Malingering of Cognitive Symptoms

Philip G. Minoudis

D.Clin.Psy. Thesis (Volume 1), 2007

University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to thank the following:

Adrian Coxell for helping guide me through the project

Peter Scragg for his pragmatic advice and numerous read-through's

Raksha Khilosia for her help with data collection

My friends and family for their support throughout the course

Overview

Volume 1 is divided into 3 parts as follows:

Part 1 (Review Paper) discusses the admissibility of psychometric evidence of cognitive malingering in UK criminal law courts. The paper opens with a historical account of psychologists as expert witnesses, highlighting significant advances relevant to malingering. This sets the context for a discussion about current developments in policy and specifically the creation of a UK standard for the admissibility of scientific evidence. The penultimate section outlines the statistical and methodological issues which challenge the development of empirical cognitive measures of malingering. The paper closes with a discussion of future directions for research and practice in presenting psychological evidence in court.

Part 2 (Empirical Paper) reports on a study testing the utility of a battery of measures to identify simulating malingerers from healthy controls and psychiatric inpatients. The battery of measures were chosen for their different approaches to detecting malingerers. An additional qualitative interview was given to the simulating malingerers to investigate the strategies they used to fake the tests. The performance of the test battery was compared to a pre-existing screening tool for malingering. The results were discussed with reference to implications for research and practice.

Part 3 (Critical Appraisal) reflects on the process of undertaking the research. It discusses the generalisability of the findings when using a simulating malingering design, the utility of measuring reaction time to detect malingering, difficulties in the recruitment of inpatients, the array of choices in selecting the test battery and the clinical applications of the research.

Table of Contents

Table of tables

Part 1: Review paper: Review of the admissibility of psychometric evidence of cognitive malingering in criminal law courts

1.1 Introduction

This review provides a variety of perspectives on the admissibility of cognitive malingering tests in law courts in the United Kingdom (UK). In doing so, it spans the history of psychology in law courts, the current political debate on the development of guidelines for admissibility of scientific evidence and the statistical and methodological issues relevant to cognitive measures of malingering.

Justification for the inclusion of these perspectives is given. An historical perspective provides background context to current issues, whilst the current policy developments in this area give an indication of the various forces of influence which directly and indirectly affect clinical and forensic neuropsychologists. Statistical and methodological issues underpin the likelihood of a test standing up to challenges in court and are crucial to decisions on admissibility. The statistical robustness of a measure also guides which tests are selected by clinical and forensic psychologists. The review concludes with a discussion of future directions for both research and the practice of presenting psychological evidence in court.

1.2 History of psychologists as expert witnesses

"Voila nos experts!"

(Robespierre, cited from Kargon, 1986)

The year 1781 saw one of the earliest introductions of science into the law courts. Robespierre, then a little-known French lawyer, associated the term 'experts' with scientists in relation to Benjamin Franklin's irrefutable discovery of the lightning rod.

The defendant had erected a lightning rod on the chimney of his house which had caused much consternation amongst his neighbours. During this protracted case, Robespierre effectively represented both his defendant and the emergence of science as an influential force in society. In doing so, he challenged the assertion that judges should not hinder the advance of science (Kargon, 1986). In some ways, this has become a template for the relationship between science and the law, which remains ambivalent and occasionally discordant today.

Munsterberg (1899) cautioned that, "Peoples [sic] never learn from history" and it has been further suggested that without understanding where forensic psychology came from, it is hard to understand where it is going (Bartol & Bartol, 2006). Thus, an introduction to the history of the profession's emergence into the law courts is crucial to appreciate its present status and likely future progression. The following section outlines the history of psychologists as expert witnesses.

Modern forensic psychology began with the emergence of psychology into the law courts. Experimental investigations into the reliability of witness testimony were the first applications of psychology (Loh, 1981; Bartol & Bartol, 2006). Cattell's (1895) informal experiments on his students suggested a wide range of individual difference in degree of accuracy and level of confidence when recounting an event. This generated interest in the psychology of testimony. Binet (1900) replicated these experiments in France and summarised other findings in the field, calling for a "science psycho-judiciaire" (Binet, 1905). William Stern (1910) developed his 'Aussage' or remembrance experiments in which he found that recall was influenced by different ways of questioning and suggestion. This finding was generalised to unreliability of courtroom witnesses. As a result of these first experiments, in Germany and Europe generally, psychologists began to work as expert witnesses in criminal cases espousing both factual and opinion testimony (Bartol & Bartol, 2006).

Hale (1980) suggested that the earliest case of an expert psychologist was in 1896, when Albert Von Schrenk-Notzing testified at the trial of a Munich man accused of murdering three women. Significant press coverage led to the notion of "retroactive memory falsification". This was described as witnesses confusing what they had seen with what they had read in press reports.

1.2.1 Reaction time to detect deception

Between the turn of the twentieth century and World War One, interest emerged in studies of "guilt deception" in Germany, Austria and Switzerland. Wertheimer & Klein (1904) postulated that a person guilty of a crime creates a 'complex' in the mind which can be detected under the right conditions. One such detection strategy used a word association and reaction time task. One hundred words were presented to the suspect, consisting of 80 'innocuous' words and 20 words pertaining to a crime. Reaction time and associated word were recorded by the psychologist. When a relevant word was read out, it was hypothesised that the suspect would either reveal his/her guilt by providing an incriminating response relating to the crime, or that it would take significantly longer to inhibit an incriminating response and provide an innocuous association. This test showed considerable promise but was overshadowed by the introduction of the polygraph (Barland, 1988). Reaction time is still considered to be a potential discriminator for malingered responses (e.g. Vendemia et al., 2005).

1.2.2 Wigmore and Munsterberg: Professional dissonance and group averages

Hugo Munsterberg arrived to the US in 1892 championing the potential applications of psychology in a wide variety of fields. This culminated in his bestseller, 'On the Witness Stand' (Munsterberg, 1908) which promoted the use of psychology in law, particularly in areas of witness accuracy, jury persuasion, hypnosis and lie detection. Munsterberg's claims were accused of being exaggerated and not grounded in

empiricism (Wigmore, 1909). In a damning paper satirising a hypothetical libel case against Munsterberg, Wigmore (1909) called into question the utility of *Aussage* experiments to facilitate court verdicts. He identified that the results of testimonial accuracy were based on group averages. In court, however, it is the reliability of an individual witness which is called into question. As such, predictive percentages of testimonial error provide misleading results when used to support a verdict. The critical difference is emphasis on frequency of testimonial error of witnesses rather than on the impact of the unreliability on the trial outcome. The focus in court becomes reliability of testimony rather than that of the verdict. Wigmore (1937) highlighted an essential dissonance between social sciences and the law and, as a result, a period of inaction between psychology and the law ensued (Loh, 1981). The presentation of evidence in court remains a point of contention today and one that has led to unjust rulings in high profile cases (e.g. Science and Technology Committee, 2005), as discussed later in this review.

Between 1911 and 1924, Fernald and Healy began to use the Stanford-Binet Intelligence Scale to test for delinquency. From this, they developed their own set of 23 performance tests, and in 1924, Healy was possibly the first psychologist to present a cognitive assessment in court, which was used to help exonerate two juveniles of murder on the grounds of developmental difficulties (Healy, 1924).

1.2.3 Marston: Jury research and presentation of scientific evidence

William Marston, a tutee of Munsterberg, identified the positive correlation between systolic blood pressure and deception, which is the theoretical underpinning of the modern polygraph (Marston, 1924). He was also the psychologist who testified in the landmark case of Frye vs. US (1923), which defined the initial standards for the admissibility of expert scientific evidence. Marston (1924) pioneered the first serious jury research, using a simulated jury design which has formed the basis for most

subsequent jury research (Winick, 1961). Several findings remain pertinent today. He suggested that one trained jury member was a more accurate fact finder than an entire jury, irrespective of gender. He indicated that females performed better as jurors than males. The professional training and experience of jurors were associated with better fact finding and written evidence was considered superior to oral evidence. Marston (1924) concluded that the self-confidence of a witness can exert more influence on a jury than the content or persuasiveness of the testimony itself. He recommended that courts incorporate his findings into their practice, however, such recommendations conflicted with procedures which are integral to an adversarial system and consequently many were not implemented (Winick, 1961).

In the period between the1940's and early 1950's, Wigmore became established as the major authority on the admissibility of expert evidence, initiating the use of test data in trials and crystallising the U.S. federal rules of evidence (Bartol & Bartol, 2006). He predicted that the use of scientific test data would develop to become as influential as expert testimony had become to establish insanity, so long as the tests were recognised as valid and feasible by the general scientific community (McCary, 1956).

From the late 1940's onward, it became more commonplace to find psychologists testifying in cases. Frequent areas where they were called upon include the influence of pre-trial publicity on witnesses and juries, the effects of pornography on adolescents, the effects of certain educational approaches on children and the likely influence of advertising on consumers (Loh, 1981). Despite this, psychological testimony was not yet readily accepted and the legal system, on both sides of the Atlantic, still treated it with caution (Bartol & Bartol, 2006).

1.2.4 Rise in the professional standing of psychology

Before the end of the Second World War, criminal law courts began to allow psychologists as expert witnesses on issues of criminal responsibility. This turnaround in the law's standpoint is considered by Loh (1981) to be due to the proliferation of mental health professionals at the time and the development of legal definitions of insanity which were commensurate with psychiatric practices. A landmark ruling which paved the way for psychological testimony on insanity occurred in the case of the People v. Hawthorne (1940). Hawthorne was being tried for the murder of his wife's lover and pleaded not guilty due to insanity. The court initially refused the psychologist to testify as an expert witness, however, on referral to the Supreme Court, this was overruled and the guideline for qualifying as an expert witness was defined as being dependent on the extent of the witness' knowledge rather than on a medical degree. Nevertheless, dissidents continued to contest that diagnosing insanity was the domain of medicine (Bartol & Bartol, 2006). A landmark ruling in 1961 in the case of Jenkins v. US led to direct (albeit conditional) support for psychologists to act as expert witnesses on questions of mental illness and remains the current guiding principle on this issue in the US today (Bartol & Bartol, 2006).

From the 1970's onward, the modern era of forensic psychology began. Research and appearances in court flourished and the first professional forensic psychology body paved the way for the emergence of a subsequent professional cognitive forensic body (Ginliano et al., 1997). The profession of cognitive forensic psychology now practices against a backdrop of the historical interchanges between psychology and the law. As such, the approach taken by forensic and clinical psychologists in court can be seen to be directly influenced by history which has demarcated both the limitations of areas in which psychologists can testify and the nature and quality that the testimony should uphold.

1.2.5 Summary and conclusions

It has been suggested that the lessons previously learned have been forgotten and psychology and the law repeat patterns of 'approach-withdrawal-rebuff' (Loh, 1981). This has hampered the development of interdisciplinary collaboration and if one is to avoid a continual cycle of this friction, then there are certain themes from history that should be remembered today.

These include the way in which psychological evidence is presented in court, first identified by Wigmore (1937) who suggested how group averages should be presented without misleading court decisions. Psychology should not overstate its usefulness, as this has delayed cooperation in the past leading to the law questioning the validity and reliability of psychological evidence (Blau, 1988). History also indicates areas of research which were not fully explored, such as the reaction time experiments of Wertheimer and Klein (1904) and Marston's (1924) investigations into jury decision-making. Finally, the profession needs to delineate the areas in which it can make contributions with clarity, so as to safeguard its future collaboration with the law.

It has taken approximately 100 years to reach a point in the UK where Ian Gibson, the Chairman of the Criminal Cases Review Tribunal has stated that lawyers and judges should not feel capable of dealing with scientific evidence themselves and should call on the assistance of scientists (Dyer, 2005, para. 14). This sets the stage for the current discussions which are taking place between science and the law, which forms the next section of this review.

1.3 Current developments in guidance on standards for admissibility of scientific evidence

In the UK, there is currently no standard protocol for deciding whether scientific evidence is admissible in court (Parliamentary Office of Science and Technology (POST), 2005). During the 2004/05 parliamentary session, the Science and Technology Select Committee (2005) discussed the necessity to develop guidelines to standardise the admissibility of scientific evidence in the UK. This could pre-emptively settle in-court disputes and allow for a common understanding of what constitutes admissible evidence. This has the potential to defuse some of the law's resistance against psychology and oblige the acceptance of psychological evidence with appropriate conditions.

The following section outlines the issues facing the development of these guidelines and considers the likely form such guidelines will take. This includes a more detailed discussion of the US guidelines which may shape the basis of the UK's future protocol (Science and Technology Committee, 2005).

1.3.1 Increasing usage of scientific evidence in court

Scientific evidence is increasingly used in court cases where there is an established scientific practice (for example, breathalysers) and in more innovative areas where the scientific basis has yet to be validated (for example, lie detectors) (POST, 2005). Scientific evidence is almost always used in murder cases and is becoming more regularly used in more frequent crimes such as burglary and car theft (POST, 2005). Of the 140,000 cases dealt with by the Forensic Science Service (FSS) in 2004-05, 2,500 required the use of FSS expert witnesses (website ref cited in POST 2005, ref no. 2 at end). In addition, the Legal Services Commission (LSC), the agency which authorises the use of experts to guarantee their payment, estimated an expenditure of £130 million per year on experts' fees (POST, 2005).

1.3.2 The complexities of controlling the quality of scientific evidence in court

Currently, individual judges must exercise their discretion over whether to accept scientific evidence in court on an ad hoc basis. However, judges have no specific training in the rigorous examination of scientific evidence. As such, there is no consideration of whether a theory is sufficiently robust and evidence-based to warrant admission in court (Science and Technology Committee, 2005). Recently, high profile cases have thrown the shortcomings of existing practice into the spotlight. The wrongful convictions of Sally Clark and Angela Cummings for murdering their babies, exposed the inadequacies, both of Professor Meadows, the expert presenting the erroneous statistical evidence, and of the court in evaluating the admissibility of the evidence. Other cases, such as the exoneration of the "Birmingham six" in 1991, have also received substantial press coverage drawing attention to the failings of our current system in appraising forensic evidence (Science and Technology Committee, 2005).

In developing a quality control for admitting expert scientific evidence, governance needs to be implemented at several levels. Expert witnesses must be vetted for their ability to provide reliable evidence, the courts, including judges and lawyers, need to be trained to make informed decisions about the validity of expert evidence and there needs to be a protocol for presenting expert evidence to the court, and particularly to jurors, in an intelligible and impartial fashion. Each of these levels present their individual problems and are discussed in turn.

The court is responsible for establishing the competency of expert witnesses (Science and Technology Committee, 2005). The Council for the Registration of Forensic Practitioners (CRFP) was founded in 1999 to provide a means of regulating this competency. Forensic practitioners are required to meet standards to ensure that their practice is up to date and competent. To join the register, applicants must

provide details of their qualifications and experience, references from colleagues and users of their services and declarations about their past and future conduct (Science and Technology Committee, 2005). This provides a means of supporting the credibility of expert witnesses.

However, there are several problems with making registration mandatory. It has been suggested that registration is too easy and does not rigorously evaluate whether the practice of its registrants is evidence-based. It is suggested that a continual process of peer review and evaluation is needed (Keogh, 2004, para. 135). Registration cannot be regarded as indisputable verification of competency (Sqibb-Williams, 2005, para.135). As such, the CRFP must itself be subjected to independent auditing of its registration process (Science and Technology Committee, 2005). In addition, emerging specialisms would be disadvantaged as assessors in their discipline may not yet exist. This could restrict the courts in being able to call upon the latest developments in scientific knowledge, a prerequisite for making all potentially applicable information available in fair trials (POST, 2005). As a result, the government response to the Science and Technology Committee stated that CRFP registration should not be made mandatory (Science and Technology Committee, 2005a).

The CPS does not provide any compulsory training for lawyers in the understanding and presentation of forensic evidence (Science and Technology Committee, 2005). Although lawyers are obliged to undergo 12 hours continuing professional development each year, there are no existing provisions necessitating training on expert evidence. Whilst specific guidance has been developed for the presentation of DNA evidence in court, further more generic guidance on presentation of scientific evidence is not specified. It falls to the Bar Council to make such provisions mandatory. The Science and Technology Committee (2005) recommended a

consultation on the training of specialist judges and barristers to increase understanding of specific areas of forensic evidence. This could provide an additional safeguard against the impartial presentation of scientific evidence and another gateway for assessing the admissibility of forensic evidence.

The adversarial system requires that an expert witness be called by either the prosecution or the defence (Science and Technology Committee, 2005). As such, the witness is effectively providing evidence to strengthen the case of one side over the other. Though the expert is obliged to remain impartial, a survey commissioned by an expert witness training company Bond Solon (2002) indicated that it is commonly accepted that lawyers do not encourage their witnesses to be truly independent. Additionally, it has been suggested that the defence will 'shop around' for an expert who will provide the most favourable evidence, though the Criminal Justice Act (2003), section 35, makes provisions which are supposed to avoid this.

A possible method of overcoming this problem is the use of a single joint expert witness who represents both parties. However, the European Convention of Human Rights entitles either party to draw on their own expert evidence on points of contention (Science and Technology Committee, 2005). Another way to address this difficulty is to prescribe methods of presenting scientific evidence so that it is not open to interpretation. Such methods have been introduced in the presentation of DNA evidence to protect against misleading presentation of probability and statistics (Science and Technology Committee, 2005). It seems possible to generalise this to all statistical presentations, thereby eliminating the element of persuasion from the presentation of scientific evidence that has marred its impartiality (Science and Technology Committee, 2005).

In summary, measures to control the quality of forensic evidence must ensure that

expert witnesses are competent and use scientifically valid techniques of an agreed standard. The basis for their interpretation in court must also be validated before being presented as evidence. Implementation of the recommendations of the Science and Technology Select Committee (2005) hinge on establishing the Forensic Science Advisory Council (FSAC). This council is to be formed by a consultation group comprising representatives from the Home Office, the Association of Chief Police Officers (ACPO) and the Association of Police Authorities (APA). To satisfy the code of conduct for an advisory committee, the FSAC must include lay members and experts among its members (Science and Technology Committee, 2006). It will be interesting to see who the ACPO and APA in conjunction with the HO recommend for the FSAC, as there already appears to be an overrepresentation of police interest in its formation. This is potentially concerning given prior accusations that the police cherry pick witnesses who provide a more convincing case, a practice that has been suggested is unlikely to lead to a fair trial (Science and Technology Committee, 2005).

1.3.3 Developing a gate-keeping protocol for the admissibility of scientific evidence

The first task of the FSAC is to develop a 'gate-keeping' protocol for the validation of scientific evidence (Science and Technology Committee, 2005a). The council will face the not insignificant task of drawing together suggestions from disparate services with vested interests: Legal Services Commission, Association of Chief Police Officers, Association of Police Authorities, Home Office, Criminal Justice System, Scientific Review Committee, Criminal Cases Review Committee, Council for Registration of Forensic Practitioners, Forensic Science Service. These competing pressures must be handled by the FSAC with the obligatory impartiality that is the code of conduct for government councils (Science and Technology Committee, 2006).

The standards that will comprise the gate-keeping test will be of particular interest to forensic psychology practitioners. These standards will ultimately decide which psychological tests will meet the requirements for admission in court, and what inferences can be made from the test results. It has been suggested that the future UK protocol will build on the existing US practice, which is based on the *Frye* and *Daubert* principles (Science and Technology Committee, 2005). The *Frye* test dates back to 1923 and dictates that evidence from novel scientific techniques can only be accepted if the technique falls into an established scientific field and is generally accepted within that community (Science and Technology Committee, 2005). The *Daubert* test superseded this and is today considered a more rigorous set of standards. As it is likely that this test will form the basis of a UK procedure, it is considered in more detail.

Since 1993, all federal courts and state jurisdictions in the US follow the Federal Rule of Evidence 702 informed by the *Daubert* case. Rule 702 offers 4 criteria by which scientific validity may be assessed. It falls to the trial judge to consider whether the evidence meets these criteria (Vallabhajosula & van Gorp, 2001):

i) The technique at issue can or has been tested.

ii) The technique has been subjected to the scrutiny of the scientific community.

iii) Whether there are standards controlling the technique's operation, and the known or potential rate of error for such technique.

iv) Whether the technique generally has been accepted in the particular scientific field.

1.3.4 Lessons learnt from the US *Daubert* guidelines

In this way, the courts are invited to evaluate the reliability and validity of the instruments used to collect data and also of the inferential methods used by clinicians to generate their opinions, including diagnoses (Fiedler, 1997). For example, one

such case which applied the Daubert guidelines rejected the admissibility of the Minnesota Multiphasic Personality Inventory - 2 (Butcher, Graham, Dahlstrom, Tellegen & Kaemmer, 1989) in relation to sex offender profiles. The heterogeneity of sex offenders rendered the instruments' classifications inadmissible based on unreliable methodology and unacceptable scientific validity (Rogers, Salekin & Sewell, 1999).

