone is able to put in their best effort is vital for understanding the results of all other tests.

4.5 Limitations of the Study

Whilst the research does have some notable strengths, a number of weaknesses can also be identified that are outlined in the next section.

4.5.1 Design considerations

A factor that could have affected the results of the current research is the different environments that testing took place in. Administration was carried out within hospital wards, in hospital clinics, in people's own homes, and in community supported housing, as well as being at different times of day. Therefore, a number of extraneous factors could have affected the results. For example, the hospital ward was often noisier than the outpatient clinic, and considering that the Inpatient participants had newly acquired injuries that would increase their distractibility, they would have potentially found assessment more difficult than they would have if testing was in a quieter environment. However, the current research attempted to keep such extraneous factors to a minimum through administering tests in a quiet room away from the main ward.

The chief investigator was not blinded to study hypotheses and methodology and administered all tests, and this may have unduly influenced participant responses. However, this helped to maintain consistency of assessment within the Epilepsy and Community subgroups, and ensured participants were provided with appropriate feedback if they were having difficulties with any of the procedures. In addition, testing directions were written down in detail and followed as closely as possible so all participants received the same guidance and prompts. However, the limited test administration within the Inpatient group is also a significant limitation of the study, as not all participants received the same tests and so consistency across administrations was not maintained.

The representativeness of the sample may also have influenced the results. Several subgroups of clinical populations with varying injuries were included in the study and analysed as an overall brain injury group. However, only three subgroup populations were recruited, and it is possible that if other populations were assessed (e.g. people with multiple sclerosis or those attending outpatient community rehabilitation) then the results of the current study could be different. However, it is likely that the population used in the current research was reasonably representative of people who have a head injury by the fact that they were drawn directly from clinical populations presenting to services.

In addition, due to the nature of the sample it was difficult to gain access to the required populations. Access to all patients required communication with consultants, ward managers and care workers. However, collaboration with staff was often the reason for failure to recruit, particularly from the Inpatient ward where staff found it difficult to find the time to identify suitable participants based on the relevant inclusion and exclusion criteria. Therefore, it is likely that consultants and staff varied in their judgements regarding who was suitable and who was asked to participate. A further difficulty regarding the representativeness of the sample was that the ethics committee also required that patients be approached by someone they were already working with before being approached by the researcher. This was to minimise any potential undue pressure to take part, but did reduce the number of participants contacted. There were also some refusals to take part across all subgroups of participants, and it is possible that those who refused were different in some way from the population who were assessed. However, people who did not wish to take part in the research were limited, and those who did refuse mainly gave reasons associated with being too busy. This is reflected in the Epilepsy sample, who were generally living independently and had work commitments, and they refused at higher rates than the other subgroups.

As mentioned previously, based on Green's (2003) descriptions it is difficult to know if the population studied in the current research matched that of the genuinely brain injured sample used to validate the WMT. It is hard to isolate scores within the WMT manual, as the studies often use a differential prevalence design to identify people who pass and fail during litigation, and then analyse the differences between these samples on a post-hoc basis. In addition, the definite brain injury group in the validation research had reported Post-Traumatic Amnesia of 16.4 days (S.D = 24). However, it was not possible to identify such scores for the current sample and it could be that the participants in the current research were much more impaired. It would be very useful to identify PTA and GCS levels in any future studies so that severity levels across various studies can be compared.

The nature of brain injury and location of brain injury were also difficult variables to assign participants to for a variety of reasons. Specifically, there are problems with the comorbidity of brain injuries with epilepsy, epilepsy surgery and tumour surgery. Also, the information available regarding the location of injury was variable, as the Community subgroup notes were not as detailed as the Inpatient and Epilepsy groups. Specific location of injury was also more likely to be recorded in the Epilepsy subgroup because the origin of seizures is more specific (i.e. left or right temporal lobe), whereas people with traumatic brain injuries are more likely to have widespread or diffuse damage with specific regions highlighted for consideration depending on brain scans and behavioural presentation (i.e. anterior brain lesions). For example, within the individual analyses section Community Participant 1 had reports of more general injuries including 'widespread atrophy', whereas Epilepsy Participant 8 had more focussed information regarding injury, specifically a tumour in the right temporal lobe.

The order of the tests was not counterbalanced in the current study because in research involving correlations it is best to standardise the order. Order of test administration can influence results (e.g. through fatigue or boredom), but randomising the order means that this variance is spread across the tasks so it is not clear where it lies, and in the current research counterbalancing may have masked the relationships that the research was interested in identifying. The standardised order could have influenced the results such as those for the Camden memory Test for Faces because this measure was administered last and therefore the risk of fatigue or boredom could have been increased. It may be that the inclusion of lots of measures within the same testing session was overwhelming, particularly for the Community participants. However, this is unlikely to have influenced WMT scores because the test is specifically designed with a half hour gap between administrations that is usually filled with other tests during a standard assessment in medico-legal settings. It would make the test highly impractical if clinicians were not able to administer further tests within this time frame. The Inpatient participants also had a break between the two WMT effort test administrations in which no other tests were performed, so testing fatigue is unlikely to account for the findings in this subgroup.

4.5.2 Testing considerations

An advantage of having one researcher assess all of the participants included in the research was that any difficulties with the measures were highlighted. For example, despite being preset to a volume judged to be reasonable by the researcher the WMT feedback noise was very off-putting for some of the participants and the volume was turned to a barely audible level after administration to the first few participants. Although this provides clinically relevant information about the participants it could also have affected the results, and it is possible that turning the feedback noise off altogether may influence the findings in patients who were already distractible and easily frustrated.

