The Test of Memory Malinger:
Simulation Studies and Clinical Validation

by

Laura M. Rees, M.A.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of

Doctor of Philosophy

Department of Psychology

Carleton University
Ottawa, Ontario
October 1996

copyright
1996, Laura M. Rees

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395, Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file / Votre référence

Our file / Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-19408-6

Name John 2445

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Psychology

0022 U.M.I.

SUBJECT TERM

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0720
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0488
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
General 0289
Ancient 0290
Linguistics 0291
Modern 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0352
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion 0318
General 0321
Biblical Studies 0319
Clergy 0320
History of 0322
Philosophy of 0469
Theology 0323

SOCIAL SCIENCES

American Studies 0324
Anthropology 0326
Cultural 0327
Physical 0310
Business Administration 0272
Accounting 0770
Banking 0454
Management 0338
Marketing 0385
Canadian Studies 0501
Economics 0503
General 0505
Agricultural 0508
Commerce-Business 0509
Finance 0510
History 0511
Labor 0358
Theory 0366
Folklore 0351
Geography 0578
Gerontology 0383
History 0386
General 0386

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0337
United States 0585
History of Science 0398
Law 0615
Political Science 0616
General 0617
International Law and Relations 0814
Public Administration 0452
Recreation 0626
Social Work 0627
Sociology 0938
Criminology and Penology 0631
Demography 0628
Ethnic and Racial Studies 0629
Individual and Family 0630
Studies 0700
Industrial and Labor 0344
Relations 0709
Public and Social Welfare 0999
Social Structure and Development 0453
Theory and Methods 0537
Transportation 0538
Urban and Regional Planning 0539
Women's Studies 0540

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture 0473
General 0285
Agronomy 0475
Animal Culture and Nutrition 0476
Animal Pathology 0359
Food Science and Technology 0478
Forestry and Wildlife 0479
Plant Culture 0480
Plant Pathology 0817
Plant Physiology 0777
Range Management 0746
Wood Technology 0306
Biology 0287
General 0308
Anatomy 0309
Botanics 0379
Cell 0329
Ecology 0353
Entomology 0369
Genetics 0793
Limnology 0410
Microbiology 0307
Molecular 0317
Neuroscience 0416
Oceanography 0433
Physiology 0821
Radiation 0778
Veterinary Science 0472
Zoology 0786
Biophysics 0760
General 0425
Medical 0996
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Nutrition 0455
Paleobotany 0426
Paleontology 0418
Paleozoology 0985
Petrology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences 0566
General 0300
Audiology 0992
Chemotherapy 0567
Dentistry 0350
Education 0769
Hospital Management 0758
Human Development 0982
Immunology 0564
Medicine and Surgery 0347
Mental Health 0569
Nursing 0570
Nutrition 0380
Obstetrics and Gynecology 0354
Occupational Health and Therapy 0381
Ophthalmology 0571
Pathology 0419
Pharmacology 0572
Pharmacy 0382
Physical Therapy 0573
Public Health 0574
Radiology 0575
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences 0485
Chemistry 0749
General 0486
Agricultural 0487
Analytical 0488
Biochemistry 0738
Inorganic 0490
Nuclear 0491
Organic 0494
Physical 0495
Pharmaceutical 0754
Polymer 0405
Mathematics 0605
Physics 0986
General 0606
Acoustics 0608
Astronomy and Astrophysics 0748
Atomic 0607
Electronics and Electricity 0798
Elementary Particles and High Energy 0759
Fluid and Plasma 0609
Molecular 0610
Nuclear 0752
Optics 0756
Radiation 0611
Solid State 0463
Statistics 0346
Applied Sciences 0984
Computer Science 0346

Engineering 0537
General 0538
Aerospace 0539
Agricultural 0540
Automotive 0541
Biomedical 0542
Chemical 0543
Civil 0544
Electronics and Electrical 0348
Heat and Thermodynamics 0545
Hydraulic 0546
Industrial 0547
Marine 0794
Materials Science 0548
Mechanical 0743
Metallurgy 0551
Mining 0552
Nuclear 0549
Packaging 0765
Petroleum 0554
Sanitary and Municipal 0790
System Science 0428
Crestotechnology 0796
Quantum Research 0795
Plastics Technology 0994
Textile Technology 0621
General 0384
Behavioral 0622
Clinical 0620
Developmental 0623
Experimental 0624
Industrial 0625
Personality 0989
Physiological 0349
Psychobiology 0632
Psychometrics 0451
Social 0451

The undersigned hereby recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis.

The Test of Memory Malingering:
Simulation Studies and Clinical Validation

submitted by

Laura M. Rees, B.A., M.A.

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chair, Department of Psychology

Thesis Supervisor

External Examiner

Carleton University
October 1996

Abstract

The Test of Memory Malingering (TOMM) is a newly developed, visual recognition test with good discriminant and face validity. The TOMM was administered as part of a battery of tests and was evaluated within four separate, yet integrated studies. Two studies used a simulation paradigm, one with university students (Study 1) and the other with traumatically brain injured patients (Study 2). Study 3 employed various patient and control groups, including TBI patients undergoing litigation. The final study (Study 4) evaluated performance on the TOMM across wide ranging and diverse clinical groups. Converging evidence from all studies indicates that the TOMM is a compelling measure of malingering/motivation. Results revealed that it met all criteria of a clinically useful measure of malingering. These include that (1) it differentiates between malingerers and non-malingerers, (2) the perceived difficulty outweighs the actual difficulty of the test, (3) it is not sensitive to non-motivational factors such as, cognitive ability, age, and education, and (4) it appears to be a legitimate measure of memory (i.e., has face validity). Analysis of debriefing questions revealed that individuals did not perceive the TOMM as an obvious measure of malingering.

Acknowledgements

There are many individuals who contributed to the successful completion of my doctoral degree. First, I would like to thank those individuals who participated in the studies, without whose interest the research project would not have been as successful as it was. Second, I would like to extend my deepest appreciation to my advisor, Tom Tombaugh, who encouraged me to think independently, who treated me as a colleague, and who appreciated my sense of humour. Last but not least, are my friends and family: Donna Chernecki who supported me throughout graduate studies and always made sure that I balanced an academic life with social activities; Tim Bennett, my best friend and confidant, who encouraged me to continue when I felt like giving up, who spent many hours discussing my research, proof reading manuscripts, and making me laugh; my parents, John and Donna, my brothers, Gord and Murray, and their families, all of who have made me realize my potential and encouraged me to succeed.

Dedication

This dissertation is dedicated to John and Donna Rees, whose love, support, and encouragement have made me who I am today. Thank you.

Table of Contents

Abstract	iii
Acknowledgements	iv
Dedication	v
Table of Contents	vi
List of Tables	vii
List of Figures	viii
Introduction	1
What is malingering?	2
Incidence of malingering	4
Terminology of malingering	6
How is malingering measured?	9
Neuropsychological invalidity indicators	10
Inconsistency between observed and expected behaviour	10
Inconsistency between reported behaviour and test performance	12
Inconsistency within and between evaluations	13
Specific Tests of malingering	14
Dot Counting Test	15
Rey Fifteen Item Test	18
Symptom Validity Testing	25
Test of Memory Malingering	37
Problems in malingering research	39
Simulation studies	40
Instructional set	44
Debriefing	46
Single measures	48
Incentive	49
Preparation time	50
Research Studies	51
Study 1: University Simulation Study	53
Method	55
Results and Discussion	60
Study 2: TBI Simulation Study	80
Method	81
Results and Discussion	85
Study 3: Litigation Study	99
Method	99
Results and Discussion	101
Study 4: Clinical Archival Data	109
Method	109
Results and Discussion	111
General Discussion	118
References	139

Appendix A: Test Description	152
Appendix B: Rating Scales	155
Appendix C: Oral Debriefing Simulation Study	161
Appendix D: Oral Debriefing TBI Simulation Study	171

List of Tables

Table 1:	Neuropsychological tests and order of administration	57
Table 2:	Group means and standard deviations for obtained and estimated performance on the TOMM for Study 1	62
Table 3:	Proportion of correct responses on the TOMM by Groups for Study 1	64
Table 4:	Group means and standard deviations for the abbreviated Hiscock for Study 1	67
Table 5:	Group means and standard deviations for the Basic Knowledge Questionnaire for Study 1	70
Table 6:	Correlations between different tests of malingering for Study 1	71
Table 7:	Group mean percentile scores of neuropsychological tests for Study 1	73
Table 8:	Classification of severity of head injury	82
Table 9:	Group means and standard deviations for obtained and estimated performance on the TOMM for Study 2	86
Table 10:	Proportion of correct responses on the TOMM by Groups for Study 2	88
Table 11:	Group means and standard deviations for the abbreviated Hiscock for Study 2	93
Table 12:	Group mean percentile scores of neuropsychological tests for Study 2	95
Table 13:	Mean and standard deviations on the TOMM by groups	103
Table 14:	Proportion of correct responses on the TOMM by Groups for Study 3	105
Table 15:	Group means and standard deviations on the TOMM for Study 4	113
Table 16:	Mean and standard deviations for percentile scores on verbal and visual tests of learning and retention	115

Table 17:	Proportion of correct responses on the TOMM by Groups for Study 4	117
Table 18:	Summary table of TOMM scores for all experiments	120

List of Figures

Figure 1:	Expected group performance on neuropsychological tests	61
-----------	--	----

INTRODUCTION

Malingering is a psychiatric construct used to identify individuals who intentionally try to deceive an examiner for external secondary gain. Although, malingering occurs within a variety of different situations, it is most prominent in the legal system. For example, pleading amnesia or insanity to avoid responsibility for one's actions or claiming disability as a result of an accident are conditions where malingering is likely to occur.

Neuropsychology's role within the legal system is expanding. More neuropsychologists are serving as expert witnesses to determine whether the individual's putative cognitive deficits are related to the accident or can account for the individual's behaviour. In order to accurately determine that an individual's complaint about a cognitive deficit is legitimate, the professional must not only have a clear understanding of how changes in brain functioning impact on behaviour, but she/he must also have knowledge of how individuals may fake neuropsychological deficits.

To detect malingering within a neuropsychological context, professionals have relied largely on clinical experience and knowledge of brain behaviour relationships. However, such approaches are not always objectively verifiable in a manner which the legal system generally demands. In order to avoid such difficulties, tests designed to detect malingering have been developed. However, there are no entirely satisfactory measures currently available.

The following dissertation is divided into five sections. The first three

sections review the literature on malingering and address the following issues: (1) what is malingering, (2) how is it measured and, (3) what problems are associated with malingering research. Based on the literature review, several research studies were conducted. The fourth section describes four integrated studies that were conducted to examine the validity of a new measure of malingering, the Test of Memory Malingering (TOMM). The final section is a general discussion and integrates findings from the current studies with the existing literature and critically evaluates the Test of Memory Malingering.

I. What is Malingering?

A well known scenario from television portrays a whiplash victim winning a compensation suit and after a few days, she/he is seen participating in a strenuous aerobics class. Lawyers have been depicted as charlatans, advising clients to wear a neck brace or use a wheelchair to increase monetary settlements. Feigning illness ranges from children playing sick to avoid school to adults avoiding work, claiming "mental health days," or "blue-collar flu" (Zielinski, 1994). These examples all illustrate feigning of illness for secondary gain, or avoiding an unpleasant aspect of life. They are all examples of malingering.

In terms of a formal psychological assessment, when someone deliberately produces low performance on a measure, the person is said to be 'malingering' (Franzen, Iverson & McCracken, 1990). Malingering is defined as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work,

obtaining financial compensation, evading criminal prosecution, or obtaining drugs" (DSM-IV, 1994, p. 683). Both the intentional nature and occurrence of an external incentive differentiate malingering from other motivational disorders. For example, Somatoform Disorders (e.g., Conversion Disorder) are not under the voluntary/conscious control of the individual. On the other hand, with Factitious Disorders (e.g., Munchausen or Ganser syndromes) the person voluntarily produces symptoms but there is no external incentive. Here, the person is thought to have a psychological need to assume a sick role. In the case of Munchausen syndrome the person has been repeatedly hospitalized (often in many different locations), is well versed with regards to the medical system, and has undergone numerous investigations including surgical procedures (Franzen et al., 1990; Rogers, 1988).

Although malingering of many different types of disorders is possible (e.g., feigning paralysis or insanity), the following discussion will focus on neuropsychological issues such as brain injury or memory impairment. As such, the paper will be limited to a review of neuropsychological assessment techniques, procedures, and research findings as they pertain to malingering. For a review of malingering involving the use of personality measures, the reader is referred to Greene (1988), Franzen et al. (1990) and/or Schretlen (1988).

The detection of malingering is an important issue in neuropsychology for a number of reasons. First, as health care costs increase and the availability of services decrease, it is imperative that health care professionals be able to identify those who truly need and will benefit from treatment. Second, neuropsychological assessment

relies on the principle that the person being assessed is motivated to do well. The validity of the assessment, therefore, becomes questionable if the person is not putting forth his/her best effort (Prigatano, 1993). However, individuals may not be motivated to perform to the best of their abilities for a variety of reasons (e.g., depression, disorientation/confusion, anxiety, or malingering), and it is the responsibility of the neuropsychologist to be able to differentiate between neurological, psychiatric, and/or intentional reasons for poor effort. Third, neuropsychologists are increasingly being asked to participate in legal matters as expert witnesses. As neuropsychologists are called upon to provide expert testimony in criminal proceedings and in cases of personal litigation, they are often asked to state to what degree the results of the individual's assessment is a valid indicator of his/her true ability. Fourth, as Heaton and others (e.g., Faust, Hart, & Guilmette, 1988; Heaton, Smith, Lehman, & Vogt, 1978; Rogers, 1988) have illustrated, it is possible to give a low performance on neuropsychological tests and not be detected. Thus, it is important to be able to detect when malingering is present. In order to accomplish this goal, neuropsychologists must know what to look for and to be able to develop measures that specifically assess the patient's level of cooperation or motivation.

Incidence

One of the difficulties when studying or diagnosing malingering is that the true incidence of the disorder is not known. Incidence is defined as "the frequency of occurrence of an event or a condition in relation to the population under examination"

(Reber, 1985, p. 349). Generally, malingering is only suspected and can not be proven with 100% certainty unless the person either admits to the deception or is caught performing the function/behaviour that she/he stated she/he could not perform. Given the nature of the disorder (i.e., to deceive without being identified), we do not know how many successful malingerers there are. We can only estimate the incidence of suspected malingering.