Two further rulings by the Supreme Court in 1997 and 1999, broadened the Daubert ruling. The first ensured the trial courts adopted a 'gatekeeping' role to exclude unreliable evidence and the second widened this gatekeeping function to permit expert witness evidence that was not based on science. Together these rulings emphasised that it is the responsibility of the trial judge to evaluate scientific validity and how the evidence can be applied to the facts in issue (Vallabhajosula & van Gorp, 2001). This may prove problematic if the UK were to follow the same procedure, as has already been discussed, the scientific knowledge of UK judges may not be adequate to make such decisions.

When implementing these guidelines, the US courts have had some difficulty in analysing the scientific validity of evidence. For example, in Chapple v. Gangar (1994), the court accepted psychometric test results for normal scores, but rejected scores significantly below the norm because the conclusion drawn that these scores related to permanent brain damage in children was not considered scientifically valid. The full methodology of the experts must be taken into account, including the conclusions drawn from the data. The *Daubert* guidelines state that the entire reasoning process must be valid (Daubert v. Merrell Dow Pharmaceuticals, 1993). The courts' difficulty in assessing scientific validity is hardly surprising. This area is fraught with complexities which vary according to the vast array of scientific evidence that can potentially be presented to a court. The following section will consider how

scientific validity is assessed in one specialised area, that of the assessment of cognitive malingering. It shows that even among specialist experts in a narrowed field, there is still controversy over assessing validity.

1.4 Current practice of presenting psychometrics to detect cognitive malingering in criminal law courts

1.4.1 Commonly used neuropsychological tests to detect malingering

Currently the most common tests of cognitive malingering which are presented to law courts in the US are cited as being the Rey Fifteen Item Test (FIT) (Rey, 1964), the Test of Memory Malingering (TOMM) (Rees, Tombaugh, Gansler & Moczynski, 1998) and the Validity Indicator Profile (VIP) (Frederick & Crosby, 2000). These tests were suggested by Vallabhajosula and van Gorp (2001) to be the most commonly used in their paper which evaluated tests that more likely meet the Daubert criteria. They also emerge as the most popular psychometric tests of malingering with US forensic neuropsychologists from a search of the LEXIS "Combined Federal and State Case Law - US" database, in which 15 out of 18 published tests which refer to cognitive malingering used them (Mossman, 2003). The author of this review replicated the above search on the UK LEXIS database, using the same search criteria as Mossman (p.231, 2003), however there were no references to the use of specific tests of cognitive malingering in UK courts.

Possible reasons for this may include reluctance on the part of UK practitioners to cite formal measures of malingering because of fears that categorising a malingerer may necessitate a subsequent charge of perjury. Also, formal criteria for assessing suitability of these measures do not yet exist, meaning that their use has not developed with the same impetus as in the US. In addition, the UK database is not as

thorough at listing references to particular scientific measures as the US version and so searches do not throw up as many examples. For whatever reason, it seems plain that there is less to instruct UK neuropsychologists on which tests to select when preparing court reports regarding symptom validity.

From a research perspective, a meta-analysis conducted by Vickery et al. (2001) suggested that the most effective neuropsychological tests of inadequate effort are the Digit Memory Test (DMT), Portland Digit Recognition Test (PDRT), 15-Item Test, 21-Item Test, and the Dot Counting (Lezak, 1983). In a book reviewing the detection of response bias in forensic neuropsychology, Hom and Denny (2002) selected papers which discussed the Victoria Symptom Validity Test (Slick, Hopp, Strauss & Thompson, 1997), the Word Memory Test (Green, Allen & Astner, 1996), the Recognition Memory Test (Warrington, 1984), the Minnesota Multiphasic Personality Inventory - 2 (Butcher, Graham, Dahlstrom, Tellegen & Kaemmer, 1989), Weschler Adult Intelligence Scale 3rd Edition (Weschler, 1997), Weschler Memory Scale - Revised (Weschler, 1987), the Halstead-Reitan Battery (Reitan & Wolfson, 1985), and the Category Test (Bolter, 1985), in addition to the tests mentioned above.

There is no shortage of neuropsychological measures of dissimulation and, indeed, exploratory research is still flourishing in this area. This may be due to a combination of factors. There are different categories of measure which may be more appropriate for different populations, there is no one measure which is absolutely categorical and indeed many researchers recommend the use of a battery of tests to corroborate findings (e.g. Rogers, 1997). In addition, each individual measure has questionable accuracy.

1.4.2 Within-test strategies for detecting malingering

The array of cognitive measures of malingering have been categorised into six

groupings according to the detection strategy used and have been evaluated for their potential validity (Rogers, Harrell & Liff, 1993). In brief, the groupings are as follows:

Floor Effect

Failure on simple tasks that even severely impaired persons would pass on defines the floor effect strategy. It is hypothesised that failure on simple informational questions (for example, "Which is bigger, a horse or a dog?") effectively differentiates malingerers. Examples of such tests include the Rey 15 Item Memory Test (Lezak, 1983) and The Wiggins and Brandt Personal History Interview (Wiggins & Brandt, 1988). The latter asks questions that even people with severe memory loss are able to answer correctly (e.g. "What is your name?"). The use of such techniques is ethically questionable when the difficulty of the task is stressed so as to tempt malingerers into feigning (e.g. Drob & Berger, 1987). It is also considered that the transparency of such a technique will not identify more sophisticated malingerers.

Performance Curve

This compares the number of easy items failed to the number of difficult items passed. The hypothesis is that malingerers will not alter their response patterns to account for differences in item difficulty. Comparing their performance curve to 'true' patient groups can potentially differentiate malingerers. Examples of such tests include use of the Rey Dot Counting and Word Recognition Test, recommended by Lezak (1983), who suggested that truthful respondents would be likely answer easier recognition tests correctly and harder recall questions with more mistakes. She hypothesised that malingerers would not have the same pattern of responses in relation to the increasing difficulty of the questions. This strategy has been utilised with success on memory tasks (Graf, Squire & Mandler, 1984) and intelligence tests (Gudjonsson & Shackleton, 1986).

Magnitude of Error

Differences in response error are suggested to distinguish malingerers from valid respondents. An example of this, which is possibly where the strategy emerged, is the Ganser Syndrome (Rogers, Harrell & Liff, 1993). This proposes that proximate answers are more likely of malingerers (i.e. near misses). The initial underlying theory supposes that malingerers who know the correct answer and suppress a truthful response, reveal their deception by answering close to the correct answer. There is still much research to be done on this area to determine whether there exist identifiable patterns of malingered responses (Rogers, Harrell & Liff, 1993). Such research includes investigations into both 'near misses' and 'gross errors' (e.g. Bruhn & Reed, 1975; Rogers, Harrell & Liff, 1993). The latter incorporates an extension to this theory that some malingered responses will be excessively distant from the correct answer. Powell (1991) has shown promising results in research measuring simulating malingerers of schizophrenia on the Mini-Mental State. Malingerers had a greater proportion of both proximal and distal answers on cognitive items.

Symptom Validity Testing

Symptom validity tests (SVT's) attempt to detect malingerers by identifying a response style that is significantly worse than would be achieved by chance. If the probability of wrong responses is significantly worse than an impaired person responding randomly then such test results indicate malingering. Pankrantz (1983) explored this strategy on a test which measured feigned deafness. Respondents were required to choose between the presence or absence of sound. During the test, sounds were presented half the time. Malingerers were found to answer more questions wrongly than would be expected by chance responding. Variations on this strategy increase the sophistication of the method, for example, by increasing the number of possible responses or examining performance across time (e.g. Hiscock and Hiscock, 1989). A strength of this approach is the lack of alternative explanations

for significantly below-average responding. A relative weakness is that in simulation designs, few simulating malingerers score significantly below average (Rogers, Harrell & Liff, 1993).

Atypical Presentation

Varying or atypical performances are thought to distinguish malingerers. These include significant differences when re-administering the same measure, or significant differences on parallel tasks which test the same ability (e.g. Wasyliw & Cavanaugh, 1988). However, both functional disorders and brain-injured patients may present similarly inconsistent response styles (Pankratz, 1988). Atypical presentations have also not been adequately tested via simulation or clinical sample designs (Rogers, Harrell & Liff, 1993) and so further support is needed before their recommended use.

Psychological Sequelae

In addition to feigned cognitive deficits, it is thought that malingerers also reported elevated levels of psychological symptoms, such as emotional disorders or personality abnormalities (e.g. Heaton et al., 1978). Schacter (1986) researched amnesiacs attitudes towards their memory loss. Simulating malingerers were found to overstate their pessimism about recovery. The main problems with this type of approach is that the psychological sequelae of genuine patient groups is largely unknown (Rogers, Harrell & Liff, 1993). It is also contended that malingerers may have insight into the likely sequelae of certain conditions, for example 'postconcussion syndrome' (Mittenberg et al. 1990). Further study is needed to establish the utility of this approach.

1.4.3 A closer examination of the *Daubert* guidelines

Once the approach has been selected, the procedures for evaluating which test to

choose when considering admissibility in court would, at the moment, most sensibly follow the Daubert guidelines. However, these are only guidelines and are not strict criteria which can be unconditionally fulfilled. The following section outlines considerations which may be taken into account under each Daubert criterion.

Daubert criterion (i): Has the procedure been tested?

In considering whether a measure has been adequately tested, one should consider firstly whether it has been evaluated by researchers. This could include whether the investigators were independent and not allied to the tool in question (Millis, 2002). The methodology used to evaluate the measure should also be scrutinised. In the case of a measure of cognitive symptom validity, one would need to weigh up whether the research includes both simulation design and clinical samples. The difficulties with validating measures using clinical samples of known malingerers are well documented (e.g. Rogers, 1997) due to problems identifying malingerers who are by definition elusive. Simulation designs may be the standard research method, but validation using this technique alone may be questioned when relating the findings to suspected 'true' malingerers in court. Simulating malingerers are not adequately representative of the response style of true malingerers and as such, cannot be used alone to influence court decisions.

The test may also be evaluated in terms of the clinical disorders that it has been tested on, e.g. both psychiatric and neurological disorders. This allows the forensic practitioner to establish an 'empirical floor' or base rate of norm-referenced diagnostically specific samples which enables cut-offs to be decided upon (Bianchini, Mathias & Greve, 2001). Hence, more accurate estimations can be made about how someone with a genuine disorder would perform on the test and above-chance estimates of feigning can be made with more confidence.

Daubert criterion (ii): Has the test been subjected to the scrutiny of the scientific community?

Scrutiny in the scientific community necessitates peer review and publication. The extent to which this has happened, denoted both by number of publications and credibility of the peer review journals will determine whether a test meets this criterion. New measures for detecting symptom validity would be of distinct disadvantage here, as has been identified by the Science and Technology Committee (2005). Novel methods would be excluded on the basis that they have not had adequate time to receive peer review. This is contrary to the law's ethos of making all evidence available to the court. Leeway should be given when debating this criterion in court to allow for innovative methods to emerge and not to stifle burgeoning areas of research. This criterion should perhaps not be used as a deciding factor for admitting a measure.

Daubert criterion (iii): What are the standards controlling the technique's operation, and what is the known or potential rate of error for such a technique?

A tests accuracy is determined by its error rate. Without an acceptable error rate, the test will be deemed unreliable (criterion 1) and will not stand up to the scrutiny of peer review or be accepted by the scientific community (criteria 2 and 4). As such, the potential of a test to reach the standards of admissibility following the Daubert guidelines can be understood to pivot on its error rate. Despite this, none of the cases which have applied the Daubert guidelines to tests of cognitive malingering have discussed the error rate in court (Mossman, 2003). This has led to calls to establish a set error rate as a standard for the scientific community, which can also facilitate determining admissibility in court (Vallabhajosula & van Gorp, 2001).

There are two main parts to this criterion when considering admissibility. Firstly, whether the measure reaches this statistical standard and secondly, how conclusions drawn from the test are presented to the court and related to the case in question. These points will be discussed in turn.

1.4.4 Statistical standards for malingering tests

The scientific validity of a test of cognitive malingering is determined by its sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Sensitivity is the probability of a positive test (i.e. a score above the cut-off on a test designed to detect malingering) among individuals who are malingering. Specificity is the probability of a negative test (i.e. a score below the cut-off) among individuals who are not malingering. The PPV is the proportion of clients who are identified by a test as having malingered who actually are malingering. The NPV is the proportion of clients who are identified by a test as not malingering who are in fact not malingering. It provides a probability that a client, who scores below a cut-off, is actually not malingering. Sensitivity and specificity are independent of base rates and directly dependent on the cut-off employed. They provide a categorical classification of malingering. PPV and NPV represent the predictive power of the test, the confidence that one can have that the classification is correct (Greve & Bianchini, 2004). Predictive power is considered to be the most clinically relevant function to assess efficacy of prediction with regard to malingering (Vallabhajosula & van Gorp, 2001; Rosenfeld, Sands & van Gorp, 1997; Rogers, 1997).

By altering the cut-off value, one can increase the sensitivity of a test. However, cut-offs designed to capture ambitiously high numbers of malingerers will likely result in higher numbers of false positives (i.e. identifying honest responders as malingerers). Similarly, altering the cut-off to maximise specificity will compromise sensitivity and

result in higher numbers of false negatives (i.e. identifying malingerers as honest responders). Given the pejorative connotations of applying the label of 'malingerer', many consider it more acceptable to allow false negative errors in favour of false positive errors (Wasyliw & Golden, 1985). Greve & Bianchini (2004) describe the 'burden of error' as being carried by the individual patient in cases of false positives and the consequences for them personally, financially and occupationally. Whereas in cases of false negatives, the burden is carried by society, for example, missing malingerers reduces the pool of resources for legitimate claimants (Franzen, Iverson & McCracken, 1990). When setting cut-off rates for use in court, one must be mindful of whether the emphasis in an individual case is on specificity or sensitivity. For example, more emphasis may be given to society in the case of a suspected malingerer in a forensic setting who is claiming diminished responsibility for a serious crime. In contrast, more emphasis may be given to the individual in the case of a claimant who has been injured in an industrial accident and is being assessed for the veracity of his reporting of posttraumatic stress symptoms. Fixed levels of sensitivity and specificity are therefore not attractive and individual cases should consider suitable cut-offs on merit.

Given the importance of specificity over sensitivity, making decisions about cut-offs should be based primarily on specificity. To minimise vulnerability to Daubert challenges the error rate should be reported transparently. Where only a single cut-off is set, it should be done to give perfect specificity. More commonly, a range of cut-offs are reported according to pre-determined specificity levels (e.g. .80, .85, .90., .95 and 1.00), and then providing the corresponding sensitivity levels (Greve & Bianchini, 2004). This allows the development of a continuum of predictive power below 100%. The predictive power associated with these cut-offs can then be related to a range of likely base rates of malingering for a given population.

It is well reported that predictive power varies with the base rate (e.g. Mossman & Hart, 1996; Heaton, Smith, Lehman & Vogt, 1978). Base rates of malingering in litigating populations has been reported to vary between 7.5 to 33 percent, whilst in forensic populations it is thought to be even higher (Vallabhajosula & van Gorp, 2001). However, it is considered unlikely that base rates would exceed .30 in most settings (Sweet, 1999). When there is no indication of base rates for a given population and only one is to be reported, it is recommended that predictive power is calculated using a base rate of .30 (e.g. Vallabhajosula & van Gorp, 2001; Mittenberg, Patton, Cannock & Condit, 2002).

Setting the cut-off is further dependent on 3 factors, the distribution of scores, the sample size and the sample composition. If the distribution of scores is skewed, it may not be possible to establish cut-offs that result in recommended specificity rates. Scores that are more continuous and less heavily skewed will be easier to allocate cut-off values. However, on many symptom-validity tests, for example the TOMM (Tombaugh, 1996), few respondents make errors which heavily skews the distribution of scores and makes it harder to assign cut-offs (Greve & Bianchini, 2004). The size of the nonmalingering control group is also crucial to determining cut-off values. Smaller nonmalingering sample sizes means that the cut-off value must also be smaller so that each individual does not represent too significant a proportion of the sample. For example, with a control group of 20 nonmalingerers, a cut-off of .99 would mean that each participant accounts for an unacceptable 5% of the sample (Greve & Bianchini, 2004). Finally, there must be at least 2 groups in a simulating malingering design as sensitivity and specificity are determined by a comparison of the performance of individuals in the groups. In addition, to determine sensitivity, one must be very confident that individuals in the malingering group are indeed malingerers. This is a common failing of designs which use 'genuine' suspected malingerers from clinical samples (Bianchini, Mathias, Greve, Houston & Crouch,

2001).

1.4.5 Presenting test results in court

Once it has been established that the data reaches an acceptable statistical and scientific standard, it falls to the expert witness to present the findings in such a way that they do not misrepresent the generalisability of the findings nor overstate the conclusions that can be drawn from them. There is an essential difference between law and psychology which leaves them fundamentally at odds. This difference can be understood from philosophical mathematics.

The detection of malingering follows the Bayesian method for determining predictive accuracy (Mossman, 2003). Expert witnesses are limited in doing no more than to provide information which contributes to the formation of opinion relevant to the matter at hand (Wagenaar, 1988). They should not provide final opinions themselves. Law, however, seeks to align disparate viewpoints and find a categorical stance (Vallabhajosula & van Gorp, 2001).

Probability can be thought of a subjective degrees of belief. An opinion is related to the probability that a hypothesis is true. Thus, opinion delivered in court can be evaluated for its relevance in terms of the probability that it is true (i.e. its error rate). Conversely, information refers to the probability of finding evidence only when the hypothesis is true. Thus if the hypothesis is equivocal, the information or evidence gleaned from it is even more shaky. Furthermore, opinion and information can be confounded and this becomes evident when judges and judiciary ask for expert opinion that transgresses their expertise. Mistakes occur when their statements are substituted as a court's opinion (Wagenaar, 1988). There is a need to make clear what conclusions can be drawn from the opinion and the limited influence that these conclusions should have on the matter at hand.

Attempts to control for this have been made in the case of the presentation of DNA evidence. A standard wording for presenting this probabilistic scientific evidence is intended to become mandatory and is followed by an explanation of how the evidence should be interpreted in relation to the case. The new recommendation for wording the presentation of DNA evidence is as follows (Science and Technology Committee, 2005):

> "The probability that an unknown person, unrelated to the defendant, would have the same profile as the crime sample is 1 in X [the relevant figure]."

This aims to standardise the presentation of scientific evidence so as to reduce the chances of juries and judiciary misinterpreting the findings. When presenting psychometric evidence to courts, the process should be equally systematic. A standard wording for presenting psychometric results could also be made mandatory.

In summary, the initial stage of selecting a type or types of test is followed by the selection of specific tests of malingering, which ideally should incorporate a variety of techniques to improve discriminatory power. Standards for selecting the tests are explained and the method for presenting conclusions is elucidated. Following such a procedure heightens the chance of a battery of tests being admitted in court and subsequently standing up to challenge.

1.5 Future of admissible psychological evidence of cognitive malingering

Future research which seeks to develop neuropsychological tests of malingering should consider past investigations which have indicated areas of useful inquiry. These can be gleaned from reviewing literature which reports salient historical experiments and research. This paper identifies certain techniques pertinent to

assessing cognitive malingering, particularly, the use of forced-choice self-report symptom inventories and the reaction time experiments of Wertheimer and Klein (1904).

In addition, when seeking to develop a battery of cognitive tests of malingering for use in court, one should consider the political context in which the battery of tests will be evaluated. This review discusses changing policy in this area which has a direct impact on forensic neuropsychologists. In the near future, it is likely that tests will be considered with far more scrutiny than has been the case thus far. When developing a new battery of tests, researchers should aim for a minimum standard denoted by the Daubert guidelines. Further, the way in which results of such tests are presented in court should be standardised, to minimise misinterpretation. The paper recommends good practice for presenting findings using a range of specificity and sensitivity levels.

As protocols controlling the admissibility of scientific evidence tighten, research in this field must respond to ensure the continued usefulness of psychological contributions to court processes. Given the capricious nature of the relationship between psychology and the law, it should be expected that forensic psychologists will be strongly challenged in court. Therefore, it is important that psychological research compensates by developing tools which will stand up to more vigorous challenge. Future research should continue to explore all the avenues that the history of empirical and experimental psychology has made available and develop tests that outperform its scientific counterparts, whilst presenting itself with an understated restraint that will not provide fodder for psychology's many presumptuous doubters.