As previously reported, it is possible that questions included on measures such as the AMI have meant that the measure itself has changed in utility over time. For example, technology has advanced to act as an aid for people's impaired memory, and people are now more reliant on mobile telephones to store and access numbers that they would once have needed to remember. Also, only one short questionnaire measure was used to identify anxiety and depression in the current sample, and this indicated significantly high rates. It is possible that a more in depth analysis of these factors could change the nature of the relationship with effort test scores.

The Brief Effort Questionnaire included at the end of the research was also affected by the nature of participant injuries. It was clear that a number of the participants were not sure what 'effort' meant, and so required additional instructions such as 'trying your best'. As the construct of 'effort' is not clear cut for participants or for researchers the answers to these questions were likely based on different concepts for different participants. As clinicians need to ensure that patients put in their best effort during testing it would be useful to find out what patients actually understand by this term. This could help to provide clearer and more straightforward ways to explain the importance of effort to participants before testing begins.

Although the current study included a variety of measures of effort, the additional effort tests used to assist in the evaluation of construct validity are also not considered to be amongst the most sensitive or specific (McCarter et al. 2009). It would be interesting to evaluate effort test performance in this group with another effort measure that has received equal research and publicity, such as the TOMM (Tombaugh, 1996).

Finally, people with brain injuries often have other co-morbid conditions. For example, people with brain injuries are known to experience higher levels of fatigue. Testing was monitored and adapted to account for this where necessary, but an objective measure of such variables was not taken and therefore it is not clear if it affected the results obtained from effort tests. However, the participants were drawn from genuine clinical populations, so any effect of factors such as fatigue is likely to be representative of people with brain injuries similar to those observed in the participants recruited to this study.

4.5.3 Statistical analyses

While most of the data conformed to the statistical assumptions required for the tests conducted on the data, some variables were identified as not normally distributed, some variables could not be used in analyses due to ceiling effects (i.e. Coin-in Hand test), some variables could not be used in particular analyses due to the nature of the data (e.g. ordinal scales with too few categories), and data for some variables was not collected for the Inpatient sample.

It is interesting to note that WMT-IR was not normally distributed, but the WMT-DR was. Such a finding indicates variability in these measures, as the assumption would be for both to be non-normally distributed and all participants to score near ceiling levels. As Green et al. (1999) point out, "Scores on the measures of biased responding on the WMT ... are not normally distributed; they show a strong ceiling effect because they are minimally sensitive to variations in ability." (p. 816). However, the WMT-IR variable contained the two participants whose administrations were halted, and it is likely that it was these low scores that skewed the data. The distribution of scores across both the WMT-IR and WMT-DR indicate that something in addition to effort could be being measured within this particular population.

Within the analyses, although some significant relationships were identified between effort test scores and the additional variables included in the study, limited relationships were identified for the WMT. The patterns observed were in accordance with previous independent research conducted with the WMT, but they were not demonstrated statistically, with the exception of the WMT memory measures and small correlations with age and processing speed. Such a finding could support the view of the test authors, in that the WMT is actually measuring something other than genuine cognitive impairment, but it could also indicate that the WMT has limited range and is not a very sensitive measure. However, despite the relatively large sample size in the current study, it is still possible that it was simply too small to identify any relationships between variables. Although the sample was a genuine clinical population with no incentives to feign, the subgroups within this population may have different reasons for failing the tests, as indicated in the results section containing individual analyses. The Inpatient participants were also not administered several of the additional measures required to draw conclusions about performance (e.g. processing speed and memory). Therefore there may not have been enough participants in the subgroups or across particular variables.

Some of the findings in the research were also based on different types of analyses. Initial pass/fail rates across effort tests were judged the most suitable data to consider in relation to the other variables, but scaled scores were also used to identify any potential relationships, particularly considering that the pass/fail rates appear unsuitable for use in the current population. Although this allowed the data to be thoroughly examined from different statistical perspectives, the conclusions drawn based on scaled scores are not directly applicable to 'failure' rates as specified by standardised cut-offs. There is also a probability of Type I error due to all the indices evaluated. However, Bonferroni corrections were applied where appropriate within the statistical analyses, and the number of failures is still more than would be expected through chance alone given that the measures are designed not to be failed by people with genuine impairments.

4.6 Recommendations for Future Research

The vast majority of the literature conducted with people who have brain injuries has been with people within litigation. Many unanswered questions remain about the performance of effort tests across populations other than those in litigation, and to what extent such tests measure variables in addition to effort. There has been virtually no previous research to date focussing on independent evaluations of the WMT in genuinely impaired populations with no incentive to feign. Although the current research has provided some key insights into the levels of effort test failure within genuine brain injured populations, further psychometric validation is urgently required in larger and different clinical groups in order to fully establish whether the WMT is measuring what it is purported to measure and is accurately classifying clinical populations.

Further research is also required to investigate performance on effort tests with larger samples of the subgroups assessed in the current study and in those where the severity of brain injuries/epilepsy has been more clearly established. Such research is vital to ensure that clinicians do not make a Type I error and conclude that a genuine patient is feigning when the results actually reflect a false positive. The current findings suggest that clinicians could be at risk of making such an error when using tests of effort with the populations assessed in the current research.