The estimated incidence of malingering varies from 1% to 50% depending on the population examined (Resnick, 1988). A higher rate of malingering is expected in medicolegal cases or within prison populations than within a general neurological setting (Franzen et al., 1990; Rogers, 1988). Values also vary depending on the instrument(s) used for the detection of malingering. In studying criminal defendants pleading insanity, Rogers (1986a) reported an incidence of 21% of suspected malingering, while Hawk and Cornell (1989) reported that in a prison population, 8% of inmates were suspected of malingering psychotic symptoms. In samples comprised of mild head injured individuals seeking compensation, the reported incidence of malingering has varied from 18% to 60% (Binder, 1993a; Binder, Villanueva, Howieson & Moore, 1993; Greiffenstein, Baker & Gola, 1994; Guilmette, Whelihan, Sparadeo & Buongiorno, 1994). Trueblood and Schmidt (1993) reported an incidence of probable malingering of 7.5% in 106 consecutive outpatient referrals for neuropsychological evaluation. All patients were seeking compensation. They also classified another 7.5% of their sample as "questionable" malingerers which brings their total reported incidence to 15%.

The reluctance of many examiners to screen for malingering makes it difficult to establish the incidence of malingering across different studies. Sceptical examiners may include a screening measure for deception where being impaired has obvious personal or financial benefit. Others will only consider assessing for deception when behaviour during testing is unusual or inconsistent. Franzen et al. (1990) suggested a set of three criteria for "determining when to assess for malingering" (p. 249): (a) if the patient is involved in legal proceedings, (b) if there is an obvious external incentive, (c) if there is a history of malingering or factitious disorder. A broader view is that since disability offers the potential for reward, screening for deception may be advisable in all disability assessments. For example, patients may feel that they may be denied rehabilitation services unless they appear severely impaired. Thus, they may exaggerate their deficits. If that is the case, it is important to screen for deception in every assessment, not just when the patient is involved in legal proceedings.

Terminology

The terminology used to describe malingering varies. As examples, dissimulation, faking, deception, compensation neurosis, exerting minimal effort, feigned disability, exaggerating symptoms, noncompliance, intentional fabrication, motivational problem, substantially less than optimal functioning, dishonest test taking, fabricating symptoms, uncooperativeness, and intentionally not putting forth best effort have all been used. It is suspected that the reason for the diverse terminology is related to the pejorative connotation attached to the term malingering

combined with the lack of certainty in diagnosing malingering.

The negative connotation, along with difficulty of diagnosis, creates a reluctance amongst practitioners to label someone as malingering. Patients themselves may become angry with the diagnosis of malingering which may leave the practitioner open to possible litigation or other more dire consequences. For example, Binder (1992) reports an incident where a patient who received a diagnosis of malingering lower back pain murdered the two diagnosing orthopedic surgeons. Typically, neuropsychologists tend to avoid a diagnosis of malingering, stating instead that the assessment is not valid due to lack of effort or poor motivation, or that the pattern of performance does not match any known neurological and/or psychiatric disorders that the examiner is aware of.

Another possible reason for the ambiguity of terms for malingering lies in the difficulty of conceptualizing the construct of malingering. Is malingering an "all-or-none" phenomenon or does it represent behaviour that lies along a continuum? For example, is the person who exaggerates symptoms fundamentally different from the person who fabricates those same symptoms?

Very few articles that have reviewed malingering have addressed the issue of how it should be conceptualized (e.g., Nies & Sweet, 1994; Rogers, 1988). The earliest work on the conceptualization of malingering comes from Lipman (1962; as cited in Nies & Sweet, 1994). He described four levels of malingering: invention, perseveration, exaggeration, and transference. Invention refers to fabrication of symptoms/deficits; perseveration refers to symptoms that were previously present but

have since subsided; exaggeration refers to symptoms that are presented as worse than they are; and transference refers to "symptoms that are fraudulently attributed to a particular injury" (Nies & Sweet, 1994, p. 502). Rogers (1988) also conceptualized gradations of malingering. He described three levels of malingering: mild, moderate and severe. Mild malingering represents exaggeration of symptoms and "plays only a minor role in differential diagnosis," whereas a moderate degree of malingering is said to occur when "the patient, either through exaggeration or fabrication, attempts to present him- or herself as considerably more disturbed than is the case" (p. 6). Severe malingering occurs when "the patient is extreme in his or her fabrication of symptoms to the point that the presentation is fantastic or preposterous" (p. 6). Some (e.g., Erickson, 1990) have even argued for the removal of the concept of malingering from clinical use as they feel it serves no purpose to diagnose the patient as malingering (Pankratz & Erickson, 1990).

In terms of assessment or identification of malingering, different degrees or gradations of the disorder are less relevant. That is, if during an assessment malingering is suspected, regardless of whether it is considered mild, moderate or severe, the assessment becomes questionable and needs to be interpreted with caution. In terms of legal issues, gradation of deception is a non-issue since any attempt at deception is a fraudulent act. However, in terms of therapeutic intervention, the distinction between different types or severity levels of malingering likely represents a fundamental difference which will impact on possible therapeutic interventions. For example, Lezak (1995) suggests that when a patient is suspected of exaggerating

his/her deficits (as opposed to fabricating) in order to ensure that they are noticed by the clinician, she will ask them to come back for more testing. At that time she "gives the explanation that 'it seems that for some reason-perhaps you were [tired, nervous, did not feel well] you did not do as well as you could" (p. 792). She then explains the importance of performing to the best of his/her ability and that in order to understand how the impairment impacts on the patient that they must try his/her best (see also Pankratz, 1979; 1983). Thus, in terms of clinical practice, differentiating between mild and more severe forms of malingering is important and relevant whilst in terms of identifying malingering it is relatively less important.

Although controversy exists in how malingering is conceptualized, certain tests have been developed to enhance detection of feigned deficits. The following section will describe how malingering is identified.

II. How is Malingering measured?

Malingering has typically been measured by examining neuropsychological invalidity indicators. These indicators are based on inconsistent performance that occurs in a variety of different situations. For example, scores are examined for inconsistency across tests measuring the same construct (attention or memory). Personality measures have also been examined for a "malingering" or "fake-bad" profile (e.g., MMPI or Rorschach). Malingering may also be assessed by employing tests specifically designed to detect malingering (e.g., Rey 15 Item Test).

The following section is a discussion of how malingering is assessed. Neuropsychological invalidity indicators will be discussed first, followed by a

description of specific tests of malingering. Personality measures will not be discussed in the current proposal and the reader is referred to Greene (1988), Franzen et al., (1990), and Schretlen (1988).

Neuropsychological Indicators

As research has shown, no test has been developed which is not susceptible to malingering (e.g., Faust et al., 1988; Heaton et al., 1978; Rogers, 1988). However, that does not mean that those who feign impairment go undetected. There are a number of indicators that signify that the individual may not be performing to the best of his/her ability (e.g., Cullum, Heaton & Grant, 1991; Larrabee, 1992; Lezak, 1995). Whether these signs indicate that the person is intentionally performing poorly is another issue which needs to be addressed once the validity of the assessment becomes questionable.

Inconsistency within an evaluation can occur (1) between test performance and the profile of the injury, (2) between reported behaviour and test performance, and (3) between tests both within the evaluation and across repeated evaluations (e.g., Binder, 1992; Franzen et al., 1990; Lezak, 1995; Nies & Sweet, 1994; Pankratz, 1988; Rogers, 1988; Trueblood & Schmidt, 1993; Schretlen, 1988). Each type of invalidity indicator will be discussed below.

Inconsistency Between Observed and Expected Behaviour. Neuropsychologists must be knowledgeable about various syndromes and neurological disorders, including the anatomical substrate involved, and how various disorders impact on neuropsychological functioning. They must also be aware of how someone of a

similar age and education should perform on various neuropsychological tests (i.e., normative performance). Once the knowledge of brain behaviour relationships is attained, it becomes possible to detect deviations from the expected profile (e.g., Pankratz, 1988; Wasyliv & Cavanaugh, 1989). For example, recovery from a closed head injury typically follows a certain progression of recovery of function depending on a number of well-documented factors (e.g., severity, focal impairment). Based on this progression gradual improvement (either as a result of recovery of function or practice effects) should occur with serial evaluations (e.g., Cullum et al., 1991). If performance deteriorates or fluctuates across evaluations, the validity of the assessment becomes questionable and the reasons for the performance need to be determined.

By understanding the relationship between a particular disorder and neuropsychological functioning, it is also possible to know how individuals should perform on various neuropsychological tests. For example, it is well documented that head injured individuals and amnesics do not show an impairment in their ability to recall autobiographical information, such as their name or date of birth (Delis, 1989; Wiggins & Brandt, 1988). Research has also demonstrated that recognition is easier than recall for various types of patients and for normals (e.g., Delis, 1989; Wiggins & Brandt, 1988). Poor effort or malingering may be suspected when this typical pattern of performance is not observed (e.g., when individuals perform better during recall than recognition) (Bernard, 1990; Millis & Ricker, 1994).

Understanding the relationship between the neurobehavioural disease and

neuropsychological test performance allows one to know what makes "neuropsychological sense" (e.g., Andrikopoulos, 1994; Mittenberg, Azrin, Millsaps & Heilbrunner, 1993). For example, simulation studies have reported that those asked to feign deficits showed greater impairment on attention/concentration than on general memory functioning (e.g., Bernard, McGrath & Houston, 1993; Mittenberg et al., 1993). The above finding is inconsistent with the finding of intact attention span in various patient groups (Butters et al., 1982; Levin et al., 1982) and it does not "make sense" since one needs to attend to stimuli in order to remember them. Therefore, attention/concentration should be equivalent to general memory functioning, if not better than memory. Another example of findings that do not make sense is the occurrence of nonlateralized sensorimotor deficits or sensory deficits crossing the midline (e.g., Mensch & Woods, 1986).

Inconsistency Between Reported Behaviour and Test Performance. Another way to assess whether behaviour and test performance are consistent is to evaluate behaviour as it occurs in everyday functioning. Unfortunately it is often not feasible to directly observe how someone functions in everyday life. However, it is possible to observe his/her behaviour during the evaluation. For example, Pankratz (1988) cites an example of a patient who "travelled several hundred miles on public transportation to be tested yet complained of orientation and memory problems that would have precluded her safe arrival" (p. 181).

Inconsistencies in everyday functioning can also be evaluated through the interview. For example, does the patient claim total disability yet manage to live

independently? Lezak (1995) gives an example of a patient who claimed that he could not go back to work because of memory problems. Yet, while on disability he became an excellent cook and volunteered for a charity organization, demonstrating that his memory impairments were not sufficiently severe to interfere with his daily life.

Inconsistencies Within and Between Evaluations. Inconsistencies within an assessment have been reported and are indications of questionable validity.

Inconsistent performance across tests measuring the same ability have been reported in several studies. Representative examples are as follows: impaired performance on the Finger Tapping Test but average performance on Digit Symbol, poor performance on Digit Span Forward yet intact ability on mental calculations, impaired primary memory capacity as measured by the Digit Span, yet performance within normal limits on word-list tasks or paired-associate learning (e.g., Franzen et al., 1990; Lezak, 1995; Pankratz, 1988; Nies & Sweet, 1994; Rogers, 1988; Schretlen, 1988; Waslyiw & Cavanaugh, 1989).

Near-misses or Ganser-like responses have also been reported to be signs of invalidity. That is, responses that are close approximations to the answer but are clearly wrong ($4 \times 2 = 7$; colour of the American Flag = red, white and green). Although most reviews of malingering research include Ganser-like responses as signs of invalidity, more recent research indicates that this variable may be part of clinical folklore and occurs rarely (e.g., Trueblood & Schmidt, 1993).

The occurrence of excessive impairment has also been noted to be a sign of

invalidity (Binder, 1990), as has impaired performance for the wrong reason. For example, Pankratz (1988) gave the example of an individual performing poorly on Trails B, not because he had difficulty shifting from one set to another, rather because he exhibited a Ganser-like response of forgetting the alphabet part way through.

Other inconsistencies within neuropsychological evaluation include failure on easy items yet passing more difficult items, better recognition performance than recall or showing a normal pattern of responses when it should be impaired. For example, claiming amnesia yet showing both primacy (unexpected) and recency serial position effects (expected), indicating that material is being transferred to long-term stores (i.e., individual is capable of learning).

Repeated evaluations are probably the best way to observe any inconsistencies both within and between evaluations. No research has been conducted examining test-retest simulated malingering behaviour, although it can be imagined that malingering would be very difficult to successfully carry out. Indeed, not only is it difficult to maintain one's deception in a consistent manner during a lengthy assessment, but it is hard to imagine how the individual could maintain continuity in deficit production over a prolonged interval.

Repeated assessments could also be evaluated for a possible lack of practice effects on measures that typically improve due to the second exposure of the stimuli (e.g., timed, novel stimuli such as Block Design). As pointed out by Lezak (1995), absence of practice effects for those who have demonstrated some learning ability, such as finding the examiner's office, again would indicate poor cooperation.

Specific Tests of Malingering

For a test to be an adequate measure of malingering several criteria need to be met. First, the test must accurately differentiate between malingerers and non-malingerers. Second, the test must appear difficult yet be quite simple. Third, the test must have face validity as a measure of cognitive function to the individual being examined. That is, the test must appear to be a legitimate test of neuropsychological functioning, otherwise the malingerer will be suspicious that the test is trying to measure motivation. Fourth, the test should not be influenced by or related to non-motivational factors, such as cognitive ability or impairment, age or education. That is, an individual with cognitive impairment (e.g., amnesia or brain-damage) or someone with low intellectual ability (e.g., mentally retarded) should be able to perform at perfect or near-perfect levels. Near-perfect performance ensures that if failure occurs on the test, it is likely attributable to one reason, that is malingering (Guilmette, Hart & Giuliano, 1993).

The following section will describe the specific measures of malingering currently available. Discussion will focus on current research findings and whether the tests are deemed valid measures of malingering.