1.6 References

Barland, G. (1988). The Polygraph test in the USA and elsewhere. In A. Gale (Ed) *The Polygraph Test: Lies, Truth and Science.* London: Sage.

Bartol, C. R. & Bartol, A. M. (2006). The history of forensic psychology. In Weiner & Hess (Eds.) *The Handbook of Forensic Psychology.* Hoboken, NJ: John Wiley & Sons, Inc.

Ben-Porath, Y., Graham, J., Hall, G., Hirschman, R. & Zaragoza, M. (Eds.) (1995). *Forensic Applications of the MMPI-2.* London: Sage.

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

Bianchini, K. J., Mathias, C. W., Greve, K. W., Houston, R. J. & Crouch, J. A. (2001). Classification accuracy of the Portland Digit Recognition Test in traumatic brain injury. *The Clinical Neuropsychologist,* 15(4), pp 461-470.

Binet, A. (1900). *La Suggestibilite.* Paris: Schleicher. As cited in Bartol, C. R. & Bartol, A. M. (2006). The history of forensic psychology. In Weiner & Hess (Eds.) *The Handbook of Forensic Psychology.* Hoboken, NJ: John Wiley & Sons, Inc.

Binet, A. (1905). The science of testimony. *L'Annee Psychologique,* 11, pp 128-137.

Blau, T. H. (1988). *The Psychologist as Expert Witness (2nd Edition).* New York: John Wiley and Sons Inc.

Bolter, J. F. (1985). Item error frequencies for the Halstead Category Test: A cross validation sturdy on a performance validity check. (unpublished manuscript).

Bond Solon (2002). *Results of an anonymous survey of 133 expert witnesses conducted in November 2002*. Retrieved on January 11, 2007, from http://www.bondsolon.com/survey.asp?survey=180

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. & Kramer, B. (1989). *Minnesota Multiphasic Personality Inventory 2: Manual for Administration and Scoring.* Mineapolis, MN: University of Minnesota Press.

Bruhn, A. R. & Reed, M. R. (1975). Simulation of bran damage on the Bender-Gestalt test by college subjects. *Journal of Personality Assessment,* 39, pp 244-255.

Cattell, J. M. (1895). Measurements of the accuracy of recollection,*Science*, 2, pp 761-766.

Chapple v. Gangar, (1924). 851 F. supp 1481 (E.D. Washington1994)

Criminal Justice Act (2003). Retrieved January 11 2007, from http://www.opsi.gov.uk/acts/acts2003/20030044.htm

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). 509 U.S. 579 (1993).

Drob, S. L. & Berger, R. H. (1987). The determination of malingering: A comprehensive clinical-forensic approach. *Journal of Psychiatry and Law*, 15, pp 519-538.

Dyer, C. (2005). *Science Baffles Juries say MP's.* Retrieved January 11 2007 from http://www.guardian.co.uk/uk_news/story/0,3604,1447122,00.html#article_continue

Fiedler, J. F. (1997). *The National Law Journal.* New York: The New York Publishing Company.

Franzen, M. D., Iverson, G. L. & McCracken, L. M. (1990). The detection of malingering in neuropsychological assessment. *Neuropsychological Review*, 3, pp 247-278.

Frederick, R. I. & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior,* 24, pp 59-82.

Frye v. United States (1923). 293 F. 1013-14 (D.C. cir. 1923).

Giuliano, A. J., Barth, J. T., Hawk, G. L. & Ryan, T. V. (1997). The forensic neuropsychologist: precedents, roles and problems. In R. McCaffrey, A. Williams, J. Fisher & L. Laing (Eds.) *The Practices of Forensic Neuropsychology.* New York: Plenum Press.

Graf, P., Squire, L. R. & Mandler, G. (1984). The information that amnesiacs do not forget. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 10, 164-178.

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, pp 533-541.

Gudjonsson, G. H. & Shackleton, H. (1986). The pattern of scores on Raven's Matrices during "faking bad" and "non-faking bad" performance. *British Journal of Clinical Psychology*, 25, pp 35-41.

Green, P. Allen, L. M. & Astner, K. (1996). The Word Memory Test: A User's Guide to the Oral and Computer-Administered Forms, US Version 1.1. Durham, NC: CogniSyst.

Hale, M. (1980). *Human Science and Social Order.* Philadelphia: Temple University Press.

Healy, W (1924). *Pictorial Completion Test II.* Chicago, Ill: C.H. Stoelting Company.

Heaton, R. K., Smith, H. H., Lehman, R. A. W. & Vogt, A. J. (1978). Prospoects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology,* 46, pp 892-900.

Hiscock, M. & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11, pp 967-974.

Hom, J. & Denny, R. L. (Eds.) (2002). *Detection of Response Bias in Forensic Neuropsychology.* London: Haworth Press.

Kargon, R. (1986). Expert testimony in historical perspective. *Law and Human Behavior,* 10(1-2), pp 15-27.

Keogh, A. (2004). Experts in the dock. *New Law Journal*, 154.

Lezak, M. (1983). *Neuropsychological Assessment.* New York: Oxford Universoty Press.

Loh, W. D. (1981). Perspectives on psychology and law. *Journal of Applied Social Psychology,* 11(4), pp 314-355.

Jenkins v. United States (1961). 307 F.2d 637 (D.C. Cir. 1962) *en banc.*

Marston, W. M. (1924). Studies in testimony. *Journal of Criminal Law and Criminology*, 15, pp 5-32.

McCary, J. L. (1956). The psychologist as an expert witness in court. *American Psychologist,* 11, pp 8-13.

Millis, S. R. (2002). Warrington's Recognition Memory Test in the detection of response bias. In J. Hom & R. L. Denny (Eds.) *Detection of Response Bias in Forensic Neuropsychology.* London: Haworth Press.

Mittenberg, W., DiGiulio, D. V., Perrin, S., & Bass, A. E. (1990). Symptoms following mild head injury: Expectations as etiology. *The Clinical Neuropsychologist,* 3, pp 297.

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

Monahan, J. & Loftus, E. F. (1982). The psychology of law. *Annual Review of Psychology*, 33, pp 441-475.

Mossman, D. (2003). Daubert, cognitive malingering and test accuracy. *Law and Human Behavior,* 27(3), pp 229-249.

Mossman, D. & Hart, K. J. (1996). Presenting evidence of malingering to courts: insights from decision theory. *Behavior, Science and Law*, 14, pp 271-291.

Musterberg, H. (1899; 1994). Psychology and history. *Psychology Review,* 101, pp 203-236.

Munsterberg, H. (1908). *On the witness stand: Essays on psychology and crime.* New York: McClure. As cited in Bartol, C. R. & Bartol, A. M. (2006). The history of forensic psychology. In Weiner & Hess (Eds.) *The Handbook of Forensic Psychology.* Hoboken, NJ: John Wiley & Sons, Inc.

Pankrantz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills,* 57, pp 367-372.

Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception.* New York: Guildford Press.

People v. Hawthorne (1940). 293 Mich. 15, 291, M.W. 205.

Parliamentary Office of Science and Technology (POST) (2005). *Science in Court.* Retrieved on January 11, 2007 from www.parliament.uk/parliamentary_office/post/pubs2005.cfm

Powell, K. H. (1991). *The Malingering of Schizophrenia.* Unpublished doctoral dissertation, University of South Carolina, Columbia.

Rees, L. M., Tombaugh, T. N., Gansler, D. A. & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, 10, pp 10-20.

Reitan, R. M. & Wolfson, D. (1985). *Halstead-Reitan Neurological Test Battery.* Tuscon, AZ: Neuropsychology Press.

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

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

Rogers, R., Harrell, E. H. & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review,* 13, pp 255-274.

Rogers, R., Salekin, R. T., & Sewell, K. W. (1999). Validation of the Millon Clinical Multiaxial Inventory for axis II disorders: Does it meet the Daubert standard? *Law and Human Behavior,* 23, pp 425-464.

Rosenfeld, B., Sands, S. & van Gorp, W. G. (1997). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, 15, pp 349-359.

Schacter, D. L. (1986). On the relation between genuine and simulated amnesia. *Behavioral Sciences and the Law,* 4, 47-64.

Science and Technology Committee (2005). *Select Committee on Science and Technology Seventh Report.* Retrieved on August 8 2006, from, http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/96/9610.htm

Science and Technology Committee (2005a). *Forensic science on trial: Government's response to the committee's seventh report of session 2004-05.* Retrieved on January 11 2007, from, http://www.publications.parliament.uk/pa/cm200506/cmselect/cmsctech/427/427.pdf

Science and Technology Committee (2006). *Science, advice, risk and evidence: How the government handles them.* Retrieved on January 11, 2007 from, http://www.publications.parliament.uk/pa/cm200506/cmselect/cmsctech/uc900-xii/uc90002.htm

Slick, D. J., Hopp, G., Strauss, E. & Thompson, G. B. (1997). *Victoria Symptom Validity Test: Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Sqibb-Williams, K. (2005). Cited in Science and Technology Committee (2005), para 151. *Select Committee on Science and Technology Seventh Report.* Retrieved on August 8 2006, from, http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/96/9610.htm

Stern, L. W. (1903). The psychology of testimony. *Journal of Abnormal and Social Psychology,* 40, pp 3-20.

Sweet, J. J. (1999). *Forensic Neuropsychology: Fundamentals and Practice.* New York: Swets & Zeitlinger Publishers.

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

Vallabhajosula, B. & van Gorp, W. G. (2001). Post-Daubert admissibility of scientific evidence on malingering of cognitive deficits. *The Journal of the American Academy of Psychiatry and the Law,* 29(2), pp 207-216.

Vendemia, J. M. C., Buzan, R. F. & Simon-Dack, S. L. (2005). Reaction time of motor responses in two-stimulus paradigms involving deception and congruity with varying levels of difficulty. *Behavioural Neurology*, 16(1), pp 25-36.

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, pp 45-73.

Wagenaar, W. A. (1988). The proper seat: A Bayesian discussion of the position of expert witnesses. *Law and Human Behavior*, 12(4), pp 499-510.

Warrington, E. K.(1984). *Recognition Memory Test Manual.* Windsor: NFER-Nelson Publishing Company.

Wasyliw, O. E. & Cavanaugh, J. L. (1988). Simulation and brain damage: Assessment and decision rules. *Bulletin of the American Academy of Psychiatry and Law*, 17(4), pp 373-386.

Wasyliw, O. E. & Golden, C. J. (1985). Neuropsychological evaluation in the assessment of personal injury. *Behavioral Sciences and The Law*, 3, pp 149-164.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale - Third Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1987). Wechsler Memory Scale - Revised. San Antonio, TX: The Psychological Corporation.

Wertheimer, M. & Klein, J. (1904). Psychologische Tatbestandsdiagnostik. *Archiv fur Kriminal-Anthropologie und Kriminalistik,* 15, pp72-113.Cited in E. B. Newman (1944) Max Wetheimer: 1880-1943, *The American Journal of Psychology*, 57(3), pp. 428-435.

Wiggins, E. C. & Brandt, J. (1988). The detection of simulated amnesia, *Law and Human Behavior,* 12(1), pp 57-78.

Wigmore, J. H. (1909). Professor Munsterberg and the psychology of testimony. Being a report in the case of Cokestone v. Munsterberg, *Illinois Law Review*, 3, 399-445.

Wigmore (1937). *The Science of Judicial Proof as Given by Logic, Psychology and General Experience and Illustrated in Judicial Trials (3rd Edition).* Boston: Little Farm.

Winick, C. (1961). "The Psychology of Juries". In H. Toch (Ed.) *Legal and Criminal Psychology*, pp 96-120. New York: Holt, Rinehart & Winston.

Part 2: Empirical Paper: Malingering of Cognitive Symptoms

2.0 Abstract

Malingering has significant consequences on social and economic levels. Studies on the detection of dissimulation recommend a variety of detection strategies across a battery of tests to maximise discriminatory power. Investigations into specific domains, such as cognitive symptomatology, are proposed as an additional focus for malingering research. This study tested the utility of a battery of measures focussing on cognitive functioning to distinguish malingerers from healthy controls and true patients using a simulating malingering design. One hundred and five participants completed a battery of tests which were compared to a pre-existing malingering screening tool. The simulating malingerers reported on the strategies they used to feign illness via a qualitative interview. Findings indicated that the battery of cognitive measures had better discriminatory power than the screening tool. The utility of using certain detection strategies are discussed. The approaches malingerers take are compared to the detection strategies used in malingering tools. Implications for future research and practice are considered.

2.1 Introduction

2.1.1 Definition and prevalence of malingering

Dissimulation refers to a variety of response styles relevant to the assessment of symptom validity. Malingering is distinctly defined as the conscious fabrication or gross exaggeration of physical and/or psychological symptoms for an external goal (American Psychiatric Association, 1994). It is further distinguished from factitious disorders in that the malingered symptomatology is not motivated by embellishing the patient role and is explicable in the context of the individual (Rogers, 1997).

The estimated prevalence of malingering ranges from 1/6th of forensic psychiatric patients (Rogers & Cruise, 2000) to between 8% and 30% of populations deemed to be at risk of malingering (Mittenberg, Patton, Canyock & Condit, 2002). In a review of the literature, Rogers, Harrell & Liff (1993) suggested 50% of claimants of neurological injury were feigning or exaggerating their cognitive deficits. In claims of diminished responsibility due to insanity, it is estimated that 21% of defendants are suspected malingerers (Rogers, 1986). Not only is malingering more prevalent than is commonly suspected, the frequency is compounded by the seriousness of the consequences of successful malingering, particularly in criminal cases (Rogers, 1997).

2.1.2 Use of psychological testing in medicolegal contexts

Neuropsychological evidence is becoming increasingly accepted in civil litigation cases where large sums of money can powerfully reinforce the temptation to malinger (Vickery, Berry, Inman, Harris & Orey, 2001). Scientific evidence is also more commonplace in criminal cases. It is becoming more regularly used in prolific crimes such as burglary and car theft and is almost always used in murder trials (POST, 2005).

Successful malingerers cost in terms of money, time and resources (Rogers, 1997). With the numbers of proceedings in which there is the potential motivation to feign impairments increasing (e.g. Reynolds 1998), the need for valid methods of identifying malingerers grows. However, detection of malingering using standard instruments, as opposed to tests specifically designed to detect malingering, is unsatisfactory (Ziskin, 1984). The American Psychiatric Association Neuropsychology Division has been recently recommended that every forensic evaluation incorporate a measure of symptom validity (http://www.div40.org/). This, coupled with the impact of successful malingerers on services, emphasises the need

to develop empirically sound tests of malingering (Rogers, 1997).

2.1.3 Methodological and conceptual issues in the development of malingering detection tools

Methodological difficulties plague the development of empirical instruments designed to detect malingering. The presentation of any one malingerer is highly idiosyncratic depending on the contextual and personal variables motivating the individual to feign symptoms (Rogers, 1990). Furthermore, the nature of the feigned symptoms will vary according to the level of understanding of the condition one is attempting to malinger (Pankratz & Binder, 1997). Heterogeneity in malingering conflicts with the conceptual underpinning of standardisation which seeks to identify common characteristics of malingerers. In order to overcome this heterogeneity, tools for measuring malingering need to include a number of different indices. The development of a tool to measure malingering must reach a compromise by seeking to minimise the number of false positives (i.e. inadvertently identifying true patients as malingerers) and identify the symptoms malingerers demonstrate which true patients do not.

Considering the likelihood of false positives, results from malingering tools which rely on self-report descriptions of psychiatric symptoms alone should be corroborated with further evidence of malingering (Rogers, 1997; Miller, 2001). A battery of tests which incorporate several detection strategies, such as floor effect, magnitude of error, symptom validity testing and performance curve, have greater validity in identifying malingerers (Bender, 2002). In addition, tests measuring malingering in a specific domain could be used to substantiate evidence of feigned mental illness (Rogers, 1997). One such method would be to focus additionally on cognitive symptoms (Halligan, Bass & Oakley, 2003). Incorporating a variety of detection strategies into a comprehensive test battery heightens the discriminatory power of the profile. Such a test battery would also more likely satisfy the *Daubert* criteria,

which outline the standards for the admissibility of scientific evidence in court (Hom & Denny, 2002). A brief review of a range of strategies developed by psychologists to detect malingering follows.

Floor effects

Floor effect indexes suggests that failure on simple tasks differentiates malingerers from severely impaired persons who normally answer correctly. Examples of such tests include the Rey 15 Item Memory Test (see Lezak, 1983) and The Wiggins and Brandt Personal History Interview (Wiggins & Brandt, 1988). However, it is considered that the transparency of such a technique does not identify more sophisticated malingerers (Rogers, Harrell & Liff, 1993). This strategy can be strengthened by adding another approach termed magnitude of error. This approach originated with Ganser Syndrome which proposes that proximate answers are more likely of malingerers (i.e. near misses) as they betray a knowledge of the true answer. Research into both 'near misses' and 'gross errors' (e.g. Bruhn & Reed, 1975) have identified this as a discriminating strategy. A test incorporating simple questions which require numeric responses facilitates recording the magnitude of error whilst simultaneously measuring floor effect.

Forced-choice measures

Forced-choice tests detect malingerers by identifying a pattern of responses that is significantly worse than chance responding. If a test required a forced-choice between two possible responses (e.g. presence or absence of symptoms), then the chance responder would be on average 50% correct. The Test of Memory Malingering (TOMM) (Willison & Tombaugh, 2006) is an example of such a method. A strength of this approach is the lack of alternative explanations for significantly below-average responding. A relative weakness is that in simulation designs, few simulating malingerers score significantly below average (Rogers, Harrell & Liff,

1993). Thus, this method tends to have low sensitivity with only the most blatant malingerers detected.

Symptom validity indices

Self-report measures that contain validity scales designed to detect if respondents are biasing their responses are known as validity indices. Examples of tests with validity scales built into them include the Minnesota Multiphasic Personality Inventory - 2 (MMPI-2) (Ben-Porath et al., 1995), the Personality Assessment Inventory (PAI) (Morey, 1991) and the Millon Clinical Multiaxial Inventory-3 (MCMI-III) (Millon & Meagher, 2004).

Problems are encountered using self-report symptom validity measures where severely impaired psychiatric patients' self-appraisal of symptomatology may be influenced by their condition (Rogers, 1997). In some cases, 'true' patients may endorse symptoms which are not present, thus undermining the theoretical utility of measuring psychiatric symptoms to identify malingerers. In addition, subjective self-report measures naturally lend themselves to the potential to malinger, facilitating the fabrication process. Halligan, Bass & Oakley (2003) recommend corroborating SVT's with measures tapping specific domains, such as cognitive symptomatology.

2.1.4 Rationale for using cognitive measures to detect malingerers of psychiatric illness

Research has indicated that malingerers of schizophrenia have attempted to feign cognitive symptoms such as attention, concentration and memory (Clark, 1988). The cognitive profile of mental illness is less overtly observable than psychiatric symptoms and therefore theoretically less easy to imitate. In addition, cognitive symptoms of mental illness are likely less well-known to the lay person. This is perhaps because the public are less aware of preserved areas functioning in

sufferers of severe impairments, making it harder to report a constellation of symptoms which are indicative of a genuine deficit. This potentially supports the use of cognitive symptom inventories to identify malingerers.

Malingering is thought to be an active process requiring a significant amount of cognitive effort (Alban, 2003). Longer reaction times indicate additional cognitive effort not normally recorded in severely impaired patients. Tasks which tap largely preserved areas of functioning such as preattentive as opposed to effortful processing (Anscombe, 1987) are better placed to distinguish dissimulation. Therefore, tests requiring automatic processing which are relatively easily performed by those with severe impairments may expose deception by virtue of the time taken to suppress a truthful response and generate a false response. An example is the Line Bisection Task (LBT) which has been used previously to differentiate malingerers from genuine responders (e.g. Khan et al., 2000) and has support as a test which partly requires preserved preattentive processing which, in turn, influences performance (Shulman et al., 2002).

2.1.5 Utility of computerised tests of malingering

Timing tasks which require preattentive processing provides a measure of excessive cognitive effort which could potentially distinguish malingerers from honest respondents. Computerising tasks facilitates comparisons between reaction time and response accuracy, and has been used successfully to improve detection of symptom validity in forensic populations (e.g. Kertzman-Semion et al., 2006). In addition, research suggests that deceptive responses require longer reaction times and are independent of practice effects (Vendemia et al., 2005). At best, computerised tests of information processing have been shown to outperform traditional forced-choice measures such as the Test of Memory Malingering (TOMM) (Willison & Tombaugh, 2006).