The current study also highlights the importance of administering multiple measures to assess effort, but it also shows that even multiple measures can potentially produce false positives. Further research could focus on developing cut-offs that are more suitable for use with people who have more severe brain injuries and cognitive impairments. There is also clinical utility in identifying those who struggle to perform well on these tests, as the very simple nature of the task would indicate that any people who fail for genuine reasons are unlikely to perform well on other neuropsychological tests.

Research including more comprehensive neuropsychological data will be of particular value when attempting to establish clear reasons for effort test failure. Longitudinal studies that collect comprehensive data over time would be a useful contribution. Retest performance on effort tests within genuine clinical populations would also be valuable to investigate, as pass and fail rates may vary over time. The current research has also only focussed on the assessment of tests based on feigning memory impairments. However, people are capable of feigning other symptoms within neuropsychological assessment, such as perceptual or motor difficulties (e.g. Greiffenstein, 2007). Future research will need to ensure that tests being developed to identify people who feign such difficulties are equally valid for use in people who have genuine impairments.

An increased focus on qualitative data such as that collated from the Brief Effort Questionnaire may also provide clearer information from a more subjective, patient perspective to inform clinicians of the reasons why people may not perform well on these tests. Further research could interview or administer questionnaires to genuine participants who have completed neuropsychological and effort testing to gather information regarding response styles during testing; this could provide information that standardised measures cannot. Further studies could also look in more detail at the structure of the WMT, as it seems plausible that people with frontal lesions may struggle on the WMT task as a result of the semantic relatedness of the test items. This would result in more false recognition of incorrect items for genuine reasons such as a failure to monitor their performance and inhibit responses.

It is possible that some of the participants in the current study were not putting in their full effort even though the sample were screened for obvious external incentives to feign and were aware that the results were confidential. It could be that the continued receiving of benefits or care is enough. Therefore, some of the results could be true positives. Alternatively, it may be that clinicians are not fully aware of all the potential secondary gains, such as maintaining autonomy and a sense of identity or minimising avoidable pain. Future research into potential alternative response styles which could lead to failures on effort tests in addition to malingering will be very valuable within clinical settings, where any secondary gains or potential response styles are not as clearly identifiable.

4.7 Conclusions

Despite evidence to suggest that effort tests such as the WMT are measuring the amount of 'effort' an individual puts into a task rather than cognitive 'ability' this study has provided evidence that people with genuine brain injuries and no incentive to feign fail effort tests at much higher rates than previously suggested. Statistical analyses showed relationships between a number of the effort tests and various demographic, injury, psychological and neuropsychological variables. This supports the concept of effort test failure as multifaceted and potentially reliant on factors in addition to effort in populations such as those included in the current research. Whilst the exploratory nature of the study cautions against over-interpretation of the results, this research does suggest that clinicians need to be wary when using established cut-off scores to interpret effort test data.

REFERENCES

Allen, M. D., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2007). Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives. *Brain Injury, 21*, 1425-1428.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders – DSM-IV-TR.* Arlington, VA: Author

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2006). Base Rates of Negative Response Bias and Malingered Neurocognitive Dysfunction among Criminal Defendants Referred for Neuropsychological Evaluation. *The Clinical Neuropsychologist, 21*(6), 899-916.

Baker, R., Donders, J., & Thompson, E. (2000). Assessment of incomplete effort with the California Verbal Learning Test. *Applied Neuropsychology, 7*, 111–114.

Baldessarini, R. J., Finkelstein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*, 569-573.

Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society, 14*, 1074–1080.

Bauer, L., & McCaffrey, R. J. (2006). Coverage of the TOMM, VSVT, and WMT on the Internet: Is test security threatened? *Archives of Clinical Psychology, 21*, 121-126.

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering Trial 1and Word Memory Test Immediate Recognition as screening tools for insufficient effort. *Assessment, 14*, 215–222.

Beck, A. T., Steer, R. A., & Brown, G. K. (*1996*). Manual for the *Beck Depression Inventory-II.* San Antonio, TX: Psychological Corporation.

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

Bender, S. D., & Rodgers, R. (2004). Detection of neurocognitive feigning: Development of a multi-strategy assessment. *Archives of Clinical Neuropsychology, 1*, 49–60.

Berry, D. T. R. & Schipper, L. J. (2007). Detection of feigned psychiatric symptoms during forensic neuropsychological examinations. In G. J. Larrabee (Ed.) *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp. 226-263).

Berry, D. T. R., Adams, J. J., Clark, C. D., Thacker, S. R., Burger, T. L., Wetter, M. W., Baer, R. A., & Borden, J. W. (1996). Detection of a cry for help on the MMPI-2: An analog investigation. *Journal of Personality Assessment, 67*(1), 26-36.

Bianchini, K. J., Greve, K. W., & Love, J. M. (2003). Definite malingered neurocognitive dysfunction in moderate/severe traumatic brain injury. *The Clinical neuropsychologist, 17*, 574-580.

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

Binder, L. M. (1993). *Portland Digit Recognition Test Manual-2nd Ed.* Beaverton, OR: Author.

Binder, L. M. (2002). The Portland Digit Recognition Test: A review of validation data and clinical use. *Journal of Forensic Neuropsychology, 2*, 27-41.

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

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research, 52*, 69-77.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.) *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective.* London: The Guilford Press. (pp29-49).