Dot Counting Test: Grouped & Ungrouped Dots (Rey, 1941). One of the earliest tests designed to detect malingering was developed by Rey (1941, as cited in Lezak, 1995). In the Dot Counting Test the patient is presented with a series of cards on which dots are arranged, first in a random fashion (ungrouped) and then in patterns (grouped). The patient is asked to count the number of dots as fast as she/he can.

Response times are recorded.

The assumption underlying the ungrouped Dot Counting Test is that as the number of dots increase, time to count the dots should also increase. The number of dots in the ungrouped phase is presented in a random order. First 11 dots are presented, followed by 19, 15, 23, 27 and 7 dots. Presenting the stimulus in a randomized order "determines whether the patient's failures are systematically associated with the altered intensity" (Nies & Sweet, 1994, p. 531) or are reflective of malingering. According to Lezak (1995) more than one deviation from the expected pattern should raise suspicion about malingering. Scores are compared to performance of normal and brain injured adults.

The grouped dot counting test consists of a series of dots arranged in various patterns (e.g., a square, a diamond). Again, the number of dots presented is varied and time taken to count the dots is recorded. The assumption underlying the grouped Dot Counting Test is that by presenting the dots in arranged patterns, time to complete the task should be less than what is seen during the ungrouped phase. Although no reference is given to empirical evidence leading to the following assumption, Lezak (1995) states the "when there is little difference or the time taken to count the grouped dots exceeds that for the ungrouped dots, the subject's cooperation becomes suspect" (p. 800).

Very little research has been conducted using the Dot Counting Test and the results that are available are equivocal with respect to the efficacy of the Dot Counting Test to detect malingering. Greiffenstein et al. (1994) found that the Dot

Counting Test did not differentiate between documented traumatically brain injured patients, patients with persistent postconcussive syndrome, and probable malingerers. Paul, Franzen, Cohen and Fremouw (1992) also found that the Dot Counting Test did not significantly differentiate between brain damaged patients and psychiatric inpatients. However, the control group performed significantly better than either of the patient groups. In the same study, Paul et al. asked the control subjects and psychiatric patients to simulate brain damage. Results suggested that there were no significant difference between the simulation groups and the control group when time was the dependent variable. However, the genuinely brain damaged group made significantly fewer errors than either of the simulated groups, which were equivalent to one another. Beetar and Williams (1995), using a computerized version of a modified Dot Counting Test, found a similar effect when error scores were reported. They found that students asked to mangle made significantly more errors than students asked to perform to the best of their ability. When time was examined, Beetar and Williams (1995) reported a significant difference between groups, with malingerers performing significantly slower than controls. However, examining the mean values revealed that the difference was too small to be clinically useful (a 1 second difference existed between groups).

Philpott and Boone (1994) and Martin, Haut, Stainbrook and Franzen (1994) assessed possible age and cognitive effects on performance on the Dot Counting Test. Neither study examined probable malingerers or simulators. Philpott and Boone (1994) examined performance in an aged sample (healthy controls and those suffering

from Alzheimer's Disease) and found that only the more severe levels of cognitive impairment ($MMSE < 20$) influenced results. No age effects were found.

Martin et al., (1994) examined performance in an adolescent sample (neurologically and psychiatrically impaired and healthy controls). They found that younger subjects (9-11 years) performed significantly more slowly on the Dot Counting Test than older subjects (12-17 years), regardless of group membership. The impaired groups (neurological and psychiatric) were also found to generally perform at lower levels than the healthy controls.

In summary, the majority of studies examined showed there was no significant difference in time taken to complete the Dot Counting task between probable or simulated malingers and control groups. Simulated malingers are found to make more errors than control subjects (e.g., Beetar & Williams, 1995; Paul et al., 1992). However, in the one study that identified probable malingers from their patient sample (Greiffenstein et al., 1994) error scores were not reported. Thus, it is unclear if error scores would differentiate between genuinely impaired individuals, suspected malingers and controls.

It would appear that the Dot Counting Test is not a reliable indicator of possible malingering. One serious problem with the Dot Counting Test is that genuine cognitive impairment lowers test scores, regardless if number of errors or time are used as the dependent measure. Thus, one can never be sure if low performance is the result of malingering or genuine impairment.

Rey Fifteen Item Test (Rey, 1964). This nonverbal memory test was developed to

test the validity of memory complaints. The patient is told that she/he will have 10 seconds to remember 15 different items. Following presentation of the items, the patient is asked to copy all of the items she/he remembers. The number of correct items and/or sets remembered is recorded. The total number of items to be remembered (15) is stressed with the assumption that individuals suspected of malingering will lower their performance. In actuality, all that needs to be remembered are three different concepts (letters, numbers, simple geometric designs). The items that are presented are: A B C; 1 2 3; a b c; circle, square, triangle; I II III.

Goldberg and Miller (1986) assessed psychiatric and mentally retarded (MR) individuals on the Rey Fifteen Item Test (Rey FIT). Results revealed that cognitive ability impacted on performance. Only 50% of the psychiatric and 0% of the mentally retarded sample recalled 15 items. The authors recommended using a cutoff score of 9 items recalled to determine malingering. However, using this cutoff would have classified 38% of the MR sample as possible malingerers. Hayes, Emmons and Lawson (1993) also examined psychiatric patient's performance on the Rey FIT. They found that performance on the test was significantly correlated with intelligence. Given this, they suggested that "the use of a single cut-off score to interpret Rey scores is inappropriate" (p. 1333). Instead they suggested the use of ability-based norms for more appropriate interpretation of Rey scores. That is, for different IQ scores they provided a list of percentile ranks of Rey scores. For example, to obtain a score that would be at approximately the 50th percentile on the Rey FIT, a person with an IQ of less than 65 would need to obtain a score of 6. In contrast, a person

with an IQ of greater than 89 would need a score of 14 to obtain the same percentile score. As can be seen, cognitive ability influences the score regardless if the person is motivated to perform well or not.

Acquired neurological impairment also influences test results (e.g., Davidson, Suffield, Orenczuk, Nantau & Mandel, 1991; Lee, Loring & Martin, 1992; Martin, Haut, Stainbrook & Franzen, 1994; Morgan, 1991; Philpott and Boone, 1994). Bernard and Fowler (1990) found that neurologically impaired subjects recalled significantly fewer items than control subjects. No significant differences were observed in the number of sets recalled. The authors stated that if they used the cutoff of 9, 89% of the brain damage sample and 100% of the control sample would have been classified correctly. However, they recommended a cutoff of 8, which resulted in 100% correct classification of both groups.

Bernard (1990) also tested two groups of simulators (incentive and no incentive) and a control group on the Rey FIT. He found no significant difference between the groups on total number correct, although control subjects remembered significantly more sets of items compared to simulators. He also reported no difference between simulation groups and suggested that the test order influenced the results. That is, the test was given last, after more difficult memory tests, which could have influenced the simulators into suspecting the nature of the test and they adjusted their performance accordingly. Bernard, Houston and Natoli (1993) compared performance of control subjects and simulators when the Rey FIT was administered at the beginning of a series of tests. Results showed that control

subjects scored significantly higher than simulators on both total number correct and number of sets recalled correctly. The authors concluded that for the Rey FIT to be effective it should be given before other more difficult memory tests. When the results were compared to those reported by Bernard (1990), the test order appeared to influence results. It should be noted however, that no study has directly tested this assumption using a counterbalanced order of test presentation. Also, test order of the Rey FIT has not been observed to influence results in other studies (e.g., Bernard & Fowler, 1990; Schretlen, Brandt, Krafft & Van Gorp, 1991). It should also be noted that in the Bernard et al. (1993) study the simulators scored significantly below the control subjects, although their mean performance was above the author's previously suggested cutoff of 8 (mean correct was 10 items). Thus, although the simulators performed at lower levels than the control subjects, the majority would not have been identified as malingering based on suggested cutoff scores.

Lee et al. (1992) tested a series of individuals with temporal lobe epilepsy ($n=100$), neurological outpatients not undergoing litigation ($n=40$) and those undergoing litigation ($n=16$). The study revealed that litigating outpatients scored significantly lower than either of the other two patient groups, which were equivalent to each other. Results indicated that 96% of the temporal lobe epilepsy group and 95% of the outpatient non-litigating group scored 8 or above, compared to only 62% of the outpatient-litigating group. Based on their findings, the authors suggested a cutoff score of 7 that was at or below the 5th percentile for their sample of non-litigating patients.

In an interesting study, Davidson et al. (1991) compared a number of different groups on the Rey FIT. The groups included: chronic pain patients ($n=50$), spinal cord injured patients ($n=25$), brain injured patients ($n=52$), controls asked to simulate malingering ($n=40$) and control subjects ($n=40$). Davidson, et al., (1991) found that simulators scored significantly below control subjects, but were not significantly different from the brain injured or chronic pain patients. In detecting simulators, the Rey FIT had a sensitivity of 60% and a specificity of 74%. Given the results, the authors concluded that the Rey FIT has low validity in detecting simulated memory impairment that should preclude using the test clinically.

Guilmette, Hart, Giuliano and Leininger (1994) compared Rey's 15 Item Test to an abbreviated version of the Hiscock Forced-Choice Procedure (to be discussed later). Subjects included moderately to severely brain damaged patients ($n=20$), severely depressed patients ($n=20$) and normal controls instructed to simulate memory impairment ($n=20$). Results revealed that the brain damaged group performed significantly worse than the other two groups on total number of items remembered on the Rey FIT. The authors also used a variety of different cutoff scores. Using the cutoff of 7 items recalled proposed by Lee et al. (1992), 40% of the brain damaged and 20% of the depressed samples would have been misclassified (i.e., suspected of malingering). Most of the simulators (95%) "performed above the cutoff and would not be suspected of malingering" (p. 290). That is, the Rey FIT would have had a sensitivity of 5% and a specificity of 75% using the cutoff of <7 items recalled. With a cutoff of <8, sensitivity is reported to have been at 15%.

whereas specificity was at 72.5%. Sensitivity increases to 25% and specificity decreases to 70% with a cutoff of <9 items recalled. As can be seen, none of the reported cutoff scores resulted in an adequate degree of sensitivity (i.e., detection of simulators as simulators).

Only two studies have examined performance of the Rey FIT with suspected malingerers (Greiffenstein, et al., 1994; Schretlen, et al., 1991). Greiffenstein et al. (1994) compared performance of verified traumatically brain injured ($n=33$), patients suffering from persistent postconcussive syndrome ($n=30$) and probable malingerers ($n=43$). Probable malingerers were persistent postconcussive syndrome patients who met additional criteria of suspected malingering (e.g., they had claims of remote memory loss, total disability after one year, improbable symptom history contraindicated by objective measures and/or two or more ratings of severe impairment on neuropsychological tests compared to appropriate norms). Results revealed that probable malingerers scored significantly lower than the other two groups of patients, whose scores were not significantly different from one another. The authors reported the sensitivity and specificity values of the three groups. A cutoff of 9 items resulted in a sensitivity of 62% and a specificity of 88% when comparing probable malingerers to traumatically brain injured. That is, 62% of the malingerers were correctly identified and 12% of the traumatically brain injured were misclassified as probable malingerers. When probable malingerers were compared to persistent postconcussive syndrome, sensitivity values were reported to be at 63% and specificity at 93%. Two weaknesses of the study need to be mentioned. The first is

that the traumatically brain injured patients were all undergoing litigation that may have influenced their test scores. Secondly, no control group was included for a comparison of impaired individuals to a non-impaired group. We can only speculate that patients' scores are lower than what would be seen in the general population.

The second study to include suspected malingerers was by Schretlen et al. (1991). In an ambitious study they compared performance of nine subject groups on the Rey 15 Item Test. Groups included community volunteers instructed to fake amnesia ($n=47$), inpatient substance abusers instructed to fake insanity ($n=22$), suspected malingerers ($n=7$), amnesics ($n=10$), mixed neuropsychiatric patients ($n=20$), traumatically brain injured ($n=55$), mixed dementia ($n=9$), severe psychiatric patients ($n=34$) and normal controls ($n=80$). Results showed that number of items recalled correlated with mental status ($r=.81$) and with IQ ($r=.55$), suggesting that the test is related to cognitive ability. No significant difference was found between controls and those faking amnesia, although those faking amnesia scored significantly higher than those with genuine amnesia (cf. Wiggins and Brandt, 1988). In fact, the authors note that those simulating impairment scored above the accepted cutoff of 9, while many patients scored below that cutoff. As depicted graphically, those suspected of faking had an overall lower mean score, although the authors report that "many" recalled at least 9 items correct (scores were not reported).

In summary, the Rey 15 Item Test is sensitive to intellectual ability, as well as to specific and diffuse neurological impairment. Given that some studies have not been able to differentiate simulators from controls on the Rey FIT (e.g., Beetar &

Williams, 1995; Bernard, 1990; Schretlen et al., 1991) while others have reported that the test is sensitive to genuine impairment (e.g., Davidson et al., 1991; Goldberg & Miller, 1986; Philpott & Boone, 1994), the Rey FIT does not appear to be a reliable measure in differentiating malingering from other diagnostic groups.

Symptom Validity Testing. More modern measures of malingering have stemmed, in part, from symptom validity testing. Symptom validity testing developed out of a need to determine, with more certainty, if a person was malingering. As the previous sections have shown, the measures that had been developed to detect malingering were sensitive to genuine cognitive impairments. Thus, it was difficult to decide if lowered performance was due to decreased effort (i.e., malingering) or to possible cognitive impairment. Symptom validity testing enables the practitioner to empirically determine if the patient was malingering, since malingering performance is determined by the laws of probability.

Symptom validity testing (SVT), a term originally coined by Pankratz and colleagues, assesses putative sensory, perceptual, or memory deficits (Pankratz, Faust & Peed, 1975; Pankratz, 1979; Pankratz, 1988). This technique requires that a stimulus (target) be presented for numerous trials and the subject's task is to choose from two-alternative choices the target previously shown. The stimulus used depends on the presenting problem. Each test is designed specifically for the type of reported deficit. For example, Pankratz et al. (1975) used this technique with an individual who claimed to have bilateral hearing loss. The patient's task was to identify which stimulus, a blue or red light, was paired with a tone. One hundred trials were

presented. Based on chance performance, the patient should be able to correctly identify 50 responses if she/he were guessing. In this case, the patient had a score of 36 correct, which is significantly below chance ($p < .004$). The only way to explain below chance performance is that the person correctly identified the stimulus but chose not to select it on test trials (i.e., reported the wrong stimulus).