2.1.6 Strategies used by malingerers

There is very little qualitative investigation into the strategies malingerers use to fake tests. Demonstrating consistency has been suggested as a response strategy on cognitive malingering tests (Demarkis, 1999). Qualitative aspects of malingering on memory tests have identified poor cooperation, aggravation, frustration, slow response times and general confusion as potential strategies used by malingerers (Iverson, 1995). Gaining knowledge of a specific condition, for example, head injury, has been indicated as another malingering technique (Huskey, 2006). However, preparation has not been found to improve malingering performance (Tan et al., 2002). Strategies have been investigated to determine the types of illness most commonly feigned (Cohn, 1995). No studies have explored a link between test detection strategy and the associated malingering style adopted in response to the test malingerers are presented with.

2.1.7 Summary

Research supports the use of a battery of tests to evaluate symptom validity (e.g. Rogers, 1997). A battery incorporating a test of floor effect, general psychiatric symptom validity, further investigation into a specific domain (e.g. cognitive symptomatology), reaction time to measure excess cognitive effort and magnitude of error would include the majority of the most reliable methods for the detection of dissimulation researched to date. Further qualitative investigation into the strategies used by malingerers according to the tests they are presented with is needed.

2.1.8 Objectives of Study

The primary aim of the research will be to assess the utility of a battery of cognitive tests to discriminate malingerers of psychiatric illness from 'true' patients and healthy controls. The performance of this battery will be compared to a standardised instrument, the Miller Forensic Assessment of Symptoms Test (M-FAST) (Miller,

2001), which tests malingering by assessing psychiatric symptoms. Secondary aims include:

- Testing the performance of three groups (simulating malingerers, true patients and healthy controls) on a battery of tests presented by computer designed to measure bogus symptom reporting and response style on simple cognitive tests.

- Comparing the discriminatory power of each of these tests, both individually and in combination as a test battery, against data from an extant instrument (M-FAST).

- To assess the degree to which true symptoms are associated with performance on the test battery and the M-FAST.

- To obtain information on the strategies used by simulating malingerers to fake the tests.

2.2 Method

2.2.1 Participants

One hundred and five participants were recruited, of which 30 were assigned to the healthy control group, 40 to the simulating malingering group and 35 psychiatric inpatients forming the true patient group. Previous research has yielded large effect sizes (e.g. McMennemin, *in preparation*). Regarding the differences between the 3 groups on the Cognitive Dysfunctions Questionnaire (Coxell), a power calculation was performed using the Zumastat 2.3 software. Means and standard deviations (in parentheses) for the 3 groups were 4.2 (.58), 4.94 (.59) and 6.14 (1) respectively.

Given a pooled standard deviation of 0.75, 17 participants are needed in each group in order to achieve a power level of 0.80, with a difference of 0.74 between the groups at an alpha = 0.05 level of significance. Additional participants were assigned to the simulating malingering group to enhance the reliability and validity of additional qualitative and quantitative measures administered to this group only.

2.2.2 Sampling

a) Healthy control and simulating malingering groups: participants were recruited using opportunity sampling. Inclusion criteria were the same for the 2 groups and are listed as follows:

- Native English speakers
- No history of severe mental illness (e.g. Schizophrenia, Bipolar Affective Disorder)
- No Learning Disability
- No history of dyslexia or other reading difficulty
- No history of severe head injury (loss of consciousness > 10 minutes)
- No history of neurological disease (e.g. epilepsy)
- No visual impairment not corrected by glasses or contact lenses

b) Patient Group: participants were psychiatric inpatients recruited from acute inpatient mental health units across 2 sites of West London Mental Health NHS Trust based at Charing Cross and Ealing Hospitals. Inclusion criteria were applied as follows:

- Native English speakers
- A current diagnosis of severe mental illness (e.g. Schizophrenia, Bipolar Affective Disorder)

- No Learning Disability
- No history of dyslexia or other reading difficulty
- No history of severe head injury (> 10 minutes unconscious)
- No history of neurological disease (e.g. epilepsy)
- No visual impairment not corrected by glasses or contact lenses
- No involvement in any medico-legal proceedings (e.g. compensation seeking)

2.2.3 Measures

i) *Weschler Test of Adult Reading* (WTAR) (Wechsler, 2001).

The WTAR is comprised of a list of 50 infrequently used words that the participant must pronounce aloud as best as possible. The WTAR has been co-normed against the Wechsler Adult Intelligence Scale III (WAIS-III) (Wechsler, 1997) and the Wechsler Memory Scale III (WMS-III) (Wechsler, 1997a), allowing an estimation of pre-morbid intelligence. Totalled raw scores are converted into standard scores which can be compared to a UK national normative sample. Raw scores only were used for the purposes of this study, namely to compare intelligence between groups rather than to the wider population.

ii) *Brief Symptom Inventory* (BSI) (Derogatis, 1975)

The BSI is a 53-item self-report questionnaire identifying psychiatric and medical symptoms. Items are rated on a 5-point scale measuring severity from (0) "Not at all" to (4) "Extremely". Respondents are asked to rate the presence of each item over the last 7 days. Scoring yields a Global Symptom Index designed to estimate severity of psychiatric illness. The questionnaire has been extensively normed against national populations and various patient samples. Reliability and validity of the instrument have been tested in more than 400 studies (http://www.pearsonassessments.com/tests/bsi.htm).

iii) *Line Bisection Task* (LBT) (Schenkenberg et al., 1980).

This version consists of a series of 26 pairs of horizontal lines which are both bisected by a vertical line. The task requires participants to identify which line is bisected closest to the midpoint of the horizontal lines. Bias in estimating the centre away from the side neglected is expected in patients with lateral neglect. This task is eminently possible for most patients barring those with specific lesions in certain areas of the brain. LBT's are simple tasks which identify malingering by a respondent showing inadequate knowledge of how a patient would respond. In addition, the task requires little conscious effort and incorporates preattentive processing (Shulman et al., 2002). Significantly greater reaction times therefore may be suggestive of a suppression of the pre-potent response, greater cognitive effort and thus serve to identify malingerers.

iv) *Ganser* (created from previous research)

This test is a computerised questionnaire comprising items which require only basic knowledge and are normally answered correctly even by the severely impaired. Scoring incorporates a magnitude of error malingering detection strategy whereby frequent incorrect answers which are close in proximity to the correct answer are hypothesised to betray a knowledge of the correct answer, thus indicating malingering. Reaction time and participant responses are recorded. Magnitude of error is computed from the absolute difference between the patient response and the correct response.

v) *Cognitive Dysfunctions Questionnaire* (CDQ) (Coxell)

This questionnaire was developed by Coxell (McMennemin, *in preparation*). It comprises 69 items of which 7 describe extremely rare symptoms and 22 describe fictitious symptoms. The rest of the items are neutral and add face validity as a genuine screening tool. Items that are endorsed are followed up after testing and

participants were asked to rate the frequency and severity of the symptoms ranging from (1) "Only once" to (6) "All of the time" and (1) "Not at all" to (5) "Severely distressing", respectively. The computerised version automatically sums the total score, total reaction time, sum scores of subscales including total distress rating, total frequency rating and total 'not sure' rating. An additional score was computed summing only the fake and rare items of the measure.

vi) *Miller Forensic Assessment of Symptoms Test* (M-FAST) (Miller, 2001).

This measure is a 25-item structured screening interview which provides information on the probability of the respondent feigning psychiatric illness. Respondents are asked whether they experience fictitious symptoms which are not commonly endorsed by genuine psychiatric patients. Scores are totalled according to the number of items endorsed. A cut-off score of 6 or more is indicative of malingering in clinical settings (Miller, 2001). Further information about malingering style can be gleaned from subscales which differentiate amongst malingerers. The use of this established screening tool for malingering afforded criterion validity.

vii) *Qualitative Questions*

After the battery of tests were administered, a series of 6 semi-structured qualitative questions were directed to the participants in the simulating malingering group. These questions were designed to:

- glean qualitative information about the techniques used to malinger.
- investigate whether the participants were able to distinguish genuine psychiatric symptom questions from false ones.
- give an indication of whether participants altered their strategy when presented with tasks utilising different malingering detection approaches.

The following questions were asked and responses recorded in note form by the

principal researcher.

a) How did you respond to the questions to appear mentally ill?

b) Did some questions appear more likely to be genuine symptoms of mental illness than others?

c) (If answered 'yes' to (b)) Why did you trust some questions over others?

d) How confident were you about appearing to be mentally ill?

e) What method did you use to malinger?

f) Did you choose a mental illness to malinger, if so which one...?

2.2.4 Design and Procedure

Malingering research typically uses a fully controlled simulation design (Schretlen, 1988). A simulating malingering group comprised of healthy volunteers is recruited and given an incentive (e.g. monetary) to feign mental illness. Their responses are compared to those of a true patient group and a healthy control group. This yields data on the malingering and true patient group allowing for discrimination between groups. By nature of their subterfuge and often current engagement in court proceedings, undetected malingerers are impractical to recruit.

All participants were required to complete the test battery in the same order, as follows:

i) WTAR

ii) BSI

iii) LBT

iv) GANSER

v) CDQ

vi) M-FAST

In addition, the simulating malingering group were then asked to complete the

following section:

vi) Semi-structured qualitative interview

The true patient group and healthy control group were asked to complete all tasks honestly. After completing the WTAR and the BSI, the simulating malingering group were asked to imagine a scenario in which they were required to convince a court that they had a mental illness by the way they responded to the questions on the rest of the tasks.

For the sake of consistency and future comparison, the malingering group were given the same instructions as had been used in a previous study testing the CDQ (McMennemin, *in preparation*). These instructions were given as follows:

"I want you to respond to all the subsequent tasks in a way that you believe a person with a severe mental illness would. You should try to portray having a serious mental illness in as realistic and convincing a manner as you can. Imagine that you are in a legal predicament (e.g. you have committed a serious offence) and you believe it is in your interest to appear to be seriously mentally ill, and therefore less responsible or not responsible for what you have done. Perhaps, for example, you want to appear to be unfit to plead at a criminal trial. I want you to answer all subsequent questions about symptoms and to perform all subsequent tasks in this way. The person whose scores on these tests best approximate those of true patients with severe mental illness will be given a prize of £50. It is therefore in your interests to answer in as realistic and convincing a manner as possible. I will not give you any further information at this stage about serious mental illness, as I am interested in your perception of this in regard to answering subsequent questions."

The simulating malingerers were given a monetary incentive of £50 to be awarded to the most convincing malingerer. This was defined as the participant whose scores on the test battery were closest to the average test battery scores of the true patient group.

Following completion of the test battery, participants were given the opportunity to ask questions about the research. Descriptive and raw data were entered into a statistical software package (SPSS). Raw paper data, consent forms and electronic data were stored in accordance with data protection protocols.

2.2.5 Data analyses

Quantitative analyses: Analyses of the data presented in the results section are in the following order. Group differences for the demographic information, including gender, age, intelligence measured on the WTAR and symptom reporting measured on the BSI are analysed using a one-way analysis of variance (ANOVA), excepting gender which is analysed using Chi Square. Levene's test was used to assess homogeneity of variance. Where it could not be assumed, a Kruskal-Wallis non-parametric test was used to corroborate findings from the ANOVAs. Post hoc contrasts were computed to investigate differences between pairs of groups.

Criterion validity was assessed using an ANOVA on the M-FAST total score. This was computed to investigate whether an established test of malingering differentiated between groups. Post hoc contrasts were performed to explore differences between pairs of groups and Levene's test for homogeneity of variance to investigate whether groups had comparable variances. Where homogeneity of variance could not be assumed, a Kruskal-Wallis non-parametric test was computed to corroborate the ANOVA. A cut-off of 6 on the M-FAST was used to create a categorical variable, separating M-FAST scores into 2 groups. A Chi Square test was performed to assess

differences between groups on this variable.

Outcome variables on the malingering test battery were as follows:

- LBT total score
- Ganser total score
- Ganser mean reaction time score
- Ganser absolute value score
- CDQ total time
- CDQ total score
- CDQ frequency total
- CDQ distress total
- CDQ 'not sure' total
- CDQ combined formula

The combined CDQ formula was computed as follows: Total CDQ score for malingering and rare symptom items only x (total frequency score for malingering and rare CDQ items x total distress score for malingering and rare CDQ items).

The Ganser absolute value was derived from the difference between the correct score and the recorded score (regardless of positive or negative values), yielding the absolute difference, or magnitude of error score. These were subjected to one-way analyses of variance to determine differences between malingering, true patient and healthy control groups. Levene's test was used to assess homogeneity of variance. Where it could not be assumed, a Kruskal-Wallis non-parametric test was used to corroborate findings from the ANOVAs. Post hoc contrasts were computed to investigate differences between pairs of groups.

The outcome variables for the malingering battery were converted into stanine scores

to afford a simple ROC curve comparison. For variables, CDQ formula, Ganser absolute total, CDQ fake and rare symptoms total and CDQ distress total, the stanine scores were inverted so that for all stanine variables likelihood of malingering was positively correlated with stanine score. A Receiver Operating Curve analysis was performed on the true patient and simulating malingering groups to examine discriminatory power of the malingering battery. These scores were compared to a Receiver Operating Curve analysis for the M-FAST, an established test of malingering.

An additional composite stanine score (Total Stanine) was computed to create a stanine that represented the malingering battery. The composite stanine consisted of the combination of malingering variables which best discriminated between groups. CDQ 'not sure' total was excluded from the composite score as the number of respondents who recorded 'not sure' answers were less than a third of the participants (N=32). CDQ frequency total was also excluded from the composite score as the variance between groups was almost double the mean (e.g. for the malingering group, mean = 270, s.d.= 433), rendering the discriminatory power of the variable unreliable. A maximum stanine score (Max Stanine) was computed from the composite (Total Stanine) variable to provide a variable which conflated all the outcome variables.

The final section of the quantitative analysis explored the confound between true pathology, demographics, M-FAST and Total stanine score. Correlations between WTAR, BSI, age, and Total stanine score for the malingering battery are presented for the true patient group only. Correlations between WTAR, BSI, age and M-FAST scores are also presented for the true patient group only.

Qualitative analyses: A semi structured interview explored the strategies simulating

malingerers used to fake the test battery. The data was analysed using a triadic elicitation technique yielding discrete categories. This technique requires selecting 3 participant answers at random. The 2 which are the most similar are matched to determine the characteristic of the category. The third is returned to the pool of answers and the process is repeated until no further categories can be elicited (Postlethwaite & Jaspars, 1986).

Questions 1 (How did you respond to the questions to appear mentally ill?) and 5 (What method did you use to malinger?) sought to elicit information pertaining to techniques used to malinger. On analysis, 7 strategies were elucidated from question 1. Analysis of question 5 yielded malingering styles which fitted the same categories as question 1, lending further support to the systematic use of these techniques.

Questions 2 (Did some questions appear to be more like mental illness than others?) and 3 (If so, why did you trust some questions over others?) regarded the apparent genuineness of the questions.

Question 4 (How confident were you about appearing to be mentally ill) investigated participants degree of confidence in feigning on the tests. Question 6 (Did you choose a mental illness to malinger, if so which one...?) explored the mental illness profile that respondents chose to portray.

2.2.6 Ethics

Following application for and receipt of appropriate ethical approval (Appendix 1), each participant was presented with an information sheet (Appendix 2) detailing the research aims, the right to withdraw at any point without consequence, their voluntary participation and what taking part involved. Consent to participate was recorded on a consent form (Appendix 3).

2.3 Results

2.3.1 Demographics

Means and standard deviations of demographic variables for malingering, true patient and healthy control groups are displayed in Table 1. Means are reported with standard deviations in parentheses, excepting sex, which is reported as number in each group with percentage of group membership in parentheses.

Table 1: Group demographic means and standard deviations

	Groups		
Variable	Malingering	True Patient	Healthy Control
Sex			
Male	17 (42.5%)	21 (60%)	17 (56.7%)
Female	23 (57.5%)	14 (35%)	13 (43.3%)
Age*	31.5 (8.3)	37.9 (11.3)	26.8 (11.1)
WTAR*	40.5 (6.7)	35.2 (8.3)	46.2 (3.5)
BSI*	22.8 (25.4)	76.2 (56.4)	6.5 (8.9)

Note: *Significant at $p < 0.01$, otherwise NS

Sex did not differ significantly between the groups (X = 2.60; df=2; NS). One-way analyses of variance found significant differences between groups according to age ($F(2,102) = 9.67$, $p < 0.01$), pre-morbid intelligence measured on the WTAR ($F(2, 102) = 22.15$, $p < 0.01$) and self-report of psychiatric symptomatology on the BSI ($F(2,102) = 33.73$, $p < 0.01$).

Post-hoc tests found that differences between groups according to age were due to difference between true patient and healthy control groups ($p < 0.01$) and malingering and true patient groups ($p < 0.01$) but not malingering and healthy control groups. Differences between groups according to a pre-morbid estimate of

intelligence found that all groups were significantly different ($p < 0.01$). Differences between groups according to self-reported psychiatric symptomatology also found that all groups differed significantly on the BSI ($p < 0.01$).

Homogeneity of variance was assessed using Levene's test and found that age, WTAR and BSI equality of variance could not be assumed. Kruskal-Wallis non-parametric tests corroborated the analyses of variance findings that there were significant differences between groups according to age, BSI and WTAR scores, but not sex.

2.3.2 Distinguishing between groups

A one-way analysis of variance showed that the M-FAST successfully discriminated between groups ($F(2, 102) = 60.22$; $p < 0.01$) (see table 2). Post hoc analyses found all groups differed significantly from each other. Levene's test for homogeneity of variance indicated equal variance should not be assumed and so a Kruskal-Wallis non-parametric test was performed to corroborate the analysis of variance ($\chi 2 = 63.77$; $df = 2$; $p < 0.01$).

Table 2: Means and standard deviations for the M-FAST

Variable	Mean	Standard deviation
M-FAST total score		
Malingerers*	12.8	5.3
True patients*	5.2	5.7
Healthy controls*	0.7	1.2

Note: * denotes significance at $p < 0.01$

A categorical variable was computed using the recommended cut-off of a total score of 6 to indicate malingering. There was a strong association between being classified as a malingerer and membership of the malingering group ($\chi 2 = 59.93$; $df = 2$; $p < 0.01$). Eleven participants in the true patient group were classified as malingerers, 36 in the simulating malingering group and none in the healthy control group (see table

3). This equates to 31% of true patients being classified as malingerers, which is just over the upper limit of expected percentages of malingerers in psychiatric patients.

Table 3: Classification of malingering on the M-FAST

Group	Classified as malingerers	Classified as non-malingerers
Malingerers	36	4
True patients	11	29
Healthy controls	0	40

2.3.3 Group differences on tests of malingering

The following section displays results from these analyses ordered by variable.

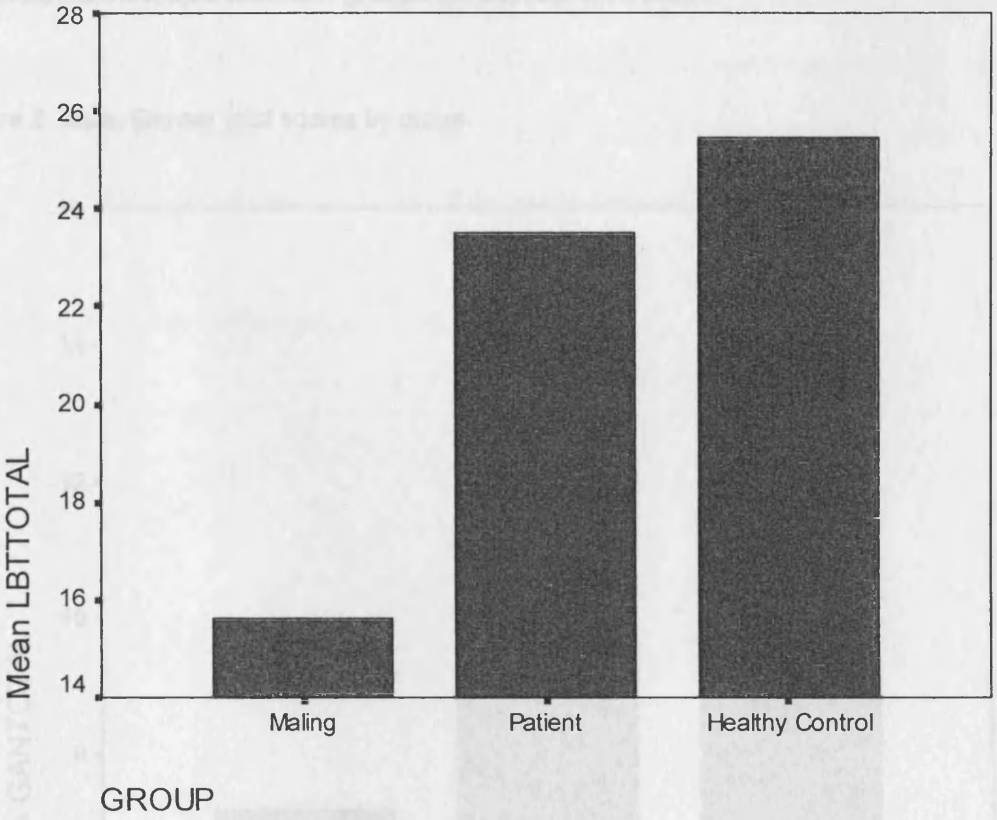
LBT total score

Mean LBT total scores differed significantly between groups ($F(2, 102) = 32.31$, $p < 0.01$), (figure 1).

Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.05$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers made significantly more errors than both the true patient and healthy control groups on the LBT.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on LBT total score.

Figure 1: Mean LBT total scores by group



Ganser total score

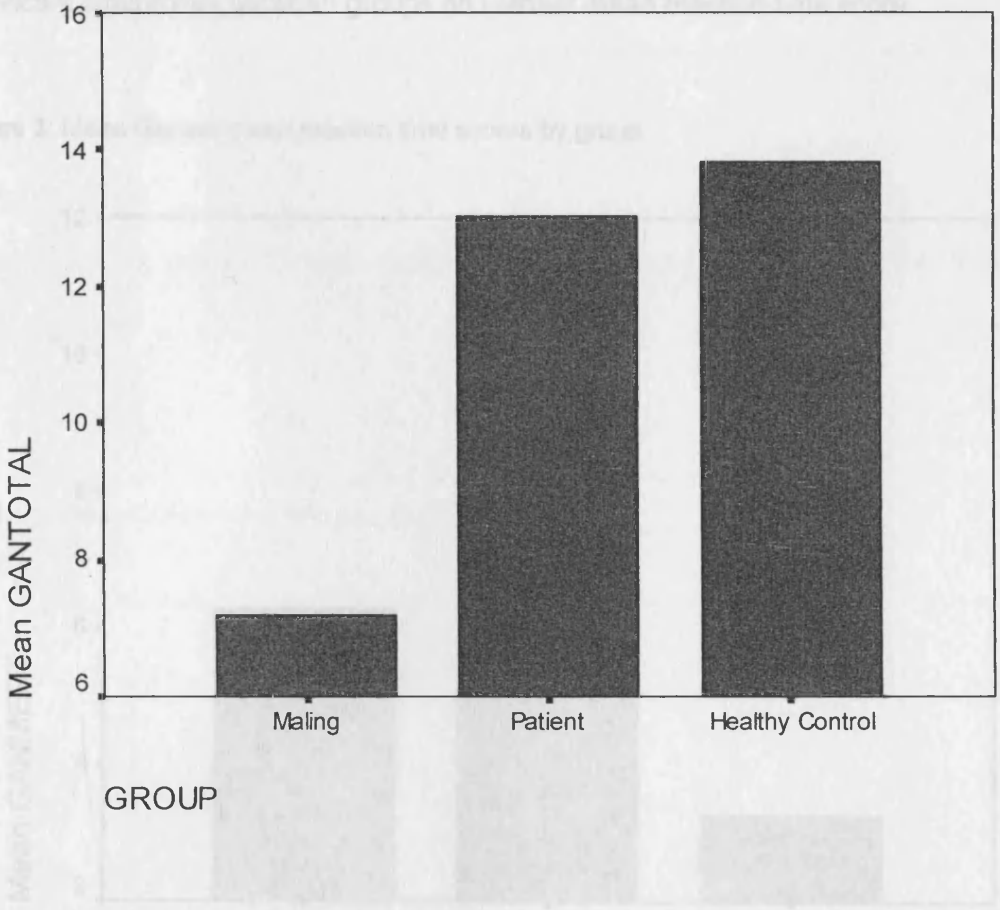
Mean Ganser total scores differed significantly between groups ($F(2, 101) = 52.53$, $p < 0.01$), (figure 2).

Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers made significantly more errors than both the true patient and healthy control groups on the Ganser.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were

computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on Ganser total score.

Figure 2: Mean Ganser total scores by group



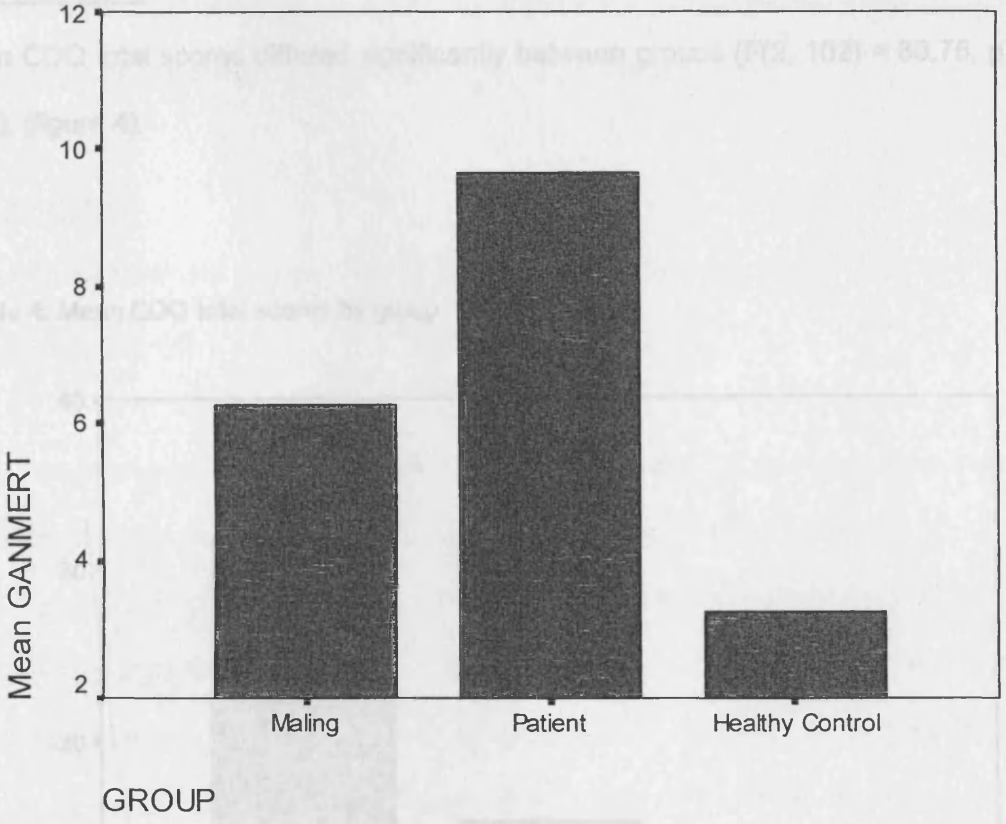
Ganser mean reaction time

Mean Ganser mean reaction time scores differed significantly between groups ($F(2, 100) = 22.78$, $p < 0.01$), (figure 3).

Post-hoc analyses indicated significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). Unexpectedly, the true patient group took longer to complete the Ganser than the other 2 groups. However, simulating malingerers took significantly longer than healthy controls, as expected.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on Ganser mean reaction time score.

Figure 3: Mean Ganser mean reaction time scores by group



Ganser absolute magnitude of error score

The data were not suitable for parametric testing. A Kruskal-Wallis non-parametric test found significant differences between groups (χ^2 = 42.79; df = 2; $p < 0.01$). Both gross errors and proximal answers are indicative of malingering for this variable. Thus, any significant deviation from the correct score suggests malingering. Extreme variance in scores makes these data hard to interpret graphically. Table 4 reports the

means and standard deviations.

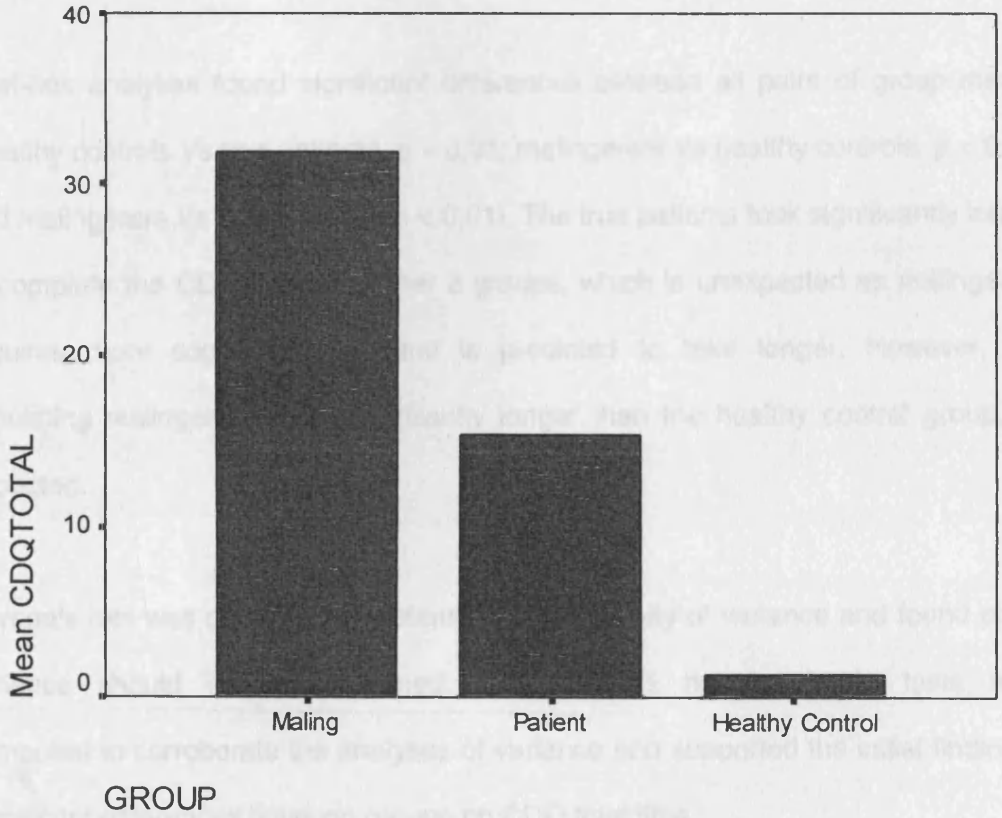
Table 4: Ganser absolute magnitude of error means and standard deviations by group

Group	Mean	Standard deviation
Malingerers	118101	73005897
True patients	41.6	88.7
Healthy controls	0.5	1.9

CDQ total score

Mean CDQ total scores differed significantly between groups ($F(2, 102) = 60.75$, $p < 0.01$), (figure 4).

Figure 4: Mean CDQ total scores by group



Post-hoc analyses found significant differences between all pairs of group means

(healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers reported significantly more cognitive symptoms than both the true patient and healthy control groups.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on CDQ total score.

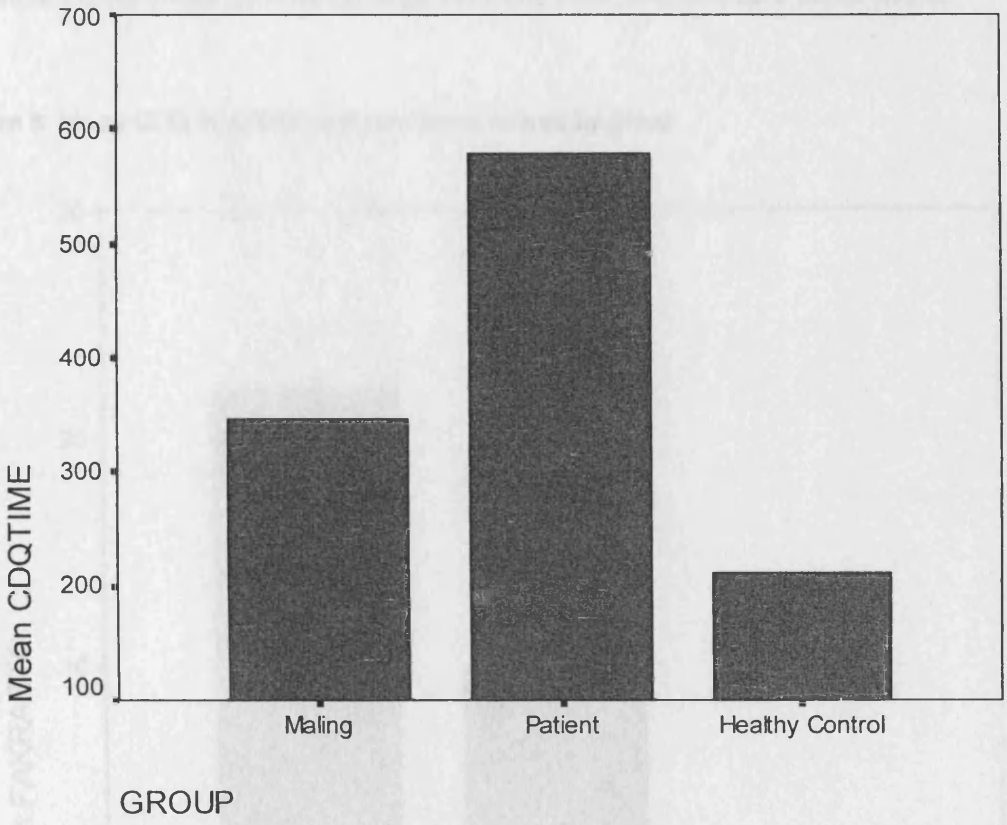
CDQ total time

Mean CDQ total time scores differed significantly between groups ($F(2, 91) = 36.09$, $p < 0.01$), (figure 6).

Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). The true patients took significantly longer to complete the CDQ than the other 2 groups, which is unexpected as malingering requires more cognitive effort and is predicted to take longer. However, the simulating malingerers took significantly longer than the healthy control group, as expected.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on CDQ total time.

Figure 5: Mean CDQ total time scores by group



CDQ total fake and rare items score

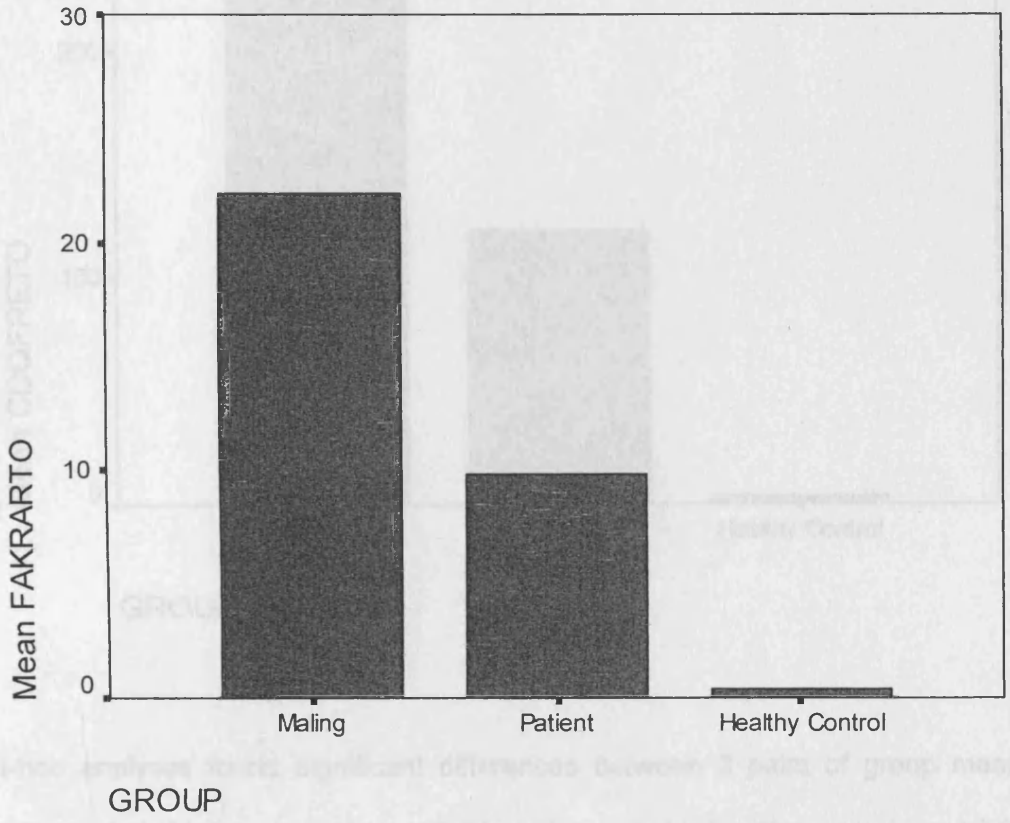
Mean CDQ total fake and rare items scores differed significantly between groups ($F(2, 102) = 56.95$, $p < 0.01$), (figure 6).

Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$ and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers reported significantly more fake and rare cognitive symptoms than both the true patient and healthy control groups.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were

computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on CDQ total fake and rare items score.

Figure 6: Mean CDQ total fake and rare items scores by group

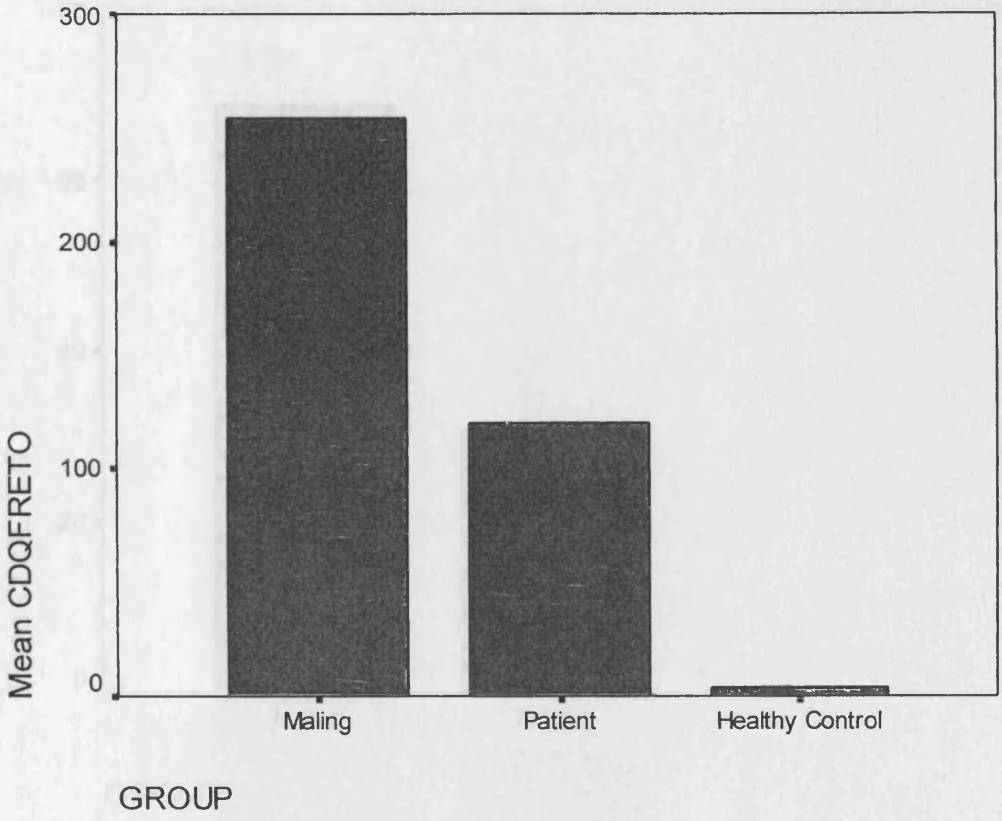


CDQ total frequency score

Mean CDQ total frequency scores differed significantly between groups ($F(2, 102) = 7.04$, $p < 0.01$), (figure 7).

Post-hoc analyses found significant differences between 2 pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; but not malingerers Vs true patients, NS). As expected, simulating malingerers reported suffering significantly more frequent cognitive symptoms than both the true patient and healthy control groups.