Boone, K. B., & and Lu, P. (2003). Noncredible Cognitive Performance in the Context of Severe Brain Injury. The Clinical Neuropsychologist, 17, 244-254.

Boone, K. B., & Lu, P. H. (2007). Non-forced choice effort measures. In G. J. Larrabee (Ed.). *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp. 27-43).

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The b Test: Manual.* Los Angeles: Western Psychological Services.

Boone, K. B., Lu, P., Sherman, D., Palmer, B., Back, C., Shamieh, E., Warner-Chacon, K., & Berman, N. (2000). Validation of a new technique to detect malingering of cognitive symptoms: The b test. *Archives of Clinical Neuropsychology, 15,* 227–241.

Bowden, S. C., Shores, A. E. & Mathias, J. L. (2006). Does effort suppress cognition after traumatic brain injury? A re-examination of the evidence for the word memory test. *The Clinical Neuropsychologist, 20,* 4, 858-872.

British Psychological Society (2009). *Assessment of effort in cognitive testing.* [unpublished consultation document]. Retrieved 20 June, 2009, from British Psychological Society: Official Site, Web site: http://www.bps.org.uk/blog/2/categories.cfm?catid=105

Brockhaus, R., & Merten, T. (2004). Neuropsychologische diagnostik suboptimalen: Leistungsverhaltens mit dem Word Memory Test [in German]. *Nervenarzt, 75,* 882–887.

Budson, A. E., Sullivan, A. L., Mayer, E., Daffner, K. R., Black, P. M., & Schacter, D. L. (2002). Supression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain, 125,* 2750-2765.

Bunnage, M., Eichinger, C., Pearce, N., Duckworth, A., & Newson, M. (2008). Criterion validity of the Word Memory Test: An audit of a sample of patients assessed for clinical, not litigious, reasons. [Proceedings of the 36th Annual Meeting of International Neuropsychological Society, Hawaii, February 2008 Abstract]. *Journal of International Neuropsychological Society, 14, (Suppl. 1),* 138–139.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20,* 419–426.

Carsky, M., Selzer, M. A., Terkelsen, K., & Hurt, S. W. (1992). The PEH. A questionnaire to assess acknowledgment of psychiatric illness. *The Journal of Nervous and Mental Disease. 180,* 458-464.

Cercy, S. P., Schretlen, D. J., & Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomena. In R. Rogers (Ed.). *Clinical assessment of malingering and deception.* (2nd ed.). New York: Guilford Press. (pp.108-129).

Chafetz, M. D., & Abrahams, J. (2005). Green's MACT helps identify internal predictors of effort in the social security disability exam. *Poster session presented at the annual meeting of the national academy of neuropsychology.*

Cochrane, H. J., Baker, G. A., & Meudell, P. R. (1998). Simulating a memory impairment: Can amnesiacs implicitly outperform simulators. *British Journal of Clinical Psychology, 37,* 31–48.

Conder, R., Allen, L., & Cox, D. (1992). Manual for the Computerized Assessment of Response Bias. Durham, NC: CogShell.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20,* 191–198.

Coughlan, A. K., Oddy, M. J., & Crawford, J. R. (2007). *BIRT Memory and Information Processing Battery (BMIPB).* London: Brain Injury Rehabilitation Trust.

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 22,* 705-722.

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

Department of Health. (2001). *Hospital episode statistics 2000-2001.* London: Department of Health.

Dodrill, C. B. (2008). Do patients with psychogenic nonepileptic seizures produce trustworthy findings on neuropsychological tests? *Epilepsia, 49*(2), 369–370.

Donders, J. & Boonstra, T. (2007). Correlates of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury, 21,* 319–326.

Doyle, A. C. (1927). *The Case-Book of Sherlock Holmes: The problem at Thor Bridge.* London: John Murray.

Drane, D. L., Williamson, D. J., Stroup, E. S., Holmes, M. D., Jung, M., Koerner, E., Chaytor, N., Wilensky, A. J., & Miller, J. W. (2006) Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia, 47,* 1879–1886.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*(5), 508-515.

Flaro, L., Green, P., & Robertson, E. (2007) Word memory test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury, 21*(4), 373-383.

Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19,* 465–473.

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology, 44,* 1013-1023.

Glaros, A., & Kline, R. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity and predictive value models. *Journal of Clinical Psychology, 44,* 1013-1023.

Goldberg, H. E., Back-Madruga, C., & Boone, K. B. (2007). In K. B. Boone (Ed.) *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective.* London: The Guilford Press. (pp. 281-309).

Gorissen, M., Sanz de la Torre, J. C. & Schmand, B. (2003). Effort in the neuropsychological evaluation of schizophrenia (abstract). *Journal of International Neuropsychological Society, 9* (4), 512.

Gorissen, M., Sanz de la Torre., J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research, 78,* 199-208.

Gorlyn, M., Keilp, J.G., Oquendo, M.A., Burke, A.K., Sackeim, H.A., & Mann, J.J. (2006). The WAIS-III and major depression: Absence of VIQ/PIQ differences. *Journal of Clinical and Experimental Neuropsychology, 28, 1145-1157.*

Green, P. (2003). *Word Memory Test for Windows: User's Manual and Program.* Seattle: Green's Publishing.

Green, P. & Allen, L. M. (1999) *Performance of neurological patients on CARB and the WMT. Supplement to the CARB and WMT manuals.* Durham, NC: CogniSyst.