Pankratz (1983) presented 3 case studies involving complaints of memory functioning. Using SVT, one test required subjects to identify which light (blue or white) was previously presented following a 15 second interval in which the subject had to complete a second, unrelated task (e.g., a distractor task). Another test required that subjects recall which 3-digit number was previously presented from two alternative choices. In two of the case studies reported, the scores indicated that the subjects were able to correctly identify the right response but choose not to select it on the test trials (i.e., below chance performance). In the third case presented, the person was found to be "cheating" in order to enhance his performance; the individual was not malingering.

Binder and Pankratz (1988) reported a case study where they used SVT to determine if a patient was exaggerating memory complaints. The task involved presenting the patient with a black pen or yellow pencil. Once the stimulus was correctly identified, it was removed and the patient was asked to count from 1 to 20. The patient was then asked to state which object she was shown. Performance was significantly below chance (63 errors out of 100 trials) indicating probable malingering.

SVT has been used with many different types of functional sensorimotor deficits, including paralysis of the legs and arm (Pankratz, 1979) and blindness (Grosz & Zimmerman, 1965; Theodor & Mandelcorn, 1973). It has also been used to assess putative memory impairment. The technique has undergone various modifications by incorporating a distractor task between presentation and recall, as well as by varying the length of the interval between presentation and recall (Binder & Pankratz, 1987; Pankratz, 1983; Pankratz, 1988).

As noted by Pankratz et al. (1975), symptom validity testing is a powerful technique because (1) "there is an intuitive perception that correctly answering half of the time is excessive for someone with a disability and (2) even if the subject realizes that he should get about half correct, small variations from the expected value becomes statistically improbable" (p. 422). However, it does have its limitations. First, the purpose of the test may be obvious to all but the most naive malingerers. As noted by Hiscock and Hiscock (1989) it is "likely that many patients would react with suspicion and resistance to such a simple and repetitive task" (p. 968). Second, a person may score above chance yet still score below expected norms and therefore be malingering. Third, given the nature of SVT, normative data cannot be determined. Each test is designed to tap the specific complaint or deficit of a particular individual. There is no "standard" symptom validity test given. Normative data cannot feasibly be collected on every possible test variation given. The individual nature of SVT is its greatest strength and its greatest weakness.

In order to compensate for the simplicity and lack of standardization of SVT,

yet maintain and enhance its strengths (e.g., forced-choice procedure and appearance of item difficulty), Hiscock and Hiscock (1989) developed their own measure of malingering. This test has been referred to as the Hiscock, the Hiscock Digit Memory Test and the Hiscock Forced-Choice Procedure. The subject is shown a 5-digit number and asked to remember it. Following a delay she/he is shown two 5-digit numbers and asked to point to the one previously seen. Eight 5-digit numbers are used in three blocks of 24 trials for a total of 72 trials. Each block of trials represents a different delay period. The first block incorporates a 5 second delay between presentation and recall, the second block uses a 10 second delay and the final block of trials has a 15 second delay. Feedback is given after each response.

The Hiscock manipulates the perceived difficulty of the test in two ways. First, although remembering a 5-digit number appears to be difficult, the task is quite easy as all one needs to remember is the first digit of the target stimulus. Second, increasing the length of the interval between presentation and recall gives the perception that the test increases the difficulty. After the first and second block of trials the person is told that because they did well (regardless of actual performance), the interval will be increased to make the task more difficult. Feedback is provided to the subject on each trial. Such feedback is thought to "make the malingering patient think he or she is performing 'too well' even if performance is not above the chance level" (Hiscock & Hiscock, 1989, p. 973).

Hiscock and Hiscock (1989) administered this test to a patient who claimed he had a severe memory impairment following a mild head injury and two control

subjects (a person with Alzheimer's Disease and a normal 5 year old child). On the first set of 24 trials, the head injured patient scored at chance level (50% correct). As the perceived difficulty of the test increased, performance deteriorated (25% correct during the 10 second interval and 12.5% correct during the 15 second interval). He received an overall score of 21 out of 72 (29%) which was substantially lower than overall performance of the demented patient (51%) and of the child (82%). It was also significantly below chance ($p < .0002$). Given that he performed at chance level on the first block of trials, he would not have been identified as malingering had the symptom validity technique been used. It was only after he was told that the task was going to become more difficult during the second and third blocks of trials that performance decreased to levels below chance.

When the score falls significantly below chance performance, malingering is highly suspected since the only explanation for such poor performance is that the person identified the correct response but choose to ignore it. However, not all individuals suspected of malingering score below chance levels. Rather, they may have a score that is above chance but significantly below the scores expected for a comparable group. Thus, it becomes important to know how other clinical groups perform on tests of malingering. That is, if malingerers are trying to simulate head injured behaviour one must know how someone with a head injury, who is not malingering, performs on the measure. For a test of malingering to be clinically useful, it must be sensitive to detecting malingerers and insensitive to genuinely impaired groups (e.g., head injured, depressed, amnesic).

Guilmette, Hart and Giuliano (1993) and Prigatano and Amin (1993) have used the Hiscock with a clinical population. Guilmette et al. (1993) tested the validity of the Hiscock with simulated malingerers (normal controls asked to fake impairment), a genuinely brain damaged group, a group of depressed patients and a group of normal controls. They found that the normal controls were 100% accurate, the brain damaged group were 96% accurate and the depressed group were 98% accurate. In contrast, the simulators performed significantly worse than the other groups (60% accurate). Of the simulators, only 35% scored below chance. Guilmette et al. suggested that a score does not have to be below chance before one suspects malingering. Rather, a score significantly below expected norms should raise suspicion of malingering. Based on their results, they suggested a cutoff of 90% correct (65 and less) as being suggestive of malingering compared to a score of 29 (40% correct) which represents significantly below chance performance. Using a cutoff of 90% correct results in both high sensitivity and specificity.

Prigatano & Amin (1993) administered the Hiscock to a group of patients suspected malingering, a group with brain damage, and a normal control group. Results indicated that the brain damaged group performed similarly to the normal controls (95.5% and 99.7% correct, respectively) while the suspected malingerers performed significantly below the brain damaged group but not below chance (73.8% correct). Again, if below chance performance was the criterion used to identify malingerers not one suspected malingerer would not have been identified, even though other test patterns were indicative of malingering (i.e., 0% sensitivity).

Guilmette, Hart, Giuliano & Leininger (1994) tested an abbreviated version of the Hiscock with moderately to severely brain damaged patients, severely depressed patients and with a group of normal controls asked to realistically simulate memory impairment. The abbreviated version of the Hiscock uses 36 trials (12 trials per block) as opposed to the original 72 trials. All other procedures remained the same. Results indicated that both the brain damaged and depressed group were 98% correct compared to 56% accuracy of the simulators. Results are similar to scores obtained with the longer version of the test even though administration time is reduced by half (e.g., Guilmette et al., 1993). Using a cutoff of 90% accurate (33 out of 36) resulted in high sensitivity (85%) and specificity (97.5%).

Guilmette, Sparadeo, Whelihan and Buongiorno (1994) administered the abbreviated Hiscock to disability claimants referred by the Social Security Administration ($n=50$). These claimants were made up of a heterogeneous diagnostic group (e.g., traumatic brain injury, alcohol abuse, seizure disorders, vascular disorders, organic mental syndromes of unknown etiologies) who were referred for neuropsychological evaluation to determine eligibility for disability. Results indicate that 18% of the sample obtained scores lower than the suggested cutoff of 90% correct. The authors also reported that another 20% of the sample received a score that was relatively low (33-35), but not below the cutoff score. Based on the observation that even severely brain damaged individuals often recall 100% of the items correctly (e.g., Guilmette, et al., 1994), the authors suggest that the validity of the assessment should be questioned when someone obtains a score lower than

expected. In the case of the disability claimants, they suggest that a score lower than 35 should raise suspicion of possible low motivation or effort.

The Hiscock test has also been modified and translated into a computer version (Victoria Revision: Slick, Hopp, Strauss, Hunter & Pinch, 1994). This procedure involves presenting 48, 5-digit numbers over three blocks of trials (5, 10 and 15 second retention trials). Digits are split evenly into easy and difficult items. Easy items occur when the foil and target item share no digit in common. On difficult items, the order of the items must be remembered as the target and foil items are identical with the exception of a transposed second and third or third and fourth digit (e.g., 74523 and 74253). Slick et al., (1994) administered this test to traumatically brain injured patients undergoing litigation ($n=10$), students asked to simulate brain damage ($n=20$) and normal controls ($n=22$). Results revealed that groups differed as a function of item difficulty and retention interval. The patients were found to be equivalent to control subjects on easy items and during the 5 second interval. During the 10 and 15 second interval, as well as on the more difficult items, the patients performed at significantly lower levels than control subjects. Simulators were found to perform at significantly lower levels than both controls and patients across difficulty level and interval levels. A criticism of this study is that they did not include non-litigating patients. Past research has shown that those undergoing litigation generally score lower than those not undergoing litigation (e.g., Binder, 1991; Binder & Willis, 1991). Future studies will need to control the issue of compensation. Results also indicate that the Victoria Revision of the Hiscock is

sensitive to genuine cognitive impairment which limits its clinical utility in the detection of malingering.

In summary, the majority of findings from different studies using the Hiscock show a relatively high degree of concordance. The two studies of Guilmette and colleagues using clinical samples (Guilmette et al., 1993; 1994) indicate that individuals with relatively severe brain damage and depression perform at high levels (98%). Similar results were reported by Prigatano and Amin (1993). Prigatano and Amin also reported results from suspected malingerers. When compared to performance of simulators in other studies (e.g., Guilmette et al., 1993; 1994), performance is higher with "true" malingerers as opposed to simulators. This highlights the need to obtain data from suspected malingerers and not to rely solely on data collected with simulators in a laboratory setting.

A criticism of the research using the Hiscock is that to date, no findings have been reported from mild head injured patients, or from patients who are seeking compensation for their disability. The one study that examined disability claimants indicated that approximately 38% of the sample may have been poorly motivated to perform well on the test (Guilmette, et al., 1994). However, no comparison group was incorporated into the design. It would have been interesting for comparison if the authors could have assessed disability claimants who had already been deemed eligible.

Based on the work by Hiscock and Hiscock (1989), Binder developed the Portland Digit Recognition Test (Portland) (Binder, 1990; 1993a; 1993b; Binder &

Willis, 1991). The Portland is different from the Hiscock in that the digits are presented auditorially, a distractor task is presented during the delay (counting backwards) and the length of the delay intervals have been increased (5, 15 and two 30 second intervals). Eighteen trials are presented in each of the four blocks of trials, for a total of 72 trials. Subjects are asked to count backwards from 20 during the 5 second interval, from 50 during the 15 second interval and from 100 during the 30 second intervals. Recall is tested via a two alternative forced-choice visual recognition task and performance feedback is given. Number correct is reported, with performance on the 5 and 15 second intervals combined to form a score of "easy" items (total 36). The items on the 30 second intervals are combined to form a score of "hard" items (total = 36).

Binder and Willis (1991) empirically studied performance on the Portland using six groups: normal control subjects; normal controls subjects asked to simulate a head injury; patients with brain damage not seeking compensation; patients with brain damage seeking compensation; patients with affective disorder; and patients with mild head trauma seeking compensation. As expected, results indicated that normal controls scored higher than all other groups (mean total correct was 64.9) and that simulators scored lower than all groups (mean total correct was 37.9). Binder and Willis suggested using a cutoff score of 39 items recalled when examining the total score (72 trials). This represents the lowest performance of the brain damaged non-compensation seeking group. Using this cutoff resulted in 26% of the mild head trauma compensation-seeking group scoring below that cutoff.

Although results were as predicted, examining the mean scores indicates that the Portland is a difficult task to complete. Controls, on average, achieved only a 90% accuracy rate on total number of items recalled. Recall that performance on the Hiscock and the abbreviated Hiscock generally results in a 100% accuracy rate for controls (e.g., Guilmette, et al., 1993; Prigatano & Amin, 1993). The brain damaged group (not seeking compensation) achieved an accuracy rate of 78% correct, indicating that the test is sensitive to genuine impairment.

In an extension of the previous study, Binder (1993a) examined performance of a group of mild head trauma patients seeking compensation ($n=47$), a group of brain damage patients seeking compensation ($n=39$) and a group of brain damage patients not seeking compensation ($n=34$) on the Portland. All groups were equated on the Rey Auditory Verbal Learning Test, a measure of auditory verbal learning. Results indicated that all groups were functioning at approximately 1.0 SD below age-appropriate norms of non-patients. When using a cutoff of 39 items recalled (on total number of trials), 33% of mild head trauma compensation seeking group and 18% of the brain damaged compensation seeking group scored below the cutoff score. Similar results were reported by Binder and Willis (1991).

A main criticism of the Portland test is that it is a time consuming task, and its score does not contribute to the neuropsychological profile of the individual (Larrabee, 1990). As noted by Lezak (1995), some patients who have performed to the best of their ability on other tests may become frustrated, losing interest in their performance. To overcome this issue Binder (1993b) proposed use of an abbreviated

version whereby the test is discontinued if a person scores 7 correct on the first 9 difficult items. Although this has had the effect of decreasing the administration time without sacrificing the sensitivity of the instrument, it still remains a time consuming test to administer.

As with the Hiscock, the Portland has been computerized (Rose, Petree & Hall, 1993; Rose, Hall & Szalda-Petree, 1995). Results showed that the computer version was equivalent to the manual version (Rose, et. al., 1993). A main advantage of the computer version is that the response latencies are measured. Rose et al. (1995) found that the detection of malingering increased by 23% when response latencies were incorporated into a discriminant function analysis (47% to 70%). However, they also reported that patients suffering from closed head injury had slower response latencies than simulators. That is, genuine impairment results in a slower response.