Figure 7: Mean CDQ total frequency scores by group

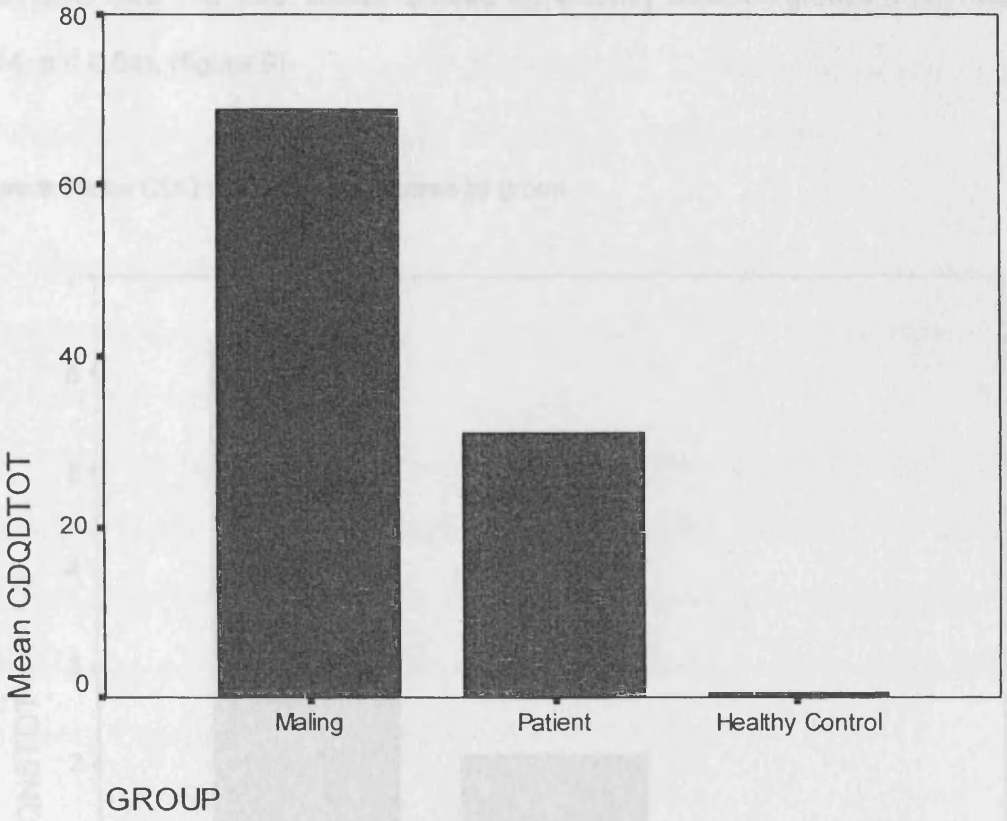


Post-hoc analyses found significant differences between 2 pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; but not malingerers Vs true patients, NS). As expected, simulating malingerers reported suffering significantly more frequent cognitive symptoms than the healthy control group.

CDQ total distress score

Mean CDQ total distress scores differed significantly between groups ($F(2, 102) = 25.20$, $p < 0.01$), (figure 8).

Figure 8: Mean CDQ total distress scores by group



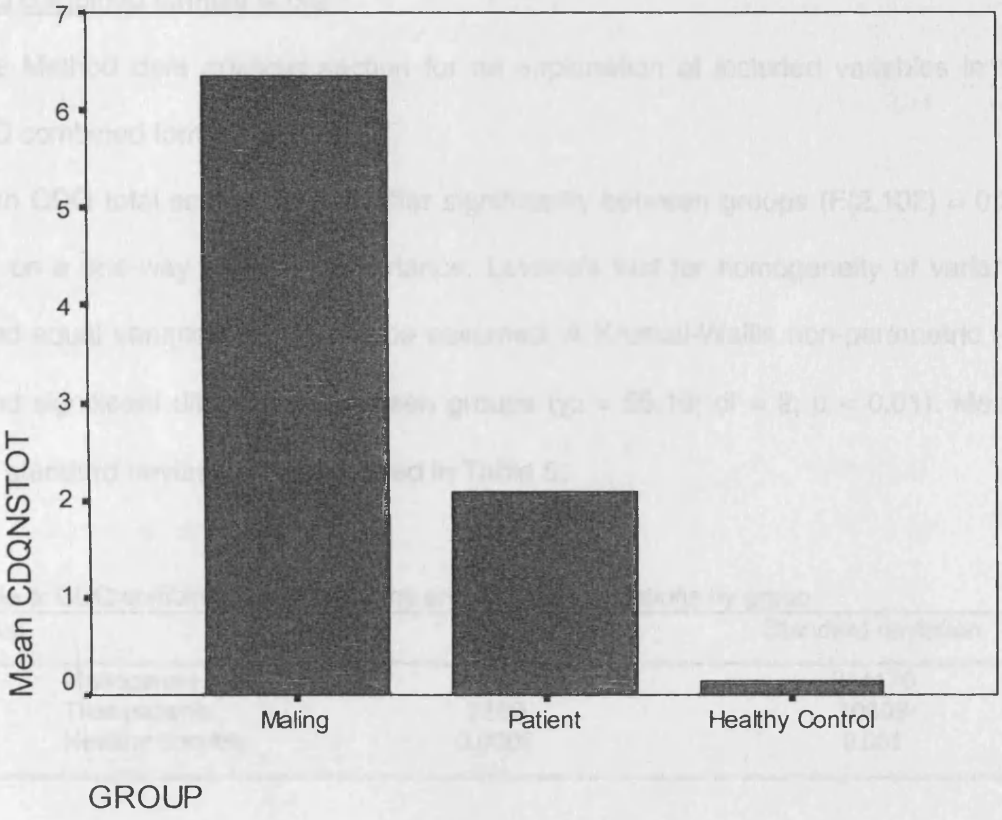
Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.01$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers reported suffering significantly more distressful cognitive symptoms than both the true patient and healthy control groups.

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on CDQ total distress score.

CDQ total 'not sure' score

Mean CDQ total 'not sure' scores differed significantly between groups ($F(2, 102) = 12.14$, $p < 0.01$), (figure 9).

Figure 9: Mean CDQ total 'not sure' scores by group



Post-hoc analyses found significant differences between all pairs of group means (healthy controls Vs true patients, $p < 0.05$; malingerers Vs healthy controls, $p < 0.01$; and malingerers Vs true patients, $p < 0.01$). As expected, simulating malingerers reported not being sure of experiencing some cognitive symptoms significantly more often than both the true patient and healthy control groups. However, numbers of respondents who used the 'not sure' option were less than a third of the participants (N=32).

Levene's test was computed to determine homogeneity of variance and found equal variance should not be assumed. Kruskal-Wallis non-parametric tests were computed to corroborate the analyses of variance and supported the initial finding of significant differences between groups on CDQ total 'not sure' score.

CDQ combined formula score

(See Method *data analysis* section for an explanation of included variables in the CDQ combined formula).

Mean CDQ total scores did not differ significantly between groups ($F(2,102) = 0.89$, NS) on a one-way analysis of variance. Levene's test for homogeneity of variance found equal variance should not be assumed. A Kruskal-Wallis non-parametric test found significant differences between groups ($\chi_2 = 65.10$; df = 2; $p < 0.01$). Means and standard deviations are reported in Table 5.

Table 5: CDQ combined formula means and standard deviations by group

Group	Mean	Standard deviation
Malingerers	41573	244170
True patients	2150	10309
Healthy controls	0.0005	0.003

Although simulating malingerers reported a combination of number, frequency and distress of cognitive symptoms to a significantly greater degree than both the true patient and healthy control groups, a parametric test did not find this difference significant. It is possible that this combined formula does not successfully distinguish between groups on an ANOVA due to the extreme variance caused by the range of responses in the malingering group. Some malingerers recorded the most extreme response for all items, whereas others were more parsimonious in their responding. This yielded a variance almost 6 times that of the mean score (mean = 41573; s.d. = 244170).

2.3.4 Receiver operating curve (ROC) analyses

Table 6 reports the area under curve statistic yielded from receiver operating curve analyses for all malingering variables.

Table 6: Area under ROC statistics for stanine malingering variables

Stanine variable	Area under curve statistic	Confidence interval (95%)
M-FAST	**0.83**	**0.74 - 0.94**
LBT total*	0.81	0.70 - 0.91
Ganser time total*	0.71	0.60 - 0.83
Ganser total*	0.83	0.73 - 0.93
Ganser absolute total*	0.75	0.64 - 0.87
CDQ frequency total	0.71	0.59 - 0.83
CDQ distress total*	0.75	0.64 - 0.87
CDQ fake & rare total*	0.83	0.73 - 0.92
CDQ 'not sure' total	0.64	0.51 - 0.76
CDQ total*	0.76	0.64 - 0.87
CDQ formula*	0.78	0.68 - 0.89
Total stanine	**0.93**	**0.87 - 0.98**
Maximum stanine	**0.92**	**0.85 - 0.98**

Note: *Variables used to compute Total stanine and Maximum stanine variables (see rationale for including these variables in Method *data analysis* section).

As can be seen from figures 10 and 11, the Total Stanine variable outperformed the M-FAST in discriminating between inpatients and simulating malingerers. ROC analysis of the M-FAST stanine (Figure 10) scores for the true patient and malingering groups yielded a significant Area Under Curve (AUC) statistic suggestive of high discriminatory power (AUC = 0.83; 95% CI: 0.74 - 0.94; $p < 0.01$). The AUC statistic corresponds to the likelihood of identifying a malingerer over a true patient according to M-FAST score.

Figure 10: M-FAST Receiver Operating Curve

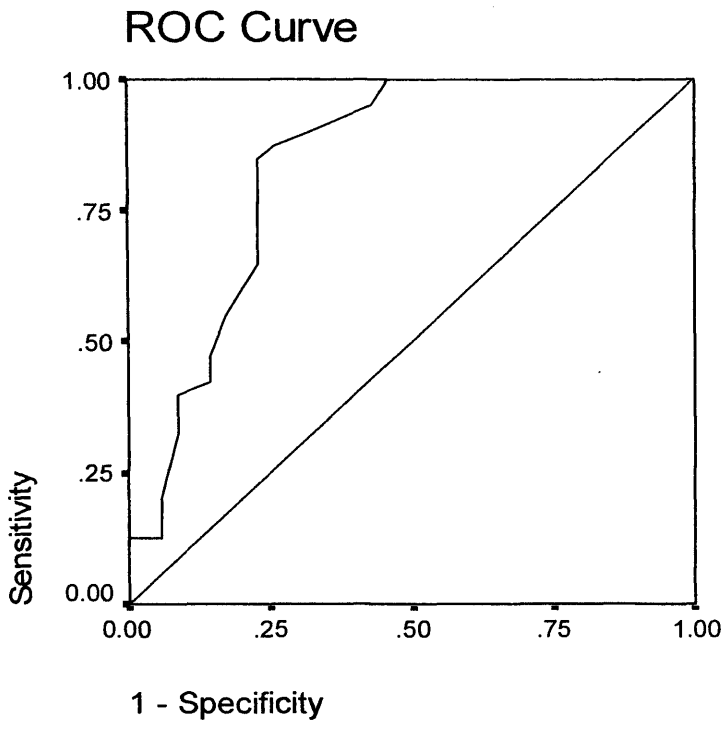
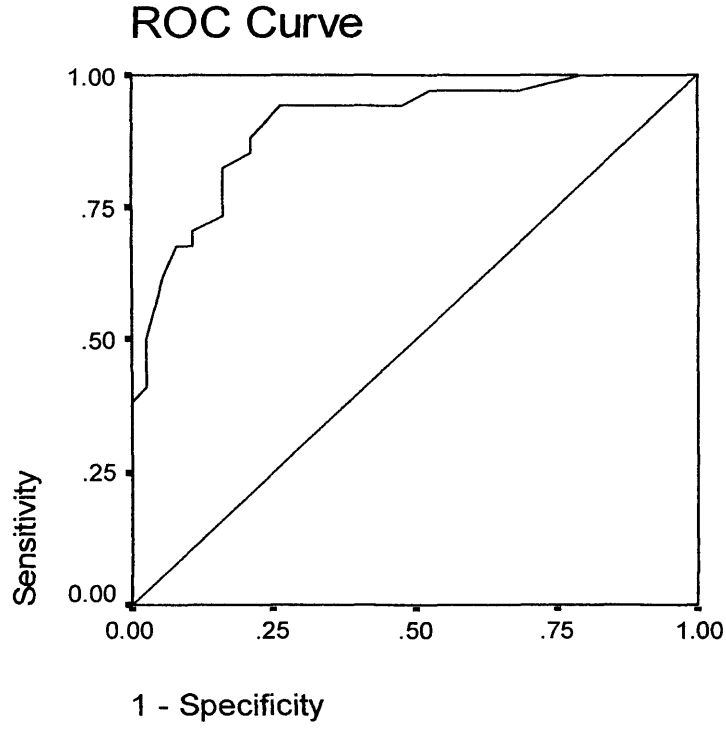


Figure 11: Total stanine Receiver Operating Curve



ROC analysis of the Total Stanine variable for the true patient and malingering groups yielded a significant Area Under Curve (AUC) statistic suggestive of superior discriminatory power (AUC = 0.93; 95% CI: 0.87 - 0.98; $p < 0.01$). The AUC statistic corresponds to the likelihood of identifying a malingerer over a true patient according to Total Stanine score. ROC analysis of the Total Stanine variable is presented in Figure 11.

2.3.5 Relationship between true pathology, age, intelligence and malingering indices

Parametric correlations were performed to indicate whether descriptive variables, WTAR, age and BSI confounded with the combined malingering score and M-FAST. Table 7 presents the correlation coefficients for these analyses. These correlations were performed on the true patient group only. To control for the likelihood of making a Type 1 error, a Bonferroni test was performed, setting significance at the alpha = 0.008 level.

Age and WTAR scores did not correlate between the Total stanine variable or the M-FAST. True pathology measured on the BSI correlated positively with the M-FAST.

Table 7: Correlations between demographic variables, M-FAST and Total stanine score

	Demographics		
Variable	Age	WTAR	BSI
M-FAST	r = -0.21	r = -0.30	r = 0.54*
Total stanine	r = 0.03	r = 0.32	r = -0.61*

Note: * significant association at $p < 0.008$

Interestingly, true pathology correlated negatively with Total stanine score. This finding suggests that the more severely ill patients were less likely to be classed as malingering according to the Total stanine variable.

2.3.6 Qualitative analysis

Analysis of the qualitative data revealed the use of a number of malingering strategies.

2.3.6.1 Malingering techniques

Respondents reported using a combination of malingering techniques, adjusting styles to particular tests and changing techniques during a single task. The strategies are described as follows.

Suppression:

Five interviewees described suppressing honest responses and substituting opposing fabricated answers in order to convey abnormality.

> "...Said the opposite of what I would normally."
>
> "...Difficult acting not what you are. Knowing not putting the correct answer."

Interviewees alluded to the cognitive effort required to suppress the pre-potent honest response.

Subtlety:

Ten interviewees expressed an attempt to convey mental illness more convincingly by not over-exaggerating their responses.

> "Not to be too obvious, not answering yes to all symptoms."
>
> "Taking the norm and slightly deviating from it, not too obviously."

A sophisticated form of this technique was used by 2 people to confound the tests by conveying preserved ability

"...Ganser truthfully, if got mental health problems, it doesn't mean memory deficit."

Imitation

Fourteen respondents attempted to mimic mental illness by exaggerating personal experiences or drawing on information gleaned from popular culture or general knowledge.

"If had own experience, tried to use that, something that might happen to me and exaggerate it."

"I imagined giving myself an ailment."

"Based on knowledge, popular knowledge, erratic, e.g. '12 Monkeys'."

Random patterns

Sixteen respondents described using a technique which ignored the content of the questions and sought to portray abnormality by answering using a random pattern or formula.

"...LBT just one button, then swapped..."

"Press the same button a lot, then stop and think on occasion."

Several of these respondents referred to using this technique on the LBT where the function of the test was less transparent.

"...First test random as couldn't work it out."

Consistency

Fourteen interviewees reported attempting to respond consistently across and within symptom inventories. This strategy was explained by some as a means of minimising the likelihood of contradicting oneself and presenting a cluster of symptoms indicative of a particular problem.

> "...Remember what had done before to answer accordingly and appear realistic."
>
> "I tried to recall patterns in the subject of the questions to have continuity and not be completely random. I thought the questions might be designed to pick holes, so I tried to keep continuity."

Inconsistency

Five participants included an element of randomness to their responding to reflect erratic behaviour that they believed was indicative of mental illness.

> "Mix of contradictory symptoms deliberately, being illogical and inconsistent."
>
> "Like I was confused about what being asked."

Global impairment

Four respondents answered yes to most symptom questions, believing that those with mental illness suffer global impairment.

> "I answered yes to everything."
>
> "Something wrong with every single one of me, everything."

2.3.6.2 Determining the veracity of the questions

82.5% (33) of simulating malingerers believed that some questions appeared more

likely to be symptoms of mental illness than others. 12.5% (5) thought that all the questions were genuine symptoms of mental illness and 5% (2) were unsure. In trying to decipher genuine symptoms from bogus questions, participants reported the following intuitions.

Guesswork

Two respondents said they guessed.

Patterns

Two respondents believed that the more questions pertained to a particular symptom, the more likely it was genuine.

> "See patterns from what asked...e.g. 'ringing in ears' was asked twice, therefore was more likely to be real."

Obscure

Four participants reported being more wary of questions which appeared too unusual.

> "...obvious trick ones, e.g. voices from right or left ear. Some were transparent, too obscure."

Stereotypes

Eighteen participants relied on their knowledge of mental health problems to guide their responding.

> "Some were more obvious, e.g. stuff you've read in books or see in movies."
> "Some seemed like true questions based on my knowledge."

Normality

Twelve respondents endorsed questions which appeared closer to normal experience.

> "Some happen to me even when normal. For example 'temper', 'yes' to these More normal experiences."
>
> "Some were closer to normality, for example, balance and memory might be exacerbated in mental illness. Or if had bothered me before."

Ninety percent of respondents (36) reported not being confident in appearing mentally ill, suggesting that the test battery was not transparent. Ten percent (4) reported a high degree of confidence in responding as if they had a mental illness. Fifty-five percent (22) of simulating malingerers did not choose a particular illness to malinger. Of those that did, 25% chose to portray a psychotic illness (e.g. schizophrenia), 5% chose depression, 7.5% (3) attempted to convey psychopathy or personality disorders and the remaining 7.5% (3) selected a neurological disorder to malinger.

2.4 Discussion

2.4.1 General discussion

The study aimed to assess the utility of a battery of tests to discriminate simulating malingerers from true patients and healthy controls. The test battery's performance was compared to that of an established malingering screening tool, the M-FAST (Miller, 2001). The individual tests forming the battery were examined individually and in combination for their discriminatory power. The extent to which true symptoms were associated with test performance was explored to give an indication of potential

confounds. The research also used a semi-structured interview to investigate the strategies simulating malingerers employed to appear mentally ill.

2.4.2 Utility of the malingering battery

In combination, the test battery outperformed an established screening tool, the M-FAST (Miller, 2001), in distinguishing between simulating malingerers, true patients and healthy controls. This supports the notion that tests utilising a variety of detection strategies increase the likelihood of identifying malingerers (Bender, 2002). It lends support to the value of cognitive malingering tests over psychiatric symptom reporting in the detection of malingering (e.g. Hom & Denny, 2002).

Individually, the LBT total score, the Ganser total score, the CDQ fake and rare symptoms score and the M-FAST test score performed with commensurate discriminatory power. A test of floor effect, measured on the Ganser, self-report of cognitive symptomatology measured on the CDQ and cognitive performance incorporating preattentive processing measured on the LBT discriminated between groups as successfully as a self-report test of general psychiatric symptomatology, the M-FAST (Miller, 2001). These individual results uphold claims of the discriminatory power of the above strategies in detecting malingering (e.g. Rogers, Harrell & Liff, 1993).

2.4.3 Group differences by reaction time

Timing of the Ganser and CDQ did not meaningfully distinguish simulating malingerers from true patients. It is hypothesised that the attentional effort required to read the questions and respond rendered the results unhelpful as a detection strategy. Indeed, some of the true patients took breaks during the testing, or became

distracted, which lengthened the time taken to complete a task. This finding is supported by the literature, which suggests that tasks which require minimal cognitive effort and tap preattentive as opposed to effortful processing will more likely discriminate between malingerers and true patients (e.g. Anscombe, 1987).

2.4.4 Correlations between malingering tests, demographic variables and true pathology

BSI scores correlated positively with scores on the M-FAST. This could be explained by the theory that severely impaired psychiatric patients' self-appraisal on symptom inventories may be influenced by their condition (Rogers, 1997). If this is the case, then it supports the use of a variety of tasks in addition to symptom reporting so as to control for this possibility.