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A users guide to the oral and computer-administered forms.* Durham, NC: CogniSyst.

Green, P., Iverson, G. L., & Allen, L. M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13,* 813–819.

Green, P., Lees-Haley, P. R., & Allen, L. M. (2002). The word memory test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 3,* 97–124.

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

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

Grieffenstein, M. F. (2007). Motor, Sensory, and Perceptual-Motor Pseudoabnormalities. In G. L. Larrabee (Ed.) *Assessment of malingered neuropsychological deficits.* New York: Oxford University Press. (pp. 100-130).

Grimes. D. A., & Schultz. K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet, 365*, 1500–1505.

Grote, C. L., & Hook, J. N. (2007). Forced-choice recognition tests of malingering. In G. J. Larrabee (Ed.) *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp. 44-79).

Gunstad, J. & Suhr, J. A. (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: Vulnerability to coaching. *The Clinical Neuropsychologist, 15*, 397-404.

Gunstad, J., & Suhr, J. A. (2001). Efficacy of the full and abbreviated forms of the Portland Digit Recognition Test: Vulnerability to coaching. *The Clinical Neuropsychologist, 15*, 397-404.

Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology Review, 5*(2), 125-148.

Halligan, P. W., Bass, C., & Oakley, D. A. (2003). Malingering and Illness Deception. Oxford, U.K.: Oxford University Press.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing: neuropsychological malingering and the word memory test. *Archives of Clinical Neuropsychology, 17*, 709–714.

Heaton, R. K., Grant, I., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. In I. Grant & K. M. Adams (Eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders.* New York: Oxford University Press. (pp. 100-120).

Henry, G. K. (2005). Probable *malingering* and performance on the test of variables of attention. *The Clinical Neuropsychologist, 19*, 121–129.

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

Iverson, G. L, & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18*, 38-51.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 15*, 829–858.

Iverson, G. L., Slick, D. J., & Franzen, M. D. (2000). Evaluation of a WMS–R malingering index in a nonlitigating clinical sample. *Journal of Clinical and Experimental Neuropsychology, 22*, 191–197.

Iverson, G. L., & Franzen, M. D. (1994). The Recognition Memory Test, Digit Span, and Knox Cube Test as markers of malingered memory impairment. *Assessment, 1*, 323–334.

Jaccard, J. & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage Publications.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia, 27*, 1043-1056.

Joy, S., Kaplan, E., & Fein, D. (2004). Speed and memory in the WAIS-III Digit Symbol-Coding subtest across the adult lifespan. *Archives of Clinical Neuropsychology, 19*, 759-767.

Kapur, N. (1994). The coin-in-the-hand test: A new 'bedside' test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery and Psychiatry, 57*, 385-386.

Kelly, P. J., Baker, G. A., van den Broek, M. D., Jackson, H., & Humphries, G. (2005).The detection of malingering in memory performance: The sensitivity and specificity of four measures in a UK population. *British Journal of Clinical Psychology, 44*, 333–341.

Kemp, S., Coughlan, A. K., Rowbottom, C., Wilkinson, K., Teggart, V., & Baker, G. (2008). The base rate of effort test failure in patients with medically unexplained symptoms. *Journal of Psychosomatic Research, 65*, 319–325.

Knight, R.G., Titov, N., & Crawford, M. (2006). The effects of distraction on prospective remembering following traumatic brain injury assessed in a simulated naturalistic environment. *Journal of the International Neuropsychological Society, 12*, 8–16.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 410–425.

Larrabee, G. J. (2007). Identification of malingering by pattern analysis on neuropsychological tests. In G. J. Larrabee (Ed.) *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp. 80-99).

Larrabee, G. J., & Berry, D. R. (2007). Diagnostic classification statistics and diagnostic validity of malingering assessment. In G. J. Larrabee (Ed.) *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp. 14-26).

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.) *Assessment of Malingered Neuropsychological Deficits.* New York: Oxford University Press. (pp 334-371).

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two component model. *Psychological Bulletin, 107*, 34-47.

Lishman, W. A. (1997). *Organic psychiatry* (3rd ed.). Oxford: Blackwell.

Lynch, W. J. (2004). Determination of effort level, exaggeration, and malingering in neurocognitive assessment. *Journal of Head Trauma Rehabilitation, 19*, 277–283.

Macciocchi, S. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology, 21*, 395–404.

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort Testing in Contemporary UK Neuropsychological Practice. *The Clinical Neuropsychologist, 20*, 1-17.

Meehl, P. E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216.

Merten, T., Bossink, L., & Schmand, B. (2006). On the limits of effort testing: symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*, 308-318.

Meyers, J. E. & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18*, 261-276.

Miller, S. (1984). *Experimental design and statistics.* London: Methuen.

Mittenberg, W. & Morgan, D. (2009) Mild traumatic brain injury in civil litigation. In J. E. Morgan and J. J. Sweet (Eds.). *Neuropsychology of Malingering Casebook.* East Sussex: Psychology Press.

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

Mittenberg, W., Puentes, G. A., Patton, C., Canyock E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 2*, 227-240.

Parkin, A. J., Bindschaedler, C., Harsent, L., & Metzler, C. (1996). Pathological false alarm rates following damage to the left frontal cortex. *Brain and Cognition, 32*, 14-27.