In summary, results from studies reviewed indicate that the Portland Digit Recognition Test is a somewhat difficult task to complete, even for healthy normal adults. Binder and Willis (1991) using normal control subjects reported an accuracy of only 90%. The test also appears to be sensitive to genuine cognitive impairment (e.g., brain damaged subjects received an accuracy rate of 78%). Another criticism of the Portland, is that much of the data has been reported from the same initial pool of subjects. That is, the original sample of Binder and Willis (1991) has been used in all subsequent work, with the augmentation of a few additional subjects. Also, the original sample did not include any mild head injured patients who were not seeking

compensation. It is important to see how this group would compare to the mild head injured compensation group, often reported in the studies.

Given these criticisms, and the fact that the test remains relatively time consuming to administer, the Hiscock, particularly the abbreviated version, appears to be the most valid measure of malingering that is currently available. However, as previously stated there are limitations with this test.

The Test of Memory Malingering (TOMM). Based on the literature review, a need for a validated measure of malingering has been outlined. No test currently available meets all criteria of a clinically useful test of malingering. The following section will describe a newly developed measure of malingering, the TOMM.

Recent research in the detection of malingering has demonstrated that recognition procedures are particularly sensitive in identifying someone who is not putting forth maximum effort. As noted previously, most recent malingering tests have used some type of symptom validity testing, a recognition methodology pioneered by Pankratz (1979) and further developed by Hiscock and Hiscock (1989) and Binder and Willis (1991). An alternative recognition paradigm for detecting malingering is suggested by cognitive research showing that picture recognition is particularly robust among normal and impaired individuals. This is dramatically illustrated in a study by Standing, Conezio and Haber (1970). Initially, individuals were shown 1100 photographs. Following a 30 minute delay they were shown a sample of 100 pairs of pictures and asked to choose which one they had previously seen (forced-choice procedure). Performance varied from 90% to 98% correct. A

second experiment using 2560 pictures, administered over two or four days, resulted in similar findings.

A procedure similar to this is employed in the Recognition Memory Test, a currently available neuropsychological test, that has been used recently in malingering research (e.g., Millis, 1992; Millis & Putnam, 1994). This test was designed "to detect minor degrees of memory deficit across a wide age range of the adult population" and was originally used to screen adult global amnesic patients (Warrington, 1984, p. 1). The Recognition Memory Test consists of two set of visual stimuli: 50 words and 50 faces. The patient is shown the stimuli (e.g., words), one after another, and immediately following presentation, she/he is shown two words and required to point to the word seen previously. The procedure is then repeated for the nonverbal stimuli (e.g., faces). Although both parts (words and faces) are typically administered together, they can be used separately.

The format allows for performance to be compared to norms for cognitively intact and cognitively impaired samples, as well as to chance performance. When the Recognition Memory Test is used for the detection of malingering, several limitations are presents. First, because the test was specifically designed to detect memory impairment it is sensitive to "true" impairment. That is, word recognition has been found to be sensitive to left hemisphere damage and face recognition has been found to be disrupted by both right and left hemisphere dysfunction. Consequently, in terms of detecting malingering, it is not possible to say that a low score does not reflect neurological impairment. Second, age, education and general cognitive ability have

all been found to be correlated with performance on the test. That is, individuals who are older, less educated and/or less cognitively able will receive a lower score, regardless of their neurological status. Thus, in terms of detecting malingering, the Recognition Memory Test it is not an ideal measure.

A second test using a visual recognition paradigm is the Test of Memory Malingering (TOMM: Tombaugh, 1996). The TOMM consists of line-drawings of 50 common objects followed by a test trial containing a series of 50, 2-choice recognition panels. Two learning trials and a retention trial are administered and volunteers are asked to estimate their performance before each trial¹.

Research has demonstrated that the TOMM has good face validity as a learning/memory test and on an earlier, more difficult, 4-choice version, performance did not vary over a wide range of ages (Hubley, McIntyre & Tombaugh, 1995). However, further research needs to be conducted under simulation conditions, to ensure that the TOMM can differentiate between individuals trying their best and individuals not trying their best. Research also needs to be conducted to ensure that low scores on the TOMM reflect lack of effort/motivation rather than the effects of other variables such as cognitive impairment or amnesia. Four studies are presented which address these research needs. However, prior to describing this research, criticisms of malingering research need to be addressed first.

III. Problems in Malingering Research

Several criticisms of malingering research have been identified. These

¹The final version does not include estimation scores

include: reliance on simulation studies, the instructional set, lack of a manipulation check, lack of assessment of the strategies employed, using a single measure of malingering, lack of an incentive, and limited preparation time (e.g., Franzen, et al., 1990; Nies & Sweet, 1994; Rogers, 1988; Schretlen, 1988). Each criticism is discussed below.

Simulation Studies

A fundamental criticism with malingering research is that studies have relied on experimental simulation paradigms that do not incorporate "true" malingerers. That is, when validating a measure of malingering, normal individuals are asked to simulate a particular type of behaviour (from this point onwards called simulators). Typically, these simulators are undergraduate students who are asked to feign brain damage. Simulator's scores are often compared to those obtained from another group of undergraduate students asked to perform to the best of their ability (control group). Results typically show that simulators perform lower than controls. However, there are several reasons why it is difficult to generalize these results to a clinical setting. Typically, these simulators are less motivated than malingerers. They have less incentive to feign an impairment and therefore are less motivated to provide a credible performance than a true malingerer. That is, the payoff for success is not as great, neither is the consequence of detection. It is also likely that undergraduate simulators are less sophisticated than malingerers.

Although it is feasible to assess probable malingerers (e.g., Greiffenstein et al., 1994; Prigatano & Amin, 1993; Schretlen et al., 1992), it is difficult to identify

them making it unrealistic to include them in an experimental study, particularly during the initial development of the measure. However, this does not mean that research is restricted to using only university students as simulators. One way to increase the clinical relevance and maintain the rigor of an experimental design is to incorporate subjects from a clinical population.

Using what Schretlen (1988) labelled as a "fully-controlled" study, it is possible to compare performance of a simulated group to two control groups: a normative group and a clinical group (e.g., head injured), both asked to perform to the best of their abilities. Schretlen (1988) argued that it would "enable one to estimate the accuracy (hit rate) with which a given test can differentiate between groups" (p. 452). By including a clinical sample, a comparison between simulated malingerers and the target clinical group is possible. Such a comparison would answer the question of whether the simulated group's performance is a realistic portrayal of behaviour observed in the target clinical group. Performance of these two groups (simulated group and clinical control group) can then be compared to a normative group. By using comparison groups, it is possible to establish a baseline of how groups perform on a specific measure. Studies employing this type of design have generally found that both control and clinical subjects outperform simulators, with control subjects performing better than the clinical group (e.g., Baker, Gola & Greiffenstein, 1994; Martin et al., 1990; Mittenberg et al., 1993; Strauss et al., 1994). However, a few studies have found that simulators perform better than their clinical counterparts (e.g., Schretlen et al., 1991).

Another way to make simulation studies more realistic is to ask clinical subjects to simulate malingering (e.g., head injured subjects asked to "fake-bad"). Not only would this avoid using students, but would also increase the generalizability of the results. The simulated clinical group can then be compared to an equivalent clinical group, and a non-clinical group asked to perform to the best of their ability. This would allow for a comparison between student and clinical simulators that would address the question as to whether students are an appropriate group to use in simulation research (i.e., is their performance similar to clinical simulators?).

A literature search revealed that only one study reported using a simulated clinical group (Paul et al., 1992). However, the main purpose of their study was to assess the reliability of a malingering measure and they did not address whether the clinical simulation group performed higher than the non-clinical simulation group. No other findings have been published where a clinical group has been asked to simulate performance. The reason for this is unclear. Perhaps, it relates to a question of ethics. That is, should researchers coach individuals, who may be able to benefit from instruction, how to malingering? This seems unlikely given that studies have used prisoners in simulation designs (e.g., Hiscock, Layman & Hiscock, 1994; Hiscock, Rustemeir & Hiscock, 1994). Although studies (e.g., Binder & Willis, 1991) have used litigating mildly head injured patients and concluded that they are malingering because their performance was lower than more severely injured non-litigating individuals, no one has asked head injured individuals to "fake-bad". By incorporating a clinical simulation group, not only does the generalizability of the

study increase, but it provides an opportunity to evaluate what strategies the clinical simulation group employs.

Although using clinical groups in simulation research would increase the validity of the study, it is not without problems. One main difficulty is that clinical groups are often heterogeneous (Nies & Sweet, 1994). For example, Morgan (1991) grouped individuals with dementia, head injury, developmental disability, stroke, and Parkinson Disease together. This procedure obscures results as to whether one type of patient shows a differential effect on the test of malingering compared to another. It also potentially confounds the effects of demographic variables, such as age and education. For example, those with dementia are typically older and therefore generally have less education compared with a head injured group who are generally younger males (see also Lee et al., 1992). It is therefore important to have strict inclusion/exclusion criteria when dealing with a clinical sample.

It has been argued that simulators are likely to be less sophisticated than malingerers. Few studies have examined the role of simulator sophistication on ability to malingering. Those that have manipulated this variable, report that sophisticated simulators are equivalent to naive simulators (e.g., Baker, Hanley, Jackson, Kimmance & Slade, 1993; Knight & Meyers, 1994; Martin et al., 1993). However, the degree of sophistication has been fairly restricted. For example, Baker et al. (1993) defined sophisticated simulators as those students who attended one lecture on amnesia. Knight and Meyers (1994) used rehabilitation workers familiar with the clinical group they were asked to simulate. They found that they were not

significantly different from students asked to simulate the same behaviour.

In summary, research on malingering usually incorporates a sample of students asked to simulate performance. Many studies have now also included multiple comparison groups (normals, various clinical groups, and probable malingerers) to evaluate the measure under investigation.

Instructions

Another criticism of simulation research involves the instructions used to create different types of simulated impairments (Rogers, 1988). Studies using simulation procedures often do not clearly outline or specify what they want subjects to simulate. For example, subjects have been asked to simulate "mental illness," "emotional disturbance," "brain-damage" or to "fake-bad" (e.g., Beetar & Williams, 1994; Benton & Spreen, 1963; Faust et al., 1988). Such vague terminology likely produces a wide range of responses from different subjects that make interpretation of results difficult. For example, simulating an "emotional disturbance" could result in one individual simulating a psychotic delusional patient and another individual simulating a depressed patient. A related problem is that instructions have generally not stated that the simulation needs to be "realistic."

More recent studies have started to outline the parameters involved within the experiment, depicting the situation in terms of a scenario. For example, scenarios often ask subjects to pretend that they have been in a motor vehicle accident and although they have started to recover, their lawyer suggests that they should pretend, in a realistic manner, that they still suffer from cognitive problems during the

assessment to increase the amount of the settlement (e.g., Bernard, 1990; Bernard et al., 1993; Davidson et al., 1991; Guilmette, Hart & Giuliano, 1993; Guilmette, Hart, Giuliano & Leininger, 1994; Knights and Meyers, 1994; Slick, Hopp, Strauss, Spellacy & Hunter, 1995). At a minimum, this type of scenario allows for replication of results from one research study to the next. By emphasizing the realistic nature of the disorder, it reduces the number of "I don't know" or "I can't remember" responses.

The reason for standardizing the instructional set (i.e., making it explicit) is to increase the likelihood that simulators will more closely approximate the performance of malingers. Given that simulators are not as invested in the outcome as malingers, experimental procedures are needed to provide subjects with information that will increase their ability to simulate in a realistic manner. A fundamental question then arises about whether coached simulators (those given a comprehensive instructional set) perform better (more successful in deception) than uncoached or naive simulators.

In studies that have compared coached and naive simulators, results have been equivocal with some studies showing that coached simulators outperform naive simulators (e.g., Hiscock, Rustemier & Hiscock, 1993; Rose, Hall & Szalda-Petree, 1995) and others finding no difference between groups (e.g., Gfeller, Cradock & Falkenhain, 1994; Hiscock, Layman & Hiscock, 1994). Part of this difficulty is that the amount or style of coaching is varied between studies. For example, Rose et al. (1995) coached simulators by telling them that "extreme performance tips off the

examiner" (p. 127). Hiscock et al. (1994) coached simulators by telling them to "avoid two common errors: (1) appearing excessively impaired and (2) answering difficult items correctly after answering easy ones incorrectly" (p. 446).

The only reported study examining the impact of different instructional set on the ability to malingering found no significant effect of instructional set (Knight & Meyers, 1994). Although it is possible that instructional set does not influence ability to simulate, another possible explanation is that the instructions need to be more explicit. For example, Knight and Meyers (1994) most explicit instruction of "you are having trouble thinking clearly" is a vague statement and does not offer any clues about how to approach the various tests. Perhaps, instructions to malingering need to state explicitly how one should approach particular tests. That is, the performance of the typical brain damaged patient is impaired on list learning tasks, but intact on digit span forward.

In summary, a primary criticism of simulation research is that the instructional set used to create different experimental groups is often brief and does not explicitly state that performance needs to be realistic. This problem has been overcome in more recent studies with the use of scenarios. However, these scenarios are often not published in research studies thereby limiting replication by other investigators.

Debriefing

Related to the issue of instructional set is the issue of subject compliance. Few studies offer a manipulation check at the end of their study to ensure that subjects complied with the experimental procedure (e.g., Mittenberg et al., 1993).

Studies that have incorporated a manipulation check into their design have usually found that some of their subjects have not complied with instructions to give a low performance. That is, they have performed to the best of their ability. For example, Heaton et al. (1978) found that 20% of their sample did not comply with the instructions to fake a low performance. Goebel (1983) found that 10% of his subjects could not malingering while Binder and Willis (1991) reported that 8% of their subjects were non-compliant. Although it is unclear why non-compliance occurs, it is important to document that it does occur. By including a manipulation check, data from subjects that were unable to follow instructions can be removed from the data analysis.

Not only is debriefing necessary to measure compliance, it also allows the opportunity to assess strategies employed by simulators. Few studies have assessed the strategies used by simulators or inquired as to whether subjects thought they were successful in their simulation attempt. By including a debriefing that examines strategy use, researchers may gain insight into how malingerers approach neuropsychological testing. This in turn would also generate knowledge for future test development.

A literature search revealed only one study which addressed the issue of strategy use. Beetar and Williams (1994) reported to identify malingering strategies on test performance. However, their conclusions about how the subjects approached the task were inferential and based on the types of errors made (e.g., random responding, delayed responding). They did not ask their subjects how they

approached each task, why they choose one strategy over another or whether they switched strategies from one test to another, and if so why. Given that malingerers rarely identify themselves as such, it is necessary to explicitly ask simulators how they approached the task of malingering if knowledge is to be advanced in this area.