In contrast, true pathology correlated negatively with the combined stanine variable. This is an interesting finding suggesting that the more psychiatrically unwell, the less likely to be classified as malingering according to combined stanine score. This opposes theories that suggest self-report in psychiatric populations renders malingering tools less sensitive (e.g. Rogers, 1997). It is unclear from the literature why this might be the case. It is possible that those classified as malingerers in the patient group only faked their answers on the cognitive malingering battery and not the measures of general psychiatric symptoms (BSI and M-FAST). If this is the case, then it would lend further support to using cognitive malingering tests in addition to measures of general psychiatric symptoms. This could potentially overturn a criticism of using self-report approaches in psychiatric populations and certainly warrants further investigation.

2.4.5 Qualitative investigation of malingering response styles

Simulating malingerers reported using a variety of strategies to fake the tests. The

strategies were employed in response to the demands of the different tests, but were also switched during an individual task. For example, strategies *subtlety* and *imitation* selectively endorsed symptoms which the respondents felt were more likely to be genuine symptoms of mental illness. If self-report general psychiatric symptomatology was the only detection strategy used, then it would be less likely to identify simulating malingerers who used this response style. This is corroborated by speculation that a weakness of floor effect strategies is that they are not sufficient to identify more sophisticated malingerers (Rogers, Harrell & Liff, 1993). This qualitative finding lends further weight to the use of multiple detection strategies to discriminate malingerers.

Certain malingering techniques lent qualitative support to the use of a particular detection strategy. For example, the response style *suppression* described the effortful processing required to suppress an honest response and substitute an opposing fabricated answer. However, the quantitative results suggest that this detection strategy is rendered indiscriminating when there is additional effort required to read the questions. Therefore tasks which require a high proportion of preattentive as opposed to effortful processing will more likely detect malingerers from patients via timing detection strategies, a finding supported by the literature (e.g. Anscombe, 1987; Vendemia et al., 2005).

Simulating malingerers alluded to attempts to convey consistency in their responses, by endorsing similar deficits within and across tests of self-report symptomatology. Respondents indicated the difficulty in remembering what they had endorsed to remain consistent. These participants were more likely to endorse similar symptoms if they had responded positively to them in prior questions. This emphasises the value of using self-report symptom inventories which incorporate questions about fake symptoms and genuine symptoms in a specific domain of functioning. The

qualitative findings suggest that this technique will be strengthened if the questions are about lesser-known deficits in mental illness, such as cognitive functioning. This is supported by Halligan, Bass & Oakley (2003) who suggest using inventories which question functioning in specific domains, such as cognitive symptomatology.

The majority of the simulating malingerers did not select a specific illness to malinger. This is perhaps an indication that the respondents had not prepared a presentation to malinger. A touted shortcoming of simulation designs is the reduced incentive to malinger compared to that of genuine malingerers (e.g. Rogers, 1997). This finding lends further support to the potential difference in presentation of simulating and genuine malingerers.

2.4.6 Strengths, limitations and future research

A strength of this research is the superior discriminatory power of a battery of tests to identify dissimulation. The test battery is underpinned by several theories about different detection strategies which are more likely to distinguish malingerers. In addition, these detection theories are corroborated by qualitative findings which indicate that different malingering styles will more likely be identified by certain detection strategies. However, the findings are only an indication of the strategies that simulating malingerers employ and cannot be considered wholly representative of the genuine malingering population. Although there is qualitative support for timing tasks to detect dissimulation, the quantitative findings did not utilise this strategy successfully.

The comparison true patient group comprised participants who were admitted to inpatient wards in crisis. A proportion of this group were in intensive care units. Considering that problems are anticipated when using self-report symptom inventories where severely impaired psychiatric patients' self-appraisal of

symptomatology may be influenced by their condition (Rogers, 1997), the discriminatory power of the test battery is all the more significant.

This study has given an indication that specific malingering techniques might be more likely identified by certain detection strategies. Future research investigating the relationship between response style and detection strategy could improve the development of test batteries to distinguish dissimulation. This research should be repeated with at risk malingering populations in order to enhance the generalisability of the findings. Timing tasks to identify malingering by virtue of excess cognitive effort needs further investigation using tasks of preattentive processing to minimise extraneous variability produced by effortful processing required on self-report inventories.

This research has practical relevance for clinical and forensic psychologists. With the growing number of cases which require an assessment of malingering, psychologists need to move towards an established gold standard of assessment of dissimulation, achievable only through research. This study has indicated that a screening tool such as the M-FAST should be supported by additional tests of malingering which incorporate a variety of detection strategies. This practical requirement was illustrated by 30% of psychiatric inpatients scoring over the cut-off for malingering on the M-FAST. This emphasises the benefit of exploring the accuracy of cut-offs on a screening tool by using a more comprehensive battery of malingering tests to avoid false positives.

Furthermore, empirically validated measures of malingering are required to protect against the economic and risk-related consequences of false negatives, particularly in criminal proceedings. It would be premature to provide cut-offs on the malingering tests on the strength of this research. From a practical point of view, clinicians need

recommended cut-offs to be able to make a categorical judgement about malingering. Further research would be useful in identifying cut-off values for the malingering battery which would be applicable in different settings. Future research could evaluate the utility of this battery of tests further by comparing to a gold standard instrument such as the Stuctured Interview of Reported Symptoms (SIRS) (Rogers, Bagby & Dickens, 1992), which investigates malingering more stringently than the M-FAST screening tool. Finally, with the expected introduction of a protocol for the admissibility of scientific evidence, research into empirically validated tools with sufficient and explicatory discriminatory power is necessary if tests of malingering are to stand up to potential Daubert-standard challenges in court.

2.5 References

Alban, A., D. (2003). Neuropsychological malingering and cognitive load: Disrupting deception of neuropsychological measures through cognitive overload. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 63(9), pp 4358.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (4th edition).* Washington D.C.: Author.

American Psychiatric Association Division of Clinical Neuropsychology (n.d.). Retrieved May 15, 2007, from http://www.div40.org/

Anscombe, R. (1987). The disorder of consciousness in schizophrenia. *Schizophrenia Bulletin*, 12, pp 241-260.

Bender, S., D. (2002). Feigning cognitive deficits on neuropsychological evaluations: Multiple detection strategies. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 62(7), pp 3372.

Ben-Porath, Y., Graham, J., Hall, G., Hirschman, R. & Zaragoza, M. (Eds.) (1995). *Forensic Applications of the MMPI-2.* London: Sage.

Bruhn, A. R. & Reed, M. R. (1975). Simulation of rain damage on the Bender-Gestalt test by college subjects. *Journal of Personality Assessment*, 39, pp 24-255.

Clark (1988). Sociopathy, malingering and defensiveness. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception* (1st edition, pp 54-64). New York: Guildford Press.

Cohn, M. G. (1995). Epidemiology of malingering studies. *Dissertation Abstracts International Section B: The Sciences and Engineering*, 55(12), pp 5562.

Demarkis, G. J. (1999). Serial malingering on verbal and nonverbal fluency and memory measures: An analog investigation. *Archives of Clinical Neuropsychology,* 14(4), pp 401-410.

Derogatis, L. R. (1975). *Brief Symptom Inventory.* Baltimore: Clinical Psychometric Research.

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

Hom, J. & Denny, R. L. (Eds.) (2002). *Detection of Response Bias in Forensic Neuropsychology.* London: Haworth Press.

Huskey, M. L. (2006). The experience of malingering: Test performance, coaching, personal knowledge, strategies and face validity. *Dissertation Abstracts International Section B: The Sciences and Engineering*, 66(7), pp 3949.

Iverson, G. L. (1995). Qualitative aspects of malingered memory deficits. *Brain Injury*, 9(1), pp35-40.

Kertzman, S., Grinspan, H., Birger, M., Shilapnikov, N., Alish, Y., Nahum, B. Z., Mester, R. & Kotler, M. (2006). Simple real-time computerized tasks for detection of malingering among murderers with schizophrenia. *Israel Journal of Psychiatry and Related Sciences,* 43(2), PP 112-118.

Khan, I., Fayaz, I., Ridgely, J & Wennberg, R. (2000). Factitious clock drawing and constructional apraxia. *Journal of Neurology Neurosurgery and Psychiatry*, 68, pp 106-107.

Lezak, M. D. (1983). *Neuropsychological Assessment (2nd Ed.).* New York: Oxford University Press.

McMennemin, J. (*in preparation). Malingered Cognitive Symptoms in Severe Mental Illness.*

Miller, H. (2001). *Miller Forensic Assessment of Symptoms Test: Professional Manual.* Odessa, FL.: Psychological Assessment Resources.

Millon, T. & Meagher, S. E. (2004). The Millon Clinical Multiaxial Inventory-3 (MCMI-III). In M. J. Hilsenroth & D. L. Segal (Eds.) *Comprehensive handbook of psychological assessment, volume 2: Personality assessment* (pp 109-121). Hoboken, NJ: John Wiley & Sons, Inc.

Mittenberg, W., Patton, C., Vanyock, E. M. & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, pp 1094-102.

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

Pankratz, L. & Binder, L. M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception (2nd Ed.).* New York: Guildford Press.

Pearson Assessments, Brief Symptom Inventory (n.d.). Retrieved May 15, 2007, from http://www.pearsonassessments.com/tests/bsi.htm

POST: Parliamentary Office of Science and Technology (2005). *Postnote: Science in court.* Retrieved January 11, 2007, from http://www.parliament.uk/documents/upload/postpn248.pdf

Postlethwaite, K.C. and Jaspars, J. (1986). The experimental use of personal constructs in educational research: The critical triad procedure. *British Journal of Educational Psychology*, 56(3), pp241-254.

Reynolds, C. R. (1998). *Detection of malingering during head injury litigation.* New York: Plenum Publishing Corporation.

Rogers, R. (1986). *Conducting insanity evaluations.* New York: Van Nostrand Reinhold.

Rogers, R. (1990). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry and Law,* 18, pp 323-33.

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

Rogers, R., Bagby, R. M. & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms (SIRS) professional manual.* Odessa, FL: Psychological Assessment Resources.

Rogers, R. & Cruise, K. R. (2000). Malingering and deception among psychopaths. In C Gacono (Ed.) *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp 269-284). New York: Lawrence Erlbaum.

Rogers, R., Harrell, E. H. & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review,* 13, pp 255-274.

Schenkenberg, T., Bradford, D. C. & Ajax, E. T. (1980). Line bisection and unilateral neglect in patients with neurological impairment, *Neurology,* 30, pp 509.

Schretlen, D. (1988). The use of psychological tests to identify malingered symptoms of mental disorder. *Clinical Psychology Review,* 8, pp 451-476.

Shulman, M. B. Alexander, M. P., McGlinchey-Berroth, R. & Milberg, W. (2002). Triangular backgrounds shift the bias of line bisection performance in hemispatial neglect. *Journal of Neurology Neurosurgery and Psychiatry*, 72, pp 68-72.

Steffan, J. S., Kroner, D. G. & Morgan, R. D. (2007). Effect of symptom information and Intelligence in dissimulation: An examination of faking response styles by inmates on the basic personality inventory. *Assessment,* 14(1), pp 22-34.

Tan, J. E., Slick, D. J., Strauss, E. & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *Clinical Neuropsychologists*, 16(4), pp 495-505.

Vendemia, J. M. C., Buzan, R. F. & Simon-Dack, S. L. (2005). Reaction time of motor responses in two-stimulus paradigms involving deception and congruity with varying levels of difficulty. *Behavioural Neurology*, 16(1), pp 25-36.

Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J. & Orey, T. A. (2001). Detection of inadequate effort on neuropsychological testing: a meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, 16(1), pp 45-73.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale - Third Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997a). Wechsler Memory Scale - Third Edition. San Antonio, TX: The Psychological Corporation

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

Wiggins, E. C. & Brandt, J. (1988). The detection of simulated amnesia. *Law and Human Behavior,* 12(1), pp 57-78.

Willison, J. & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neurpsychology*, 21(1), pp 41-52.

Ziskin, J. (1984). Malingering of psychological disorders. *Behavioural Sciences and the Law,* 2, pp 39-50.

Part 3: Critical Appraisal

This paper reflects on the research undertaken and reported in this thesis volume. It covers themes including the generalisability of the findings when using a simulating malingering design, the utility of measuring reaction time to detect malingering, difficulties in the recruitment of inpatients, the array of choices in selecting the test battery and the clinical applications of the research.

3.1 Generalisability in simulation designs:

Although the majority of studies utilise a simulation design and this is recognised as the standard method for studying malingering (e.g. Rogers, 1997), the major limitation of this approach is the generalisability of the findings. The simulating group in the current study were offered a monetary incentive to feign illness. However, this motivation is not considered comparable to the significant incentive genuine malingerers may experience in order to receive compensation or influence outcome in a criminal trial. During the testing phase, some of the simulating malingering group gave the impression that, whilst they may have enjoyed the process, they were not seen to demonstrate seriousness in presenting a persuasive profile of illness. This observation is consistent with the view that simulating malingerers are far from comparable to true life malingers in their motivation to feign.

A related drawback became apparent when testing the simulating group. In recruiting participants, it was necessary in some cases to provide advance warning, if, for example, the researcher was booking in time to see people at their workplace. As a result, some participants may have known the purpose of the study a month in advance and others may have not known until immediately before testing. This created differences in the opportunity to prepare a response style. One participant mentioned researching a condition to improve the profile of responses. This would

not have been possible to do for some participants. It is considered that this is an additional weakness in the design and that the simulating malingering group should all be provided with advance warning of the study aims so as to allow the opportunity to research mental illness, if they so choose.

Within those simulators who had the opportunity to research mental illness, there appeared to be differences in the amount of preparation that participants engaged in. This was not asked explicitly and on reflection may be a useful point of inquiry for qualitative research. However, it would support the notion that there is heterogeneity in response styles of malingerers, who may vary significantly in the approach they take prior to the testing process. As Pankratz & Binder (1997) contended, there may be differences in the nature of malingering according to the level of understanding of the condition one is attempting to feign.

Generalisability remains a significant problem for malingering research and tests of malingering should be replicated using an at risk malingering group. Nevertheless, a comparative strength of this research design was the inclusion of a clinical comparison group. A substantial amount of research using a simulating design only use a simulating group and a healthy control group (Rogers,1997). These designs do not explore the possibility that the differences in simulating malingerers would not also be found in patient groups. The interpretation of the findings from this study are therefore strengthened by the inclusion of a clinical comparison.

3.2 The utility of reaction time as a discriminating variable

Reaction time is suggested to be a effective method for detecting malingerers (e.g. Vendemia et al., 2005). The method is made more attractive by virtue of the difficulty in faking reaction time. If malingering requires a significant amount of cognitive effort (Alban, 2003), then tasks which require little effortful processing may expose

malingerers by the added time it takes to initiate a malingered response. This idea has been of interest since Wertheimer & Klein's (1904) experiments and yet has rarely been demonstrated successfully (e.g. Vendemia et al., 2005). This is possibly because there are extraneous factors which create additional variability in this fine measurement.

During the testing, the most noticeable difficulties emerged when administering the timed tasks to the inpatient group. Extraneous factors which rendered reaction time ineffective on the test battery included the length of time the test takes to complete. The longer a measure is, the harder it is for a patient to maintain concentration and the more likely distractions will occur. This was noted on several occasions when patients broke off from the task to engage in conversation, stop for a break or leave the room for unspecified reasons. Timing tasks which require significant cognitive effort may also render this detection strategy insensitive. Having to read questions requires effort and also creates variance according to reading speed. Longer and more complex questions created more variance. Patients took significantly longer on all tests which measured reaction time.

The Line Bisection Test (LBT) had the most potential to discriminate malingerers by reaction time. Due to programming difficulties, the task was not timed in this study. However, during the testing process, a qualitative observation indicated that most inpatients maintained concentration during the LBT and completed it swiftly. This maybe due to the little cognitive effort required to answer each item. Indeed, it has been suggested that the LBT is predominantly a test of preattentive processing, with the demands of each item processed automatically and the only effortful requirement being to find the centre of the line (Shulman et al., 2002). The LBT was also the first computerised task in the test battery and so levels of concentration may have been higher at this point.

In summary, the research has corroborated earlier findings that reaction time requires fine measurement. Processes which may improve the discriminatory power of reaction time include administering reaction time tasks at the beginning of a test battery, using tasks which have a low number of items and, most importantly, using tasks which require minimal effortful processing.

3.3 Patient group

The clinical group were recruited from inpatient acute psychiatric units in order to be sure of approaching potential participants who had a diagnosis of severe mental illness. The principle reason for using a clinical comparison group is to ensure that differences in the simulating malingering group are not replicated by patient with genuine mental health problems (Rogers, 1997). It has also been suggested that patients with severe mental illness may not report their symptoms accurately (Rogers, 1997). It is therefore important to include patients with severe symptomatology to ensure that the test battery distinguishes simulating malingerers from patients with erratic reporting styles.

On some wards, sampling selectively recruited patients who were less severely ill. Ward staff warned the researcher about approaching patients who were particularly unwell. Of those approached, the less severely ill patients appeared more likely to consent to the testing than the more unwell. There may be a combination of reasons for this. Less well patients may not have felt able to commit to an hour of testing, they may have not wanted to engage with a researcher who they viewed as a representative of the NHS and they may have been suspicious of the aims of the research. The latter point should be considered in the wording of the information sheet.

Some patients with psychosis questioned the integrity of the research and voiced

suspicion that the research was evaluating the veracity of their symptoms. Reassurances provided by the researcher were not always sufficient to allay their fears. It is possible that mention of malingering or the genuineness of symptom reporting activated paranoia in some patients. Although there were no visible repercussions as a result of approaching patients during the course of this research, it remains possible that the nature of the research could unsettle patients with paranoia. Researchers have an ethical obligation to remain explicit about the research aims and can continue to stress their voluntary participation and right to withdraw at any point, should patients become disturbed before, during or after the testing.

As a result, the inpatient group is at risk of consisting of the less severe patients on wards. However, a strength of this research was that recruitment spanned several units (totalling more than 14 wards). Included in these were patients recruited from two intensive care units. Patients on these wards were often highly aroused and in crisis. The researcher approached patients with trepidation, and although the ratio of positive responses to participate were lower, several patients on these wards took part in the research. Participants from the most severe inpatient units were therefore included in the patient group. On balance, the degree of illness severity is considered adequate to be representative of inpatients in the psychiatric system. Researchers should be aware of the potential to be directed away from the more severely ill patients and can compensate for this by visiting high dependency wards and questioning ward staff about the patients they can approach.

3.4 Test battery selection

A strength of the research lay in the range of detection strategies included in the test battery. Rogers, Harrell & Liff (1993) emphasised the importance of incorporating a variety of malingering measures to heighten discriminatory power. All detection

strategies, excepting reaction time, successfully distinguished between groups. However, it was the combined test results that differentiated most powerfully between groups. Ideas for improving reaction time detection are provided above.

This research did not use a performance curve technique in the test battery. This could possibly improve detection further as performance curve measures receive positive reviews of their discriminatory power (Rogers, Harrell & Liff, 1993). However, there is a trade off in selecting a battery of tests between the time taken to administer them and the discriminatory power. The length of the test battery will have an impact on the uptake of this method in clinical practice. Whilst the test battery outperformed an established screening tool in distinguishing simulating malingerers, the administration time was on average approximately 45 minutes, compared to 5 minutes for the screening tool. Clinicians will be mindful of this when selecting malingering measures in practice. However, in thorough investigations of malingering, for example, for court reports, clinicians may well prefer to use a more comprehensive battery such as the one used in this research, to lend more credence to the results and interpretations.

3.5 Clinical applications

Research such as this has good clinical utility as there is a vast array of tests available to choose from. This provides Clinical Psychologists with a clear indication of a useful battery of tests which have been demonstrated to distinguish malingering with superior discriminatory power compared to a shorter screening measure. Furthermore, Clinical Psychologists may not ordinarily be aware of the utility of including a cognitive symptom inventory to detect malingering in cases where cognitive symptomatology is not the primary focus (for example, Posttraumatic Stress Disorder). This research provides strong support for the inclusion of a specialised cognitive symptom inventory to improve the discriminatory power of a malingering

test battery.

An aim of this research was to evaluate the utility of a new cognitive symptom inventory, the Cognitive Dysfunctions Questionnaire (CDQ). The CDQ was found to have good discriminatory power independently and be more effective when used in conjunction with other detection strategies. The CDQ could potentially be shortened without losing sensitivity or specificity. The 'not sure' responses did not add to the power of the test battery when the combination stanine scores were computed. In addition, the 'frequency' and 'distress' ratings add a significant amount of time to the administration time without significantly improving to the discriminatory power. These three variables could be removed, shortening the test, with little alteration to the test's effectiveness.