Parkin, A. J., Ward, J., Bindschaedler, C., Squires, E. J, & Powell, G. (1999). False recognition following frontal lobe damage: the role of encoding factors. *Cognitive Neuropsychology, 16*, 243-265.

Pernicano, K.M. (1986). Score differences in WAIS-R scatter for schizophrenics, depressives, and personality disorders: a preliminary analysis. *Psychological Reports, 59*, 539 543.

Rey, A. (1941). L'examen psychologie dans las cas d'encephalopathologie traumatique. *Archives de Psychologie, 28*, 286-340.

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

Richman, J., Green, P., & Gervais, R. (2006). Objective tests of symptom exaggeration in independent medical examinations. *Journal of Occupational and Environmental Medicine, 48,* 303-11.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory, & Cognition, 21, 803-814.

Rogers, R. (1997). Introduction. In R. Rogers (Ed.). *Clinical Assessment of Malingering and Deception* (2nd ed.). New York: Guilford Press. (pp. 1-19).

Rogers, R. (2008a). An introduction to response styles. In R. Rogers. (Ed.). *Clinical Assessment of Malingering and Deception.* (3rd Edition). New York: The Guilford Press. (pp. 3-13).

Rogers, R. (2008b). Detection strategies for malingering and defensiveness. In R. Rogers. (Ed.). *Clinical Assessment of Malingering and Deception.* (3rd Edition). New York: The Guilford Press. (pp. 14-38).

Rogers, R., Jackson, R. L., & Kaminski, P. L. (2004). Factitious psychological disorders: The overlooked response style in forensic evaluations. *Journal of Forensic Psychology Practice, 3,* 115-129.

Rose, F. E., Hall, S., Szalda-Petree, A. D., & Bach, P. J. (1998). A comparison of four tests of malingering and the effects of coaching. *Archives of Clinical Neuropsychology, 13,* 349–363.

Rosenhan, D. (1973). On being sane in insane places. *Science, 172,* 250-258.

Ross, S. R., Putnam, S. H, & Adams, K. M. (2006). Psychological disturbance, incomplete effort, and compensation-seeking status as predictors of neuropsychological test performance in head injury. *Journal Clinical and Experimental Neuropsychology, 28,* 111-125.

Rugg, M. D., Fletcher, P. C., Firth, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain, 119,* 2073-2083.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). Edinburgh: Churchill Livingston.

Schacter, D. L., Verfaellie, M., Anes, M. D., & Racine, C. (1998). When true recognition suppresses false recognition: evidence from amnesic patients. *Journal of Cognitive Neuroscience, 10,* 668-679.

Schmitter-Edgecombe, M. & Nissley, H. M. (2000). Effects of divided attention on automatic and controlled components of memory after severe closed-head injury. *Neuropsychology, 14,* 559–569.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15,* 431-349.

Sharland, M. J., Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22,* 213–223.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18*(6), 911–922.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13,* 545–561.

Slick, D., Tan, J., Strauss, E., & Hultsch, D. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19,* 465–473.

Smith, G. E., Cerhan, J. H., & Ivnik, R. J. (2003). Diagnostic validity. In D. S. Tulsky & D. Saklofske (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III.* San Diego, CA: Academic Press. (pp. 273–301).

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2005). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). Edinburgh: Churchill Livingston.

Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. S., & Hultsch, D. F. (2002). Intraindividual variability as indicators of malingering of head injury. *Archives of Clinical Neuropsychology, 17*, 423–444.

Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15(5)*, 415-424.

Sweet, J. J. (1999). Malingering: Differential Diagnosis. In J. Sweet (Ed.). *Forensic Neuropsychology: Fundamentals and Practice.* Lisse, The Neatherlands: Swets & Zeitlinger.

Sweet, J., J., Condit, D. C., & Nelson, N. W. (2008). Feigned amnesia and memory loss. In R. Rogers. (Ed.). *Clinical Assessment of Malingering and Deception.* (3rd Edition). New York: The Guilford Press. (pp. 218-236).

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*, 495–505.

Tombaugh, T. N. (1996). *TOMM, Test of Memory Malingering.* New York: Multi Health Systems Inc.

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS-R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology, 16*, 597–607.

Trueblood, W., & Binder, L. M. (1997). Psychologist's accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology, 12*(1), 13-27.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology, 15*, 578–590.

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

Vickery, C. D. Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*, 37-48.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16*, 45-73.

Warrington, E. K. (1996). *The Camden Memory Test Battery*. Hove: Psychology Press.

Watt, S., Shores, E.A., & Kinoshita, S. (1999). Effects of reducing attentional resources on implicit and explicit memory after severe traumatic brain injury. *Neuropsychology, 13*, 338–349.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale* (3rd ed.). San Antonio: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading*. San Antonio: The Psychological Corporation.

Wehner, D. T., Ahlfors, S. P., & Mody, M. (2007). The influence of semantic processing on phonological decisions in children and adults: A magnetoencephalography (MEG) study. *Journal of Speech, Language, and Hearing Research, 50*, 716–731.

Wiggins, E., & Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behaviour, 12*, 57–78.

Wood, R. L. (2002). Understanding Neurobehavioural Disability. In R. L. Wood & T. M. McMillan (Eds.). *Neurobehavioral Disability and Social Handicap following Traumatic Brain Injury*. East Sussex, UK: Psychology Press. (pp. 3-28).