It is also important to ask simulators about the tests they were given. That is, which tests they felt were easy or difficult to fake, and which ones they thought were designed specifically to detect malingering and so on. This will give researchers insight into how tests are perceived by the lay public.

In summary, it is necessary to include a manipulation check at the end of simulation studies to ensure that subjects have complied with the instructions. Studies that have included this have shown that not all subjects complied with the instructions to simulate a low performance. It is also advisable to directly inquire about the type of strategies that simulators used to feign impairment on the tests, and whether they thought they were successful in their attempt to deceive.

Single Measure

Some studies have compared how groups perform on a single measure of malingering without including other neuropsychological tests (e.g., Baker et al., 1993; Benton & Spreen, 1963; Bruhn & Reed, 1975; Millis, 1994; Millis & Putnam, 1994). This methodology does not approximate a neuropsychological examination and results in a smaller range of test scores to base responses on. As stated by Lezak (1995), "single tests typically assess only one brain function and therefore may not be relevant to a specific patient's complaints" (p. 792). For example, the person who

malingers a memory deficit will likely perform adequately on what she/he perceives as a measure of visuospatial functioning. Another reason for including multiple tests is that it is probably easier to fake on one measure than it is to fake on a variety of measures administered over the course of an assessment. It is possible that performance on a single measure will not be the same when that measure is included in the context of other tests, thereby limiting generalization of test results. That is, inclusion of multiple measures may influence how simulators approach a particular test. Incorporating many tests within a simulation study also provides converging evidence of possible malingering performance.

Researchers are often limited by the time constraints of participants. However, it is possible to include several measures that more closely approximates a neuropsychological assessment without unduly prolonging the test sessions. For example, a study that takes between 2-3 hours to complete is not unreasonable, and it offers more valuable information than what is obtained by using a single measure (e.g., Bernard, 1990).

Incentive

Another main criticism of simulation research is that incentives offered to simulators are generally not comparable to those that a successful malingerer would achieve (e.g., compensation through litigation). Also, the penalty for detection is often lacking in simulation research. It has often been proposed that studies should more closely approximate what occurs in the medicolegal context. However, in terms of offering a comparable incentive, it is difficult to imagine how one would

realistically approximate the possible gains from litigation. A possible alternative is to offer a differential reward. Most studies offer course credit as an incentive to volunteer for the study. A more adequate incentive would be to also include a differential monetary reward to the one who can simulate performance the best.

Intuitively one would expect that the greater the reward in simulation research the better the incentive for simulation. However, studies that have examined the role of the incentive on simulation performance have reported no effect of incentive (Bernard, 1990; Martin et al., 1993). Although it would be tempting to say that incentive has no role on motivation to simulate, further research needs to be conducted to examine this question more closely. Offering differential rewards to students who are participating to fulfil course requirement credits may not be an added incentive. Research needs to address whether a differential reward influences voluntary participation in research (e.g., non-student participation).

Preparation Time

Finally, few studies have allotted time for the participants to prepare for the simulation task. That is, subjects are given a scenario to read outlining the parameters of the experiment and are then asked to begin. No study reviewed in the literature has given simulators adequate time to prepare for their task. Although it is unclear if increasing preparation time would result in better simulation performance, it would more closely approximate what occurs in the clinical realm. That is, malingerers often have a sufficient time to prepare for the assessment and have given thought to how they will perform during testing.

IV. Research Studies

As shown in the preceding literature review on malingering, qualitative indices of test performance have been difficult to objectively verify and do not occur frequently or consistently. Existing tests of malingering have been found to be affected by non-motivational factors (e.g., cognitive impairment, age, education) and have limited face validity as tests of learning/memory. Based on the previous review, no completely adequate measure of malingering is currently available.

Simulation research on malingering has been criticized primarily because the findings do not readily generalize to a clinical setting. This is largely because simulation research has relied primarily upon asking students to simulate cognitive impairment. Moreover, they usually have been given only one test, have had little time to prepare and, have had little incentive to perform in a realistic manner. Also, simulation research has not addressed the type of strategies subjects have used to simulate cognitive impairment.

In view of these findings, there is a need for an empirically demonstrated test that will meet the criteria for a clinically useful test of malingering. Preliminary research with the TOMM proved promising. Data from this test showed that it was viewed as a legitimate measure of learning/memory, and it was not influenced by the effects of age or education (Hubley, McIntyre & Tombaugh, 1995). It was also found to be perceived as more difficult than it actually was. However, in order to be considered a clinically useful measure of malingering, the TOMM needs to demonstrate that it can differentiate between malingerers and non-malingerers.

Performance on the TOMM must also be shown to be insensitive to various types of cognitive impairment.

In order to validate the TOMM four integrated studies were conducted. The first two studies used simulation designs. Suggestions to overcome possible limitations of simulation research were incorporated into both designs and strategy use was evaluated. Study 1 consisted of university students and Study 2 used traumatically brain injured patients. The third study examined performance on the TOMM across a selected group of patient and control subjects, including individuals suspected of malingering. Study 4 examined performance on the TOMM across a diverse clinical sample and forms a comprehensive normative data base.

Study 1

University Simulation Study

As previously outlined, a clinically useful measure of malingering must meet four major criteria: (1) the test must differentiate between those trying their best and those not trying their best (i.e., between malingerers and non-malingerers), (2) the perceived difficulty of the test must outweigh actual difficulty, (3) the test must be perceived as a legitimate measure of neuropsychological functioning (i.e., have face validity), and (4) the test must be insensitive to non-motivational factors, such as cognitive ability and/or impairment, age or education.

The purpose of the first study was to determine if the TOMM could differentiate between university students asked to simulate the performance of head injured patients and students asked to perform to the best of their ability. Based on the previous literature, it was expected that simulators would score substantially lower on the TOMM than control subjects, although they would not score below chance performance.

In order to evaluate whether the perceived difficulty of the test outweighed actual difficulty, estimation scores were compared to obtained scores across all trials of the TOMM. If the test is initially perceived by all students as being more difficult than it actually is, estimation scores on Trial 1 should be equivalent between simulators and controls. On Trial 2, estimation scores should substantially increase for control subjects in relation to their performance on Trial 1, and compared to

simulators performance on Trial 2. This would indicate that control subjects realized after Trial 1 that it was a relatively easy test to successfully complete. Simulators may increase their estimate scores on Trial 2, but not in a substantial manner.

Face validity of the TOMM was evaluated through use of rating scales. It was hypothesized that the TOMM would not be identified as a test of malingering due to the nature of the task demands (remember 50 pictures).

Rating scales were used to explore whether simulators could identify which tests were malingering tests and how convincing they thought they were in their attempts at deception. Simulators were also verbally asked how they lowered their performance on each test given. That is, what their strategy was for simulating on each test.

The abbreviated Hiscock and the Basic Knowledge Questionnaire, both tests of malingering, were included in the study for comparison purposes. It was hypothesized that the abbreviated Hiscock would differentiate between simulators and controls and that simulators' performance would deteriorate as the length of the intervals increased. However, it was expected that due to the simplicity of the test, it would be identified as a test of malingering. Performance on the Basic Knowledge Questionnaire was expected to be equivalent between groups and that it would be identified as a test of malingering due to the obvious nature of the questions.

Study 1 also incorporated suggestions to overcome the previously mentioned criticisms of simulation research. That is, the instructional set emphasized that simulation of cognitive deficits needed to be realistic, simulators were given one week

to prepare for the study, a substantial differential reward was offered, tests of malingering were embedded into a neuropsychological battery of tests, and an extensive debriefing was carried out where compliance and strategy use was evaluated.

Method

Participants

Forty-four Introductory Psychology students were randomly assigned to one of two groups: *Simulation Group* ($n=25$) or *Control Group* ($n=19$). Mean age of *Simulators* was 21.5 years ($SD=2.3$), with 17 males and 8 females. Mean age of the *Control Group* was 21.8 years ($SD=4.3$), with 12 males and 7 females. Each participant received course credit for participation.

Materials

Materials consisted of commercially available neuropsychological tests that were selected on the basis of four cognitive domains (memory, attention, executive function and general ability) and malingering. Memory (learning and retention) was assessed by the California Verbal Learning Test (CVLT: Delis, Kramer, Kaplan & Ober, 1987) and the Visual Reproduction subtest of the Wechsler Memory Scale-Revised (WMS-R: Wechsler, 1987). Attention was assessed by the Mental Control and the Digit Span Forward subtests of the WMS-R, and the Trail Making Test, part A (Lezak, 1995). Executive functioning was measured by the Trail Making Test, part B (Lezak, 1995), the Digit Span Backward subtest of the WMS-R, Ruff Figural Fluency Test (RFFT: Ruff, 1988), and Controlled Oral Word Association (FAS &

Animals: Lezak, 1995). General ability was measured by the Information subtest from the Wechsler Adult Intelligence Scale-Revised (WAIS-R: Wechsler, 1981) and the Finger Tapping Test (Lezak, 1995). Tests designed to detect malingering included the Test of Memory Malingering (TOMM: Tombaugh, 1996), the Basic Knowledge Questionnaire (Knight, 1988) and the abbreviated Hiscock (Guilmette et al., 1993). Appendix A contains a complete description of each test. The order of test administration is shown in Table 1.

 Insert Table 1 about here

Exploratory rating scales were developed to assess subjects' compliance with the experimental condition and to determine whether they perceived themselves to have been successful in their deception. Additional scales were developed to address how the subjects approached the tests, to determine strategies used, and to assess whether they could correctly identify tests that were specifically designed to detect simulation (see Appendix B)².

Procedure

Participants were randomly assigned into one of two groups (*Simulation* or *Control*). Demographic information (e.g., age, education) and medical history were obtained from all participants. Those participants with a medical history suggestive of

²A special note of appreciation to J. Knight for his assistance in developing these scales.

Table 1
Neuropsychological Tests and Order of
Test Administration

Cognitive Domain	Order of Presentation
Attention	
Mental Control	1
Trails-A	4
Digit Span Forward	7
Memory	
California Verbal Learning Test	
Learning	2
Retention	9
Visual Reproduction of WMS-R	
Learning	3
Retention	10
Executive Function	
Trails-B	5
Digit Span Backward	8
Ruff Visual Fluency	13
Verbal Fluency	14
General Abilities	
Finger Tapping	6
Information WAIS-R	12
Malingering	
TOMM	
Learning	11
Retention	15
Abbreviated Hiscock	16
Basic Knowledge Questionnaire	17

central nervous system impairment were excluded from data analysis (e.g., those with a prior history of head injury, alcohol abuse, current clinical depression, etc.). Those assigned to the *Simulation Group* met with the examiner one week before testing.

During that time, the nature of the experiment was described. *Simulators* were given the following brief scenario describing an individual who was seeking compensation for a head injury caused by a motor vehicle accident. The subjects were asked to *realistically* simulate the performance of this individual on neuropsychological tests. However, they were not given any specific instructions on how this should be accomplished. Volunteers were told that they would have one week to prepare, and they could use any resource available to them (e.g., books, lectures, videos).

Scenario

In this study you will be asked to complete a set of tasks that are often used to measure a variety of changes that occur in people who have brain-damage.

As you take each test, we would like you to assume the role of someone who has experienced some brain-damage from a car accident.

Pretend that you were involved in a head-on collision. You hit your head against the windshield and were unconscious for 15 minutes. You were hospitalized over night for observation and then released. Gradually, over the past few months, you have started to feel normal again. However, your lawyer has informed you that you may get a larger settlement from the court if you look like you are still suffering from brain-damage. In the real world, the usual purpose of the tests you are about to take is to determine if the accident has produced any impairments in your abilities due to brain damage.

As you portray the above person, try to approach each test as you imagine this person would respond if they had been given the same instructions from their lawyer or someone else hoping to influence the amount of the settlement. Try to create responses on the tests that will convince the examiner that you are truly brain-damaged, keeping in mind that settlement monies depend upon your being diagnosed as cognitively impaired on these tests. *Also be aware that having a lawsuit pending often raises the suspicion that people may try to*

exaggerate their difficulties. That means your impairments resulting from the head injury must be believable. Major exaggerations, such as not being able to do anything, remembering absolutely nothing, or completely failing to respond, are easy to detect.

As an incentive, we are offering \$50.00 to the person who produces the most convincing profile of test scores as a head-injured individual. Determination of who gets the money is based on test scores alone, and **NOT** on your behaviour during the test session.

If you have any questions about the instructions or the study, please ask them now.

Participants returned one week later to be tested. Before testing, informed consent was obtained. Participants were asked to reread the scenario and then to demonstrate they fully understood what was expected (i.e., they were asked to paraphrase the instructions to the examiner). Neuropsychological tests previously discussed were administered in the order shown in Table 1. Testing took approximately two hours. At the end of testing, participants were asked to answer a series of rating scales in an honest manner. Debriefing followed and any questions were addressed.

The procedure was the same for the *Control Group* with the exception that subjects were asked to perform to the best of their ability. Participants in the *Control Group* did not receive the scenario. Testing took approximately 1.5 hours. The volunteers were debriefed as to the nature of the study, and then were then asked to rate each test as to whether they felt it would have been easy or difficult to fake and whether it was designed specifically to detect possible malingerers (Rating Scales 5 & 6).

At the end of the study, test performance of the *Simulators* was examined to

determine the one whose profile most closely corresponded to what is typically seen in a head injured, cognitively impaired population (see Figure 1). That person received the \$50.00 prize. For the *Control Group*, a random draw for \$50.00 was made.

 Insert Figure 1 about here

Results and Discussion

For clarity of presentation, data analyses are presented separately for the malingering tests, neuropsychological tests, and rating scales.

Tests of Malingering

TOMM: Mean performance of obtained scores is presented in Table 2. Inspection of the table shows that scores for Trial 1, Trial 2, and the Retention Trial are markedly different for the two groups, with *Controls* scoring higher than the *Simulators*.