Finally, perhaps the ultimate test of a malingering battery's utility is the potential to stand up to Daubert challenges in court. Clinical Psychologists require tests of malingering that have a strong evidence base to warrant their use in court. In order to do this, a test battery needs to be evaluated in research, it needs to be peer reviewed and published, it has to meet an acceptable threshold for sensitivity and specificity and is required to receive general acceptance in the scientific community (Hom & Denny, 2002). This research goes some way to establishing a new test battery that, following publication, could pass Daubert challenges in court. With a new protocol for the admissibility of scientific evidence in UK courts pending (Select Committee for Science and Technology, 2005), research such as this will give confidence to Clinical Psychologists in choosing a battery of tests to assess malingering.

3.6 References

Alban, A., D. (2003). Neuropsychological malingering and cognitive load: Disrupting deception of neuropsychological measures through cognitive overload. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 63(9), pp 4358.

Hom, J. & Denny, R. L. (Eds.) (2002). *Detection of Response Bias in Forensic Neuropsychology.* London: Haworth Press.

Pankratz, L. & Binder, L. M. (1997). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception (2nd Ed.).* New York: Guildford Press.

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

Rogers, R., Harrell, E. H. & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, 13, pp 255-274.

Science and Technology Committee (2005). *Select Committee on Science and Technology Seventh Report.* Retrieved on August 8 2006, from, http://www.publications.parliament.uk/pa/cm200405/cmselect/cmsctech/96/9610.htm

Shulman, M. B. Alexander, M. P., McGlinchey-Berroth, R. & Milberg, W. (2002). Triangular backgrounds shift the bias of line bisection performance in hemispatial neglect. *Journal of Neurology Neurosurgery and Psychiatry*, 72, pp 68-72.

Vendemia, J. M. C., Buzan, R. F. & Simon-Dack, S. L. (2005). Reaction time of motor responses in two-stimulus paradigms involving deception and congruity with varying levels of difficulty. *Behavioural Neurology*, 16(1), pp 25-36.

Wertheimer, M. & Klein, J. (1904). Psychologische Tatbestandsdiagnostik. *Archiv fur Kriminal-Anthropologie und Kriminalistik,* 15, pp72-113.Cited in E. B. Newman (1944) Max Wetheimer: 1880-1943, *The American Journal of Psychology*, 57(3), pp. 428-435.

Part 4: Appendices

4.1 Appendix: Ethical approval form

NHS

Ealing & West London Mental Health Trust Research Ethics Committee

06 November 2006

Mr Philip Minoudis
Trainee Clinical Psychologist
Sub Dept. of Clinical and Health Psychology
University College London

Dear Mr Minoudis,

Full title of study: **Malingering of Cognitive Symptoms**
REC reference number: **06/Q0410/48**

Thank you for your letter of 2 November 2006, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The favourable opinion applies to the research sites listed on the attached form.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application		14.08.06
Application		14.08.06
Application		14.08.06
Investigator CV		15.08.06
Investigator CV		15.08.06

Protocol	1	15.08.06
Covering Letter		
Covering Letter		
Covering Letter		16.08.06
Covering Letter		16.08.06
Peer Review	UCL	
Peer Review	UCL	
Participant Information Sheet	1	
Participant Information Sheet	1	
Participant Information Sheet: Participants	3	27.10.06
Participant Information Sheet: Participants	3	27.10.06
Participant Information Sheet: Patients	1	2710.06
Participant Information Sheet: Patients	1	2710.06
Participant Information Sheet	1	15.08.06
Participant Information Sheet	1	15.08.06
Participant Consent Form	1	15.08.06
Participant Consent Form	1	15.08.06
Participant Consent Form: Participant	3	27.10.06
Participant Consent Form: Participant	3	27.10.06
Participant Consent Form: Patient	1	27.10.06
Participant Consent Form: Patient	1	27.10.06
Participant Consent Form	1	
Participant Consent Form	1	
Response to Request for Further Information		
Response to Request for Further Information		
Supervisor		15.08.06
Supervisor		15.08.06

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

06/Q0410/48	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely,

Chair

Email:

4.2 Appendix: Information sheets

Information sheet for participants (Date: 27/10/2006, Version 3)

(Ealing & West London Mental Health Trust Research Ethics Committee registration number: 06/Q0410/48)

You are being asked to participate in a research project. The following information sheet explains what the research is about, why it is being carried out and what will be asked of you if you agree to take part. *If you do not wish to take part then you do not have to.* If you decide not to take part, then this will not affect your rights in any way. If you have any questions about the research, then please feel free to ask.

This sheet is your copy to keep.

Title of Research
Assessment of Cognitive Symptoms

Aim of Research
The research compares people with genuine mental health problems to people who are asked to pretend they have a mental illness. This helps researchers to develop questionnaires which can tell between genuine mental health problems and faked ones.

What would you have to do?
If you decide to take part then you will be asked a series of questions about symptoms of mental health problems. You will also be asked to complete some tasks on a computer. We would like you to answer these questions as honestly as possible and to complete the tasks to the best of your ability. All together, this should take approximately 1 hour. We have no reason to believe that you will experience any difficulties as a result of taking part.

This research does *not* involve any physical examinations or medications. Whatever you decide will not affect the care you receive in any way.

Taking part in this research is entirely voluntary. If you decide you do not want to take part, you can withdraw from the research at any time.

**This research has been reviewed and approved by Ealing & West London Mental Health Trust Research Ethics Committee.*

Information sheet for patients (Date: 27/10/2006, Version 1)

(Ealing & West London Mental Health Trust Research Ethics Committee registration number: 06/Q0410/48)

You are being asked to participate in a research project. The following information sheet explains what the research is about, why it is being carried out and what will be asked of you if you agree to take part. *If you do not wish to take part then you do not have to.* If you decide not to take part and you are a patient, then this will not affect your current or future treatment in any way. If you have any questions about the research, then please feel free to ask.

This sheet is your copy to keep.

Title of Research
Assessment of Cognitive Symptoms

Aim of Research
The research compares people with genuine mental health problems to people who are asked to pretend they have a mental illness. This helps researchers to develop questionnaires which can tell between genuine mental health problems and faked ones.

What would you have to do?
If you decide to take part then you will be asked a series of questions about symptoms of mental health problems. You will also be asked to complete some tasks on a computer. We would like you to answer these questions as honestly as possible and to complete the tasks to the best of your ability. All together, this should take approximately 1 hour. We have no reason to believe that you will experience any difficulties as a result of taking part.

This research does *not* involve any physical examinations or medications. Whatever you decide will not affect the care you receive in any way. Nothing will be recorded in your notes, whether you choose to participate or not.

Taking part in this research is entirely voluntary. If you decide you do not want to take part, you can withdraw from the research at any time.

**This research has been reviewed and approved by Ealing & West London Mental Health Trust Research Ethics Committee.*

4.3 Appendix: Consent forms

Participant Consent Form (Date: 27/10/2006, Version 3)

(Ealing & West London Mental Health Trust Research Ethics Committee registration number: 06/Q0410/48)

Title of Research
Assessment of Cognitive Symptoms

Name of Researcher
Philip Minoudis

I confirm that I have read and understood the information sheet dated 27th October 2006 about the above research and have been given the opportunity to ask questions.

I understand that taking part in this research is voluntary and I am free to withdraw at any time, without having to give a reason. Withdrawing from the research will not affect my rights in any way.

I agree to take part in this study.

I understand that if I tell the researcher anything that suggests a risk of harm to myself or to others, the researcher will inform appropriate services.

Name of Participant	Date	Signature
.....	.....	.....
Name of Researcher	Date	Signature
.....	.....	.....

REMEMBER THAT YOU MAY WITHDRAW FROM THIS RESEARCH AT ANY TIME WITHOUT ANY CONSEQUENCE

Patient Consent Form (Date: 27/10/2006, Version 1)

(Ealing & West London Mental Health Trust Research Ethics Committee registration number: 06/Q0410/48)

Title of Research
Assessment of Cognitive Symptoms

Name of Researcher
Philip Minoudis

I confirm that I have read and understood the information sheet dated 27th October 2006 about the above research and have been given the opportunity to ask questions.

I understand that taking part in this research is voluntary and I am free to withdraw at any time, without having to give a reason. Withdrawing from the research will not affect my treatment or rights in any way.

I understand that sections of my medical notes may be looked at by the researcher where it is relevant to taking part in the research. I give permission to the researcher to have access to my medical notes.

I agree to take part in this study.

I understand that if I tell the researcher anything that suggests a risk of harm to myself or to others or an intention to leave the hospital without permission, the researcher will inform the nursing staff.

Name of Participant	Date	Signature
.....	.....	.....
Name of Researcher	Date	Signature
.....	.....	.....

REMEMBER THAT YOU MAY WITHDRAW FROM THIS RESEARCH AT ANY TIME WITHOUT ANY CONSEQUENCE

4.4 Appendix: Measures

1) Weschler Test of Adult Reading (WTAR) (Wechsler, 2001).

WTAR Word List - UK pronunciation guide

Say, **I will show you some words that I will ask you to pronounce.** Place the WTAR Word Card in front of the examinee. As you point to the card, say, **Beginning with the first word on the list, pronounce each word aloud. Start with this word** (point to item 1), **and go down this column, one after the other, without skipping any. When you finish this column, go to the next column** (point to the second column). **Pronounce each word even if you are unsure. Do you understand?** When you are sure that the examinee understands the task, say, **Ready? Begin.**

	Item	Pronunciation	Score (0, 1)		Item	Pronunciation	Score (0, 1)
1.	again	ah-GEHN ah-GAIN *or* uh-GEHN or uh-GAIN		26.	conscientious	con-shee-EN-shss	
2.	address	ah-DRESS *or* uh-DRESS		27.	homily	HOM-ih-lay *or* HOM-ih-lee	
3.	cough	kawf *or* kof		28.	malady	MAL-uh-day *or* MAL-uh-dee	
4.	preview	PREE-vyue		29.	subtle	SUH-tl	
5.	although	awl-THO		30.	fecund	FE-cund *or* FEE-cund	
6.	most	mohst		31.	palatable	PAL-ah-tuh-bul *or* PAL-uh-tuh-bul	
7.	excitement	eck-SITE-munt *or* ik-SITE-munt		32.	menagerie	meh-NA-juh-ree	
8.	know	noh *or* no		33.	obfuscate	OB-fuh-skate	
9.	plumb	plum		34.	liaison	lee-AY-zon *or* lee-AY-zn	
10	decorate	DEK-oh-rate *or* DEK-uh-rate		35.	exigency	eks-IH-jen-say *or* eks-IH-jen-see	
11.	fierce	fee-us *or* feerss		36.	xenophobia	zen-oh-FO-bee-uh	
12.	knead	need		37.	ogre	OH-gur	
13.	aisle	iyle		38.	scurrilous	SKUR-ih-lus *or* SKUR-uh-lus	
14.	vengeance	VEN-jnss		39.	ethereal	ih-THEE-ree-ul *or* ih-THEER-ee-ul	
15.	prestigious	pre-STIJ-us *or* pre-STEEJ-us		40.	paradigm	PAH-rah-dime	
16.	wreathe	reeTH		41.	perspicuity	per-spuh-KYEW-uh-tee	
17.	gnat	nat		42.	plethora	PLETH-oh-rah *or* PLETH-eh-rah	
18.	amphitheatre	AM-fih-thee-uh-ter		43.	lugubrious	loo-GOOB-ree-uss *or* loo-GOO-bree-uss	
19.	lieu	loo *or* l(y)oo		44.	treatise	TREE-tiz *or* TREET-iz	
20.	grotesque	gro-TESK		45.	dilettante	DILL-ih-tan-tay *or* DILL-uh-tahnt	
21.	iridescent	ihr-ih-DESS-unt *or* ihr-uh-DESS-unt		46.	vertiginous	ver-TIDJ-in-iss	
22.	ballet	BA-lay *or* ba-LAY *or* bal-ay		47.	ubiquitous	you-BIC-wuh-tiss *or* you-BIC-wuh-tus	
23.	equestrian	eh-KWESS-tree-un *or* ih- KWESS-tree-un		48.	hyperbole	hy-PER-bul-lay *or* hy-PUR-bul-lay	
24.	porpoise	PAW-pss *or* POR-poyz (Scots)		49.	insouciant	in-SOO-see-yunt	
25.	aesthetic	ess-THET-ik *or* ees-THET-ik		50.	hegemony	heh-GEM-o-nee *or* heh-JEM-o-nee *or* HEH-geh-mon-ee	

WTAR Raw Score	
WTAR Standard Score	

2) Brief Symptom Inventory (BSI) (Derogatis, 1975).

A B C D

3) Line Bisection Task (LBT) (Schenkenberg et al., 1980).

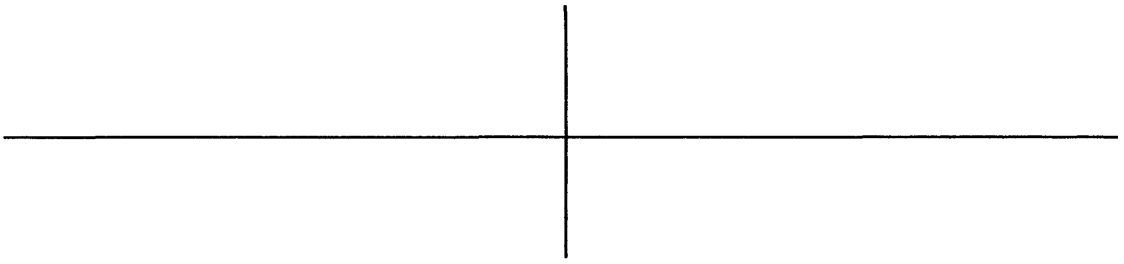
This task was presented on a laptop. Participants were read the following instructions aloud while presented with an example item.

Instructions:

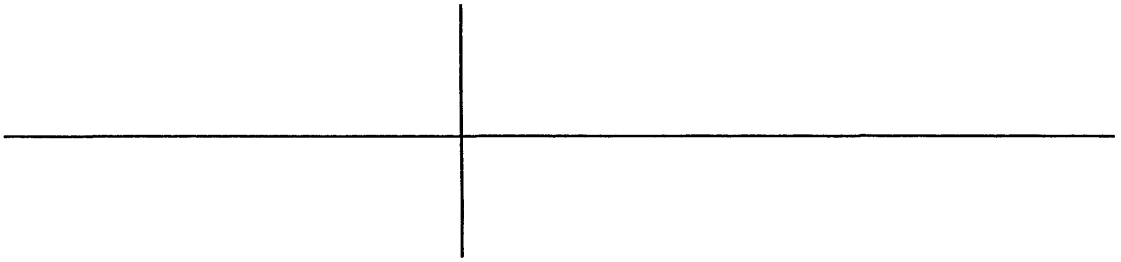
There are two horizontal lines on the screen. Each line has another line that goes through it. Your task is to decide which of these lines that go across the screen has a line that goes through the middle of it. You should press either the key with the 'L' sticker on it if it is the left line or the key with the 'R' sticker on it if you think it is the right hand line.

Example of one of the items:

1 A



1 B



4) Ganser questionnaire (created from previous research).

This questionnaire was presented on a laptop. Only numerical responses were required. Each question was presented individually. Response and reaction time were recorded.

Instructions

Please answer the following questions:

1. How many days are there in a week?
2. How many weeks are there in a year?
3. How many sides does a square have?
4. What is 1 plus 2?
5. How many legs does a dog have?
6. What is 5 – 1?
7. How many hours are there in a day?
8. How many seconds are there in a minute?
9. How many things are there in a dozen?
10. How many legs do people have?
11. How many wheels does a bicycle have?
12. How many months are there in a year?
13. How many minutes are there in an hour?
14. How many things are there in a pair?

5) Cognitive Dysfunctions Questionnaire (CDQ) (Coxell).

This questionnaire was presented on a laptop. Responses and reaction times were recorded.

Participants were presented with the following instructions on screen:

People can report experiencing a wide range of mental and sensory problems. I am going to ask you if you have had any of the following problems in the last three months. Please press the key marked 'yes' if you have had any of these problems, and the key marked 'no' if you have not had any of these problems. If you are not sure if you have had these problems please press the key with the '?'.

Each question was presented on screen individually and prefixed with:

In the last three months have you experienced...

1) Your sense of taste being much stronger than usual
2) Feeling flooded by tastes
3) Not being able to taste one thing because of being flooded by different tastes
4) Often getting a vinegar-like taste in your mouth for no obvious reason
5) Not being able to taste things on one side of your tongue.
6) Your sense of smell being much more powerful than usual
7) Feeling flooded by smells
8) Not being able to smell one thing because of being flooded by different smells
9) Often being able to smell something like burning rubber for no reason
10) Getting the taste or smell of something just by touching it (for example touching an apple and tasting it in your mouth before you eat it)
11) Reaching for something but your hand missing it by more than a couple of inches
12) A muscle or muscles jumping or twitching
13) Loss of the ability to write down words on paper
14) Your handwriting having changed a great deal
15) The impression that you have lost the ability to control your left hand at times and it seeming to have a "mind of it's own".
16) Finding it very difficult to judge how heavy things are when you pick them up with your hand
17) Finding it very hard to tell if things are hot or cold when you are touching them.
18) Finding it very hard to tell if things are rough or smooth when you are touching them.
19) Finding it very hard to tell how big things are when you are touching them
20) Finding it very hard to tell what shape things are when you are touching them.
21) Loss of the ability to feel sensations down one side of your body.
22) The feeling that a *part* of your body has got much larger.
23) Loss of the ability to feel things on your face
24) Having great trouble keeping your balance when walking
25) Your hands trembling when you start writing
26) Not being able to control your arms or hands like you used to.
27) Not being able to do things like tying up shoelaces or doing up buttons

29) Things that you are holding seeming much bigger or smaller than they usually do.
30) The feeling that a part of your body has got much smaller
31) Not being able to do simple things like drawing simple shapes
32) Sounds being louder or more intense
33) Feeling flooded by sound
34) Not being able to concentrate on one particular sound because of feeling flooded by other sounds
35) Not being able to say words that you used to be able to say
36) Everyday sounds sounding somehow different to before
37) Repeating things that other people say even though you do not mean to do so
38) Repeating yourself a lot
39) Buzzing followed by ringing in the ears
40) Hearing the things that people say to you repeated over and over in your head.
41) Not being able to repeat something that somebody has said.
42) Starting to say something and then forgetting what you meant to say
43) Thinking that people talking in your language are talking in a different language
44) Not being able to remember the names of things
45) Jumbling your words when you speak
46) Your vision being much more powerful than usual
47) Feeling flooded by visual images
48) Not being able to concentrate on one visual image because of feeling flooded by lots of visual images
49) Things just not seeming to have any colour when you look at them.
50) Not being able to recognise things properly when you are looking at them.
51) Double vision
52) Things seeming to change colour when you look at them for more than a couple of seconds
53) Not being able to tell the difference between colours like you could before.
54) Things that you look at having a greenish colour to them
55) Not being able to recognise the faces of people that you know.
56) Things looking much smaller than they used to.
57) Things seeming to look much closer than they did before
58) Things looking much bigger than they used to.
59) Visual things being repeated. For example, a person will walk past you, and then a few minutes later you will get the impression of seeing them walk past you again.
60) Things that you look at having a reddish colour to them
61) Not being sure if you are right or left handed
62) Not understanding how to do even simple maths like adding up or taking away
63) Not being able to remember what you had for breakfast
64) Not being able to remember the first name of one of your parents
65) Not being able to remember your date of birth
66) Not being able to remember what happened yesterday
67) Not being able to remember anything for a reasonably long period of your life (for example not remembering anything from secondary school)
69) Getting confused between left and right
70) Loss of the ability to tell the time when you look at a clock

At the end of this section, participants were asked to rate the frequency and distress of each item that they had endorsed.

In the last few months how often have you experienced <present item>?

1) Only once
2) A few times
3) About half of the time
4) More than half of the time
5) Nearly all the time
6) All of the time

How much distress has <present item> caused you?

1) None
2) Mild
3) Moderate
4) Severe
5) Very severe

6) Miller Forensic Assessment of Symptoms Test (M-FAST) (Miller, 2001).

PAR Psychological Assessment Resources, Inc. • 16204 N. Florida Avenue • Lutz, FL 33549 • 1.800.331.8378 • www.parinc.com

9 8 7 6 5 Reorder #RO-4649 Printed in the U.S.A.