Wynkoop, T. F. & Denny, R. L. (2001). Test reviews: Computerised test of response bias (CTRB), Word Memory Test (WMT) and Memory Complaints Inventory (MCI). *Journal of Forensic Neuropsychology, 2*(1), 71–77.

Zigmond, A. S., Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67,* 361-370.

APPENDICES

5.1 Information Sheet

Information about the Research

You are being invited to take part in a research project. Before you decide, it is important to understand why the study is being done and what is involved. Please take a minute to read this brief information sheet.

Study title:

Memory Testing in Neurological Conditions

What is the purpose of the study?

People with neurological conditions sometimes have problems with their memory and concentration. A number of new assessments have been introduced to assess how serious these problems are and to identify whether you are able to put in your best effort during the assessments. However, very little work has been carried out looking at how people with neurological conditions perform on these new measures.

We are researching how people perform so that we can help doctors select the best tests for people at different stages of recovery.

Why have I been chosen?

You have been invited because you have been identified by your doctor as having a neurological condition. A total of 60 people will also be involved in the study.

What will happen if I take part?

If you do take part then we will ask you to try your best to complete some tests of memory and concentration. The tests will take 20-40 minutes, with an additional break in the middle.

Please note that the results of the tests are strictly anonymous and confidential. Your name will not be on the tests.

Assessment can take place at Chapel Allerton Hospital or St James' Hospital in Leeds, or at your home if more convenient. If you live at home, but would prefer to come to the hospital then you can receive travel expenses of up to £10.

At the end of the study we will prepare a brief summary of the results. This will be available if you would like to read it.

Further information:
Please contact Natalie Hampson, Psychologist in Clinical Training, Leeds University. Tel: 0113 343 2732

5.2 Consent Form

CONSENT FORM

Title of Project: Memory Testing in Neurological Conditions
Name of Researcher: Natalie Hampson

Please mark box

1. I confirm that I have read and understood the information sheet and this consent form. I have had the opportunity to consider the information, and ask questions. ☐

2. I understand that my participation is voluntary and I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

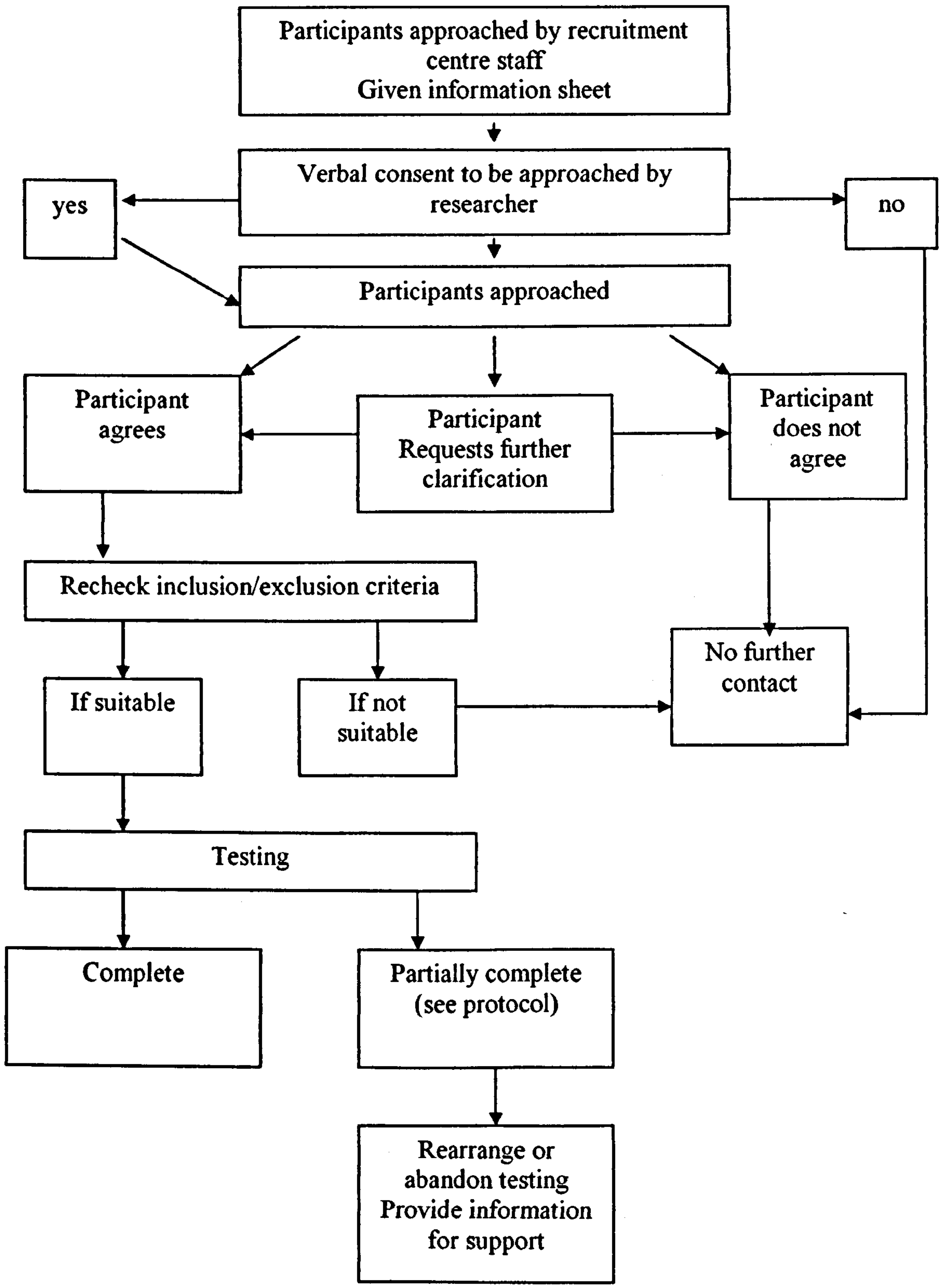
3. I understand that any written records will be used only for research and/or teaching purposes and that I will not be identified by name. ☐