 Insert Table 2 about here

A repeated measures multivariate analysis of variance on the TOMM revealed a significant Group and Trial effect ($F(1,42) = 68.93, p < .0001$; $F(2,84) = 12.77, p < .0001$, respectively). The Group x Trial interaction was not significant ($p > .05$). One-way ANOVAs revealed that the *Simulation Group* scored significantly lower than

Figure 1. Expected Group Performance on Neuropsychological Tests

Table 2
Group Means and Standard Deviations for Obtained
and Estimated Performance on the TOMM

Measures	Group	
	Simulation	Control
Obtained		
TOMM 1	32.6 (7.8)	48.8 (1.7)
TOMM 2	35.7 (8.9)	49.9 (0.2)
Retention	34.3 (8.2)	50.0 (0.0)
Estimated		
TOMM 1	24.9 (9.6)	33.4 (9.1)
TOMM 2	36.2 (9.8)	44.0 (7.6)
Retention	31.0 (10.6)	44.1 (7.6)

the *Control Group* on each trial of the TOMM (TOMM 1: $F(1,43) = 78.71$, $p < .0001$; TOMM 2: $F(1,43) = 48.41$, $p < .0001$; Retention Trial: $F(1,43) = 69.72$, $p < .0001$).

Examining the distribution of scores for each group on each trial (see Table 3) revealed that by Trial 2, all subjects in the *Control Group* received a score of 49 or 50 correct. Thus, using a score of 49 as the cutoff score results in 100% specificity (i.e., percentage of control subjects correctly identified). In comparison only 2 simulating subjects (8%) received a score of 49 or 50 (92% sensitivity)(i.e., percentage of simulating subjects correctly identified). Using the more conservative cutoff score of 45 (90% correct responding) suggested by Tombaugh (1996), shows that specificity remains at 100%, while sensitivity decreases to 84%. That is, all *Control* subjects are correctly identified while 21 of the 25 *Simulators* are correctly identified. On the Retention Trial, using a cutoff score of 45, specificity remains at 100% and sensitivity increases to 88%.

Insert Table 3 about here

The distribution of scores was also analyzed according to chance levels of responding. Chance performance on the TOMM is a score of 25. Based on the binomial equation (Siegal, 1956) and the 95% confidence interval, a score of 19 is significantly below chance performance. On Trial 2, 12% of the *Simulators* received a score of 25 or lower (including 8% who scored below 19). On the Retention Trial,

Table 3
Percent of Subjects Obtaining Different
Scores on the TOMM

Groups	Number of Correct Responses							
	50	49	48	47	46	45	44-40	<40
Simulation								
Trial 1			4%			4%	16%	76%
Trial 2	4%	4%		4%	4%	4%	12%	68%
Retention		4%		8%			16%	72%
Control								
Trial 1	37%	42%	5%	11%		5%		
Trial 2	95%	5%						
Retention	100%							

12% scored at or below 25 (including 4% who scored less than 19). In contrast, no subject in the *Control Group* received a score below 49 on Trial 2 of the TOMM.

As can be seen by the above description, using cutoff scores (either 49 or 45) results in better classification than relying on chance level performance only. That is, 12% of *Simulators* were identified using chance level of responding compared to 84% using a conservative cutoff score of 45.

Mean performance of estimation scores is presented in Table 2. As shown in the table, estimation scores are similar between groups on Trial 1. On Trial 2 and on the Retention Trial, *Simulators* have lower scores compared to the *Control* subjects.

Estimation scores for each trial of the TOMM were analyzed using a repeated measures multivariate analysis of variance. Results revealed significant Group, Trial, and Group x Trial effects (Group: $F(1,42) = 11.40, p < .002$; Trial: $F(1,42) = 30.74, p < .0001$; Group x Trial: $F(1,42) = 8.15, p < .001$).

One-way ANOVAs of estimation scores on each trial revealed no significant difference between *Simulators* and *Controls* on Trial 1 ($F(1,43) = 1.92, p > .05$). However, estimation scores on Trial 2 and on the Retention Trial revealed that *Simulators* estimated their performance to be significantly lower than *Controls* (Estimate TOMM 2: $F(1,43) = 9.04, p < .01$; Estimate Retention: $F(1,43) = 20.73, p < .001$).

The finding that estimation scores were equivalent between groups on Trial 1 highlights an important criterion for a test of malingering. That is, the perceived difficulty of the test must outweigh the actual difficulty. If estimation of performance

reflects perception of test difficulty, both groups perceived the degree of difficulty of the test in a similar manner. After exposure and feedback (Trial 1), *Control* subjects adjusted their estimate scores upwards, reflecting that the test was no longer perceived to be as difficult as initially perceived. Although *Simulators* adjusted their estimate upwards on Trial 2, scores were significantly lower than the *Control* group's estimation. Such findings indicate that *Simulators* artificially kept their scores low, even though exposure to the test would reveal that the test is not as difficult as initially perceived.

In summary, performance on the TOMM revealed (1) that it has good sensitivity and specificity and (2) that it has good face validity due to the perceived difficulty of the test.

The Abbreviated Hiscock: Examining mean scores on the abbreviated Hiscock (see Table 4) reveals that *Simulators'* performance progressively declined as the length of the interval increased. In contrast, *Control* subjects maintained perfect performance across all three intervals. Subsequent analysis confirmed these observations.

 Insert Table 4 about here

A repeated measures multivariate analysis of variance on all trials of the abbreviated Hiscock revealed significant Group ($F(1,42) = 32.59, p < .0001$), Trial ($F(2,84) = 17.30, p < .0001$) and Group x Trial effects ($F(2,84) = 17.30, p < .0001$).

Table 4
Group Mean and Standard Deviations on the
Abbreviated Hiscock (A-Hiscock)

Measure	Group	
	Simulation	Control
A-Hiscock--5	10.8 (1.7)	12.0 (0.0)
A-Hiscock--10	9.4 (1.9)	12.0 (0.0)
A-Hiscock--15	8.4 (2.4)	12.0 (0.0)
A-Hiscock--Total	28.5 (5.4)	36.0 (0.0)

Subsequent one-way ANOVAs revealed that the *Simulation Group* scored significantly lower than the *Control Group* on all three trials of the abbreviated Hiscock (abbreviated Hiscock 5: $F(1,43) = 9.08, p < .005$; abbreviated Hiscock 10: $F(1,43) = 30.74, p < .001$; abbreviated Hiscock 15: $F(1,43) = 37.83, p < .001$).

These findings are consistent with the hypotheses outlined, as well as with other research findings (e.g., Guilmette, Hart, Giuliano & Leininger, 1994; Guilmette, Sparadeo, Whelihan & Buongiorno, 1994). The recommended cutoff score of 33 (Guilmette, Hart et al., 1994) results in 100% specificity and 76% sensitivity, indicating that the abbreviated Hiscock differentiates between *Simulators* and *Control* subjects with a respectable degree of accuracy.

The distribution of scores was also examined according to chance level of responding. Chance performance on the abbreviated Hiscock is calculated on the sum of all trials (total = 36), therefore chance is 18. Based on a 95% confidence interval, a score of 13 is significantly below chance performance. Scores ranged from 19 to 36. No subject in the *Simulation Group* scored below chance level. As with the TOMM, the results show that it is better to use a cutoff score than to rely on chance level responding in identifying possible malingerers.

Basic Knowledge Questionnaire: Examining mean scores on the Basic Knowledge Questionnaire (Table 5) suggests that *Simulators* and *Control* subjects were similar in their responses to the questionnaire (one point difference). However, a one-way ANOVA revealed that performance of the *Simulation Group* was significantly lower than the *Control Group* ($F(1,43) = 7.46, p < .01$).

 Insert Table 5 about here

It was hypothesized that performance by the two groups would be equivalent, as it was thought that the Basic Knowledge Questionnaire was an obvious measure of malingering and would be recognized as such. However, findings did suggest that performance of the groups was sufficiently different, although the clinical usefulness of this test remains to be demonstrated.

Correlations: A correlational analyses of all malingering tests is presented in Table 6. All tests are significantly correlated ($p < .0001$). Examination of the table reveals a stronger relationship between the abbreviated Hiscock and the TOMM than between these two tests and the Basic Knowledge Questionnaire. Thus, indicating that the utility of the Basic Knowledge Questionnaire is questionable.

 Insert Table 6 about here

Neuropsychological Tests

Mean percentile scores for each test are presented in Table 7. Examination of Table 7 reveals that *Simulators'* performance on neuropsychological tests is markedly lower than *Control* subjects' performance. A multivariate analysis of variance (MANOVA) was significant. Subsequent univariate analyses showed that *Simulators* performed significantly lower than *Controls* on each neuropsychological test.

Table 5
Group Means and Standard Deviations on the Basic
Knowledge Questionnaire

Measure	Group	
	Simulation	Control
Basic Knowledge Questionnaire	16.2 (2.6)	17.9 (0.5)

Table 6
Correlations Between Different Tests of Malingering

Tests	Abbreviated Hiscock	Basic Knowledge Questionnaire
Basic Knowledge Questionnaire	.70	
TOMM 1	.78	.59
TOMM 2	.76	.65
TOMM Retention	.82	.60

 Insert Table 7 about here

As stated in the introduction, there are a number of neuropsychological indicators of malingering, including (1) inconsistencies between *Simulators'* test performance and what is typically observed following a traumatic brain injury (TBI), and (2) inconsistencies between and within tests measuring the same construct.

Although TBI patients show slightly different strengths and weaknesses in terms of neuropsychological functioning, some generalities exist. Given the facts presented in the scenario (LOC < 20 minutes, recovering) the level of severity of the head injury would be considered mild (see Figure 1). Typically, those with mild TBI suffer from slowed or inefficient information processing which may show up as attentional deficits, including poor concentration, heightened distractibility, difficulty doing more than one thing at a time, and complaints of short term memory failure (reduced acquisition and recall of new information). Self-monitoring and verbal retrieval may also be reduced (Lezak, 1995, p.182).

In terms of tests administered, *Simulators'* performance is inconsistent with what is typically seen following mild TBI. *Simulators* showed excessive impairment on the California Verbal Learning Test (a measure of learning and retention), on the Finger Tapping Test (a measure of manual dexterity), and on the FAS (a measure of verbal fluency). Performance on Visual Reproduction, a measure of visual learning

Table 7
Group Means for Percentile Scores
on Neuropsychological Tests

Measure	Group	
	Simulation	Control
Learning and Retention		
CVLT		
Learn	< 1st	16th
Retention	< 1st	16th
Visual Reproduction		
Learn	19th	78th
Retention	15th	70th
Attention		
Mental Control	16th	50th
Trail Making-A	17th	72nd
Digit Span Forward	10th	56th
General Ability		
Finger Tapping		
Preferred Hand	<1st	30th
Nonpreferred Hand	<1st	35th
Information	16th	50th
Executive Functioning		
Trail Making-B	20th	59th
Digit Span Backward	19th	63rd
FAS	< 1st	42nd
Animal Naming	16th	50th

and retention, was lower than expected, but not markedly so. Mental Control was also reduced. Other measures of attention (Trail Making Test part A and Digit Span forward) were within normal limits, as were measures of executive functioning, with the exception of the Ruff Figural Fluency Test (e.g., Trail Making Test part B, Digit Span backward, Animal Naming). On the RFFT, *Simulators* performance was excessive compared to a mild TBI sample. Performance on the Information subtest, a measure of general knowledge, was borderline. That is, it was reduced more than expected, although not significantly so.

Simulators' performance was also inconsistent in regards to recognition versus recall procedures. That is, *Simulators* did not show significant improvement during recognition procedures compared to free recall procedures. *Simulators* also showed significant deterioration over delayed recall which is not consistent with mild TBI.

In summary, although *Simulators* complied with the instructions, they were unable to realistically approximate the cognitive impairment following a mild TBI. Most responded at a more impaired level than expected, particularly on tests of verbal retention and manual dexterity. Perhaps in future studies one way to overcome such confusion is to explicitly state the level of severity of the brain injury in the scenario. However, given that *Simulators* had one week to prepare for the evaluation, it is not likely that specifying the level of severity will overcome this difficulty.

Overall, these findings were consistent with the literature (e.g., Beetar & Williams, 1995; Bernard et al., 1993; Franzen et al., 1990; Paul et al., 1992). The literature generally reports that *Simulators* lower their performance across all

measures, which is typically inconsistent with the neuropsychological profile they are attempting to simulate. Such findings emphasize the unsophisticated nature of undergraduate students when attempting to simulate. That is, they do not have a good understanding of the nature of the disorder they are trying to mimic. Such results point to a main criticism of malingering research: relying on simulation data from undergraduate students may be misleading when malingering tests are interpreted in the context of a clinical situation. Such findings illustrate the importance of examining how bonafide TBI patients simulate performance (i.e., exaggerate deficits in a realistic manner).

All Tests

A stepwise discriminant function analysis (DFA) revealed that of all tests used, including tests of malingering, performance on Trial 1 of the TOMM was a better discriminator of group membership than any other variable. Considering all tests together, five tests provided the best linear combination for predicting group membership: TOMM 1, Trail Making Test part A, free recall intrusions and performance on List B of the California Verbal Learning Test, and the retention trial of Visual Reproduction. In terms of classification, based on the selected discriminant function, sensitivity was found to be 96% and specificity was 100%. Overall, 98% of the cases were correctly classified.

Rating Scales

The majority of *Simulators* (72%) indicated that they tried moderately hard to follow the simulation instructions (Rating Scale 1). One subject indicated that she put

forth minimal effort and was subsequently removed from all data analyses. A majority of *Simulators* (64%) also felt they were successful in their attempts at deceiving the examiner (Rating Scale 2). Sixty-four percent indicated that they had no background knowledge of the cognitive effects of brain injury, and they did not use the week prior to testing to increase their knowledge (Rating Scale 3).

Participants were also asked to endorse how convincing they thought they were in their attempt at deception on each test (Rating Scale 4). The scale ranged from not convincing at all (0%) to very convincing (100%). The frequency distribution of responses showed that, on average, subjects did not feel strongly one way or the other about how convincing they were in their attempts at deception ($M=49.9\%$, $SD=25.0$). Basically, subjects selected a neutral position, even though the majority felt they were successful in their attempt at deception.

Rating Scales 5 and 6 were given to participants in both the *Simulation* and *Control Groups*. Scale 5 addressed the "fakeability" of the tests. That is, the degree to which subjects thought they could fake the test without detection. Four categories were given (1 = no obvious way to fake test, 2 = test too simple to fake, 3 = test too difficult to fake, and 4 = obvious way to fake the test). Mean scores reflect that approximately half of the subjects thought the tests were too simple to fake and the other half felt that the tests were too difficult to fake (*Simulators*: $M=2.5$, $SD=1.2$; *Controls*: $M=2.5$, $SD=1.0$).

Scale 6 attempted to address whether subjects could identify the true intent of the tests (i.e., whether they could differentiate between malingering and non-

malinger tests). Scores ranged from 1 to 10 (1 = test identified correctly, 10 = misperception of nature of test). Overall, mean performance for *Simulators* across all tests was 4.8 ($SD=3.2$). Performance by the *Control* subjects was similar ($M=4.6$, $SD=3.0$). No consistent trend emerged and no test was strongly endorsed as an obvious measure of malingering.

Although it is difficult to interpret the findings from the different rating scales, particularly given the large variability of scores, it is interesting to note that participants in both groups had a similar outlook regarding the fakeability of the tests (Scale 5) and the identification of the true intent of the tests (Scale 6). Thus, group membership did not seem to influence subject's perception of the tests.

The final scale given produced the most useful data. On this scale, subjects were asked to verbally report what strategy they used to malingering for each test (see Appendix C). Results were analyzed in a qualitative manner and revealed the use of seven strategies. Each strategy is listed below followed in parentheses by the test where that strategy was most prominent.

1. Reduce speed of performance on timed tests (Mental Control, Trail Making Test, Finger Tapping Test, Ruff Figural Fluency Test).
2. Repetition of items or responses (FAS, Ruff Figural Fluency Test).
3. Change in the sequence of items (Trail Making Test, Digit Span).
4. Report "I don't know" answers (Information subtest).
5. Randomly omit items (California Verbal Learning Test, Digit Span).
6. Report false, yet related items (errors of intrusion on California Verbal Learning Test).
7. Change the details of the designs on Visual Reproductions.

Overall, these strategies appeared to be instruction-dependent. That is, subject's based their main strategy on information that was contained in the test

instructions. For example, if it was stated that the subject should proceed as quickly as possible, she/he slowed their performance down. If instructed to state different items, she/he repeated items.

On tests of malingering (TOMM, abbreviated Hiscock, and the Basic Knowledge Questionnaire) subjects commented that the abbreviated Hiscock and the BKQ were too easy to fake, although almost all reported doing so. On the TOMM, no one reported that it was too easy to fake, although a few were surprised at how easy it would be to do well on.

In summary, the University Simulation Study showed that the TOMM is a useful measure for discriminating between *Simulators* and *Control* subjects. The TOMM meets three of the four criteria for a good test of malingering: (1) the TOMM accurately differentiated the *Simulators* from the *Control* subjects, (2) given that no one identified the TOMM as a measure of malingering, it has face-validity as a neuropsychological measure of memory and, (3) based on findings from the estimate scores, the perceived difficulty of the test outweighs the actual difficulty.

In contrast, the abbreviated Hiscock was perceived to be too easy to fake, although empirically it did differentiate between *Simulators* and *Controls*. Performance on the abbreviated Hiscock highlights the necessity of including a debriefing regarding subject's strategy use. That is, even though subjects complied with simulation instructions and faked performance on the abbreviated Hiscock, the majority commented that it was too easy to fake.

Results of this study also highlighted the advisability of including a clinical

group in simulation research. University students were unable to realistically simulate the cognitive deficits seen following a mild traumatic brain injury. By including a group of traumatically brain injured patients asked to intentionally lower their performance, researchers may be able to better approximate or generalize their findings to the clinical situation.

Study 2

TBI Simulation Study

One possible criticism of Study 1 is that it did not employ participants who were knowledgeable about traumatic brain injury. Thus, we do not know if they were able to simulate performance of TBI individuals in a realistic manner. The purpose of Study 2, therefore, was to evaluate how traumatically brain injured people, asked to simulate or exaggerate possible cognitive deficits, perform on the TOMM and other neuropsychological measures. Study 2 is unique as no other published study has asked traumatically brain injured patients to simulate performance. Given that this group forms the group often suspected of malingering, findings regarding their performance will be clinically relevant and will be extremely useful in terms simulation research.

As with Study 1, performance on the TOMM was evaluated in terms of whether it meets the criteria of a clinically useful measure of malingering. It was hypothesized that TBI simulators would perform at substantially reduced levels on the TOMM compared to TBI controls, although not below chance performance. It was also expected that TBI controls would perform at perfect or near-perfect levels on the TOMM, indicating that the TOMM is not influenced by cognitive impairment.

Strategy use and compliance were also evaluated using the same procedures outlined in Study 1.

Method

Participants

Forty-one subjects were divided into two groups: a *TBI Control Group* ($n=25$) and a *TBI Simulating Group* ($n=16$). Mean age of the *TBI Control Group* was 40 years ($SD=14.0$), with 15 males and 10 females. Mean education was 13.6 years ($SD=2.7$). Mean age of the *TBI Simulators* was 44.9 years ($SD=11.2$), with 9 males and 7 females. Mean education was 12.8 years ($SD=2.3$).

Severity of injury was measured by the length of unconsciousness (see Table 8) (Lezak, 1995, p.755). Three categories were employed (1 = mild, 2 = moderate, and 3 = severe). The *TBI Control Group* contained 10 individuals with mild head injuries, 1 with a moderate injury, and 14 with severe injuries ($M=2.2$, $SD=1.0$). The *TBI Simulating Group* had 6 individuals with mild injuries, 2 with moderate injuries and 8 with severe injuries ($M=2.1$, $SD=1.0$). Groups did not differ with respect to age ($p > .05$), education ($p > .05$) or severity of injury ($p > .05$).

Insert Table 8 about here

Subjects were recruited through advertisements placed in the Ottawa Citizen, Carleton University Alumni Magazine, Ottawa-Carleton Head Injury Association Bulletin, and on Roger's Community T.V. Channel. Subjects were also recruited from St. Vincent's Pavilion, the Smiths Falls Brain Injury Association, and from a local neuropsychologist's private practice.

Table 8
Classification of Severity of Head Injury

Classification	Length of Unconsciousness
Mild	<20 minutes
Moderate	>20 minutes but <6 hours
Severe	>6 hours

Materials and Procedure

The following procedure was used for the community-dwelling, head injured individuals assigned to the *TBI Control Group*. Interested volunteers were asked to call the investigator at Carleton University to obtain information about the study (i.e., that we are looking at how individuals who have sustained a head injury perform on various measures of learning and memory). If they agreed to participate, volunteers were requested to come to the University at a mutually agreed upon time. Participants were then asked to read and sign an informed consent. They were then asked to sign a medical release form (Form 14) so that the investigator could access their medical records pertaining to the head-injury (e.g., radiological and/or neuropsychological reports). Next, participants were asked questions about their head injury (e.g., description of the event, length of coma, various cognitive and/or physical problems following the injury) using the Philadelphia Head Injury Questionnaire. Following this, testing began and proceeded in the same order as outlined in Study 1 with the following exception: the Beck Depression Inventory (Beck, 1987), a screening test for possible depression, replaced the Ruff Figural Fluency Test. All tests followed standard administration procedures. Testing took approximately 2-3 hours to complete. Following completion of the study, participants were debriefed and questions were answered. Other than parking, no financial remuneration was provided.

For hospital-based patients, the procedure was modified. Once ethical approval for the research study was obtained from the hospital, a staff member

(Psychologist) of the facility made initial contact. Of those who agreed to participate, their name was given to the examiner who contacted them and set up an appointment to meet them at the hospital. The participants signed an informed consent to participate and a medical release form (Form 14) to permit review of their charts for pertinent information. Testing took between 30 and 90 minutes depending on whether subjects had previously undergone a neuropsychological evaluation. If subjects had undergone neuropsychological testing within the past year and the data were available, participants were administered only the Mini Mental Status Exam (MMSE: Folstein, Folstein & McHugh, 1975), TOMM, abbreviated Hiscock and, the Beck Depression Inventory. If participants had not previously undergone a neuropsychological evaluation, they also completed the California Verbal Learning Test and the Visual Reproductions subtest of the WMS-R.

Subjects in the *TBI Simulating Group* were recruited following the same procedure outlined for the community-dwelling TBI subjects in the *Control Group*. Interested volunteers were asked to contact the examiner and the study was described to them (i.e., that they would be given various tests of learning and memory and would be asked to intentionally lower their performance). If they agreed to participate and were not currently involved in litigation, the scenario used in Study 1 was mailed to them and an appointment was scheduled for approximately two weeks later. Before testing, informed consent was obtained. Participants were asked to reread the scenario and then to demonstrate they fully understood what was expected (i.e., they were asked to paraphrase the instructions to the examiner). *TBI Simulators*

received the same neuropsychological tests as the *TBI Control* subjects with one exception. Subjects did not receive the Beck Depression Inventory. Testing was conducted at Carleton University or at the Smiths Falls Brain Injury Association offices in Smiths Falls and took approximately 2-3 hours. At the end of testing, participants were asked to respond to three rating scales in an honest manner (Rating Scales 4 & 6 and oral debriefing regarding strategy use). Debriefing followed and questions were addressed.

Results and Discussion

Tests of Malingering

TOMM: Mean performance of obtained scores are presented in Table 9.

Inspection of the table shows that scores for Trial 1, Trial 2, and the Retention Trial are markedly higher for the *Control Group* than for the *Simulation Group*.

Subsequent analysis confirmed this observation.

Insert Table 9 about here

A repeated measures multivariate analysis of variance on the TOMM revealed a significant Group and Trial effect ($F(1,39) = 30.03, p < .0001$; $F(2,78) = 22.55, p < .0001$, respectively). The Group x Trial interaction was not significant ($p > .05$). One-way ANOVA's revealed that the *TBI Simulation Group* scored significantly lower than the *TBI Control Group* on each trial of the TOMM (TOMM 1: $F(1,39) = 28.88, p < .0001$; TOMM 2: $F(1,39) = 25.35, p < .0001$; Retention Trial: $F(1,39) = 27.39,$

Table 9
Group Means and Standard Deviations for Obtained
and Estimated Performance on the TOMM

Measure	Group		
	Simulation TBI	Compliant TBI	Control TBI
<u>Obtained</u>			
TOMM 1	35.1 (10.3)	28.1 (7.8)	47.1 (3.6)
TOMM 2	39.4 (10.1)	32.1 (7.3)	49.6 (1.3)
Retention	38.8 (10.3)	31.6 (8.0)	49.6 (1.0)
<u>Estimated</u>			
TOMM 1	31.1 (14.2)	31.4 (18.0)	34.9 (13.2)
TOMM 2	35.8 (13.5)	33.6 (15.9)	42.3 (11.2)
Retention	37.7 (13.3)	32.1 (16.0)	44.9 (8.8)

$p < .0001$).

Examining the distribution of scores for each group on each trial (see Table 10) revealed that by Trial 2 only one subject in the *Control Group* received a score lower than 48. If these data are used to establish a cutoff score of 48, 96% of the *Control Group* are correctly identified (i.e., 96% specificity). In contrast, six *Simulators* scored above 48 on Trial 2 (63% sensitivity). The more conservative cutoff score of 45 suggested by Tombaugh (1996) results in 96% specificity and 56% sensitivity. On the Retention Trial, specificity and sensitivity increase to 100% and 63%, respectively.

 Insert Table 10 about here

Examining the data reveals that almost half of the *Simulation* sample ($n=7$) scored above the recommended cutoff score of 45 on Trial 2. That is, a bimodal distribution occurred with approximately half of the subjects scoring below and half scoring above the cutoff score of 45. Examining the questionnaire data of the *Simulation Group* revealed that a number of individuals ($n=8$) had reported that they had tried their best throughout testing (i.e., that they did not simulate performance). Given that these subjects (*Non-compliant Simulators*) indicated they were unable to intentionally lower their performance, their data was removed from the analysis. Analysis on the data of the *Compliant Simulators* resulted in 100% sensitivity when using a cutoff score of 45. That is, a cutoff score of 45 correctly classified all the

Table 10
Percent of Subjects Obtaining Different
Scores on the TOMM

[illegible]

Compliant TBI Simulators on all three trials of the TOMM.

A statistical reanalysis of performance across all trials of the TOMM with data from only the *Compliant Simulators* revealed similar findings to those originally reported with the entire sample. For example, a repeated measures multivariate analysis revealed significant Group and Trial effects ($F(1,31) = 48.14, p < .0001$; $F(2,62) = 13.39, p < .0001$, respectively). The Group $\times$ Trial interaction was not significant ($p > .05$). One-way ANOVA's revealed that the *Compliant TBI Simulators* scored significantly lower than the *TBI Control Group* on each trial of the TOMM (TOMM 1: $F(1,31) = 91.87, p < .0001$; TOMM 2: $F(1,31) = 139.68, p < .0001$; Retention Trial: $F(1,31) = 127.48, p < .0001$).

The distribution of scores for all subjects was analyzed according to chance levels of responding. Chance performance on the TOMM is a score of 25. Based on a 95% confidence interval, a score of 19 is significantly below chance performance. No subject in the *Control Group* scored below chance on any Trial of the TOMM. In contrast, on Trial 1, four *Simulators* (25%) scored at or below 25 (no subject scored at or below 19). On Trial 2, only one *Simulator* (6%) scored below chance (score = 19), and on the Retention Trial, three *Simulators* (19%) scored at or below chance. Thus, relying on chance level performance results in a high misclassification rate. As shown, using chance level as the criterion identifies only 6% of all the *Simulators* compared to 56% of all *Simulators* when using a conservative cutoff score of 45 (or 100% when only *Compliant Simulators* are used).

Mean performance of estimate scores on the TOMM are presented in Table 9.

Inspection of the table shows that estimate scores for *Simulators* are lower than scores for the *Control Group*. Estimation scores for each trial of the TOMM were analyzed using a repeated measures multivariate analysis of variance. Results revealed only a significant Trial effect ($F(2,76) = 19.07, p < .0001$). The Group effect and the Group x Trial interaction were not significant ($p > .05$). One-way ANOVA's of estimation scores on each trial revealed no significant differences between the groups on Trials 1 and 2 ($p > .05$). However, on the Retention Trial, *TBI Simulators