4. I give permission form my medical records to be seen by the main researcher, who is a Psychologist in Clinical Training. ☐

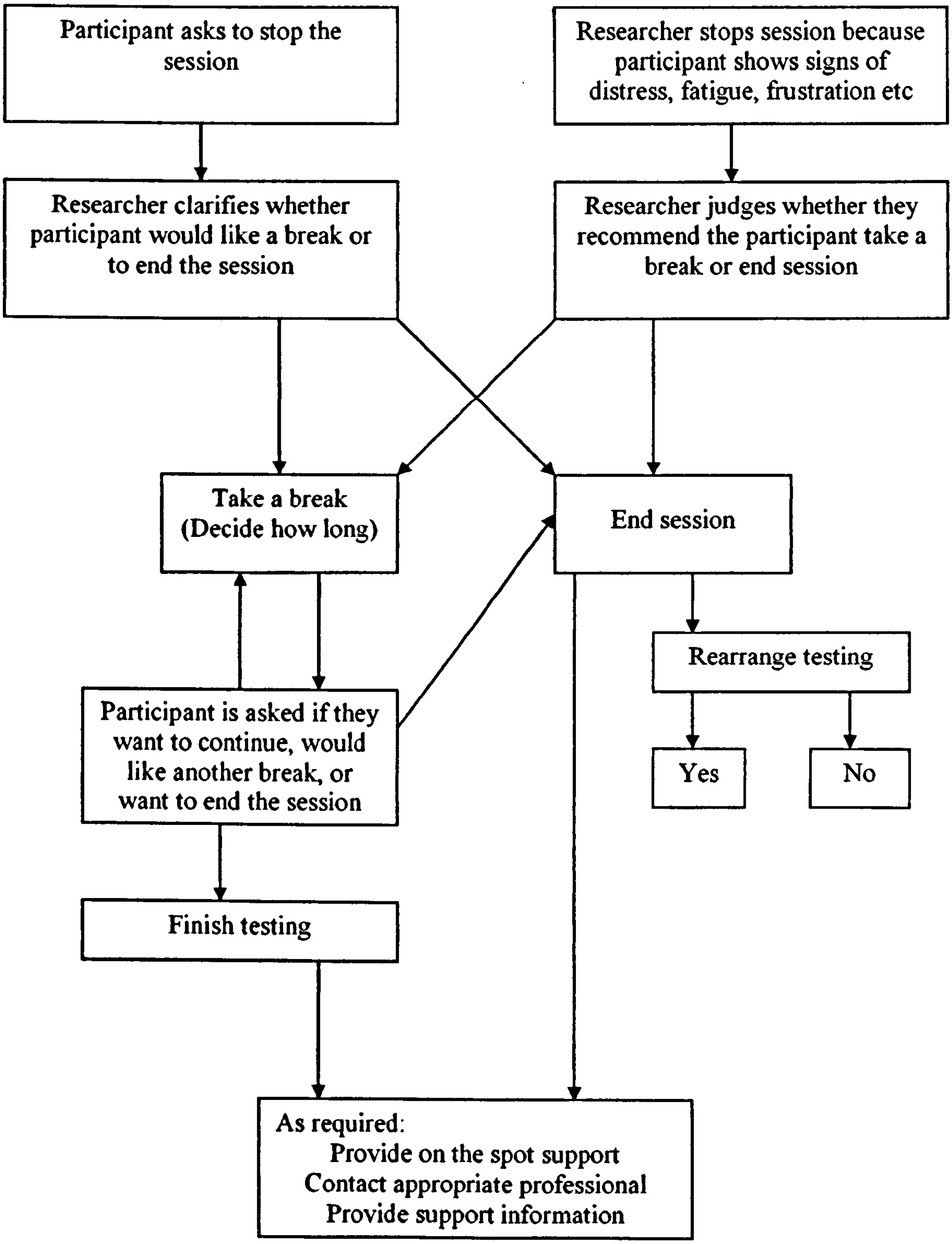
5. I agree to take part in the study. ☐

Name of Participant	Date	Signature
Name of Researcher	Date	Signature

5.3 Recruitment Protocol



5.4 Protocol for Stopping the Assessment



5.5 Additional Analyses

Table 30: T-test calculations between the current research and Green's (2003) validation research

Study	WMT-IR			WMT-DR		
	T	df	Sig.	T	df	Sig.
Moderate to severe brain injuries (N=57) vs current research (N=47)	3.94	102	p<.02	2.87	65	p<.02
Neurological with impaired memory (N=20) vs current research (N=47)	2.73	65	p<.02	2.23	65	p<.03
Neurological with normal memory (N=20) vs current research (N=47)	2.76	65	p<.02	2.94	65	p<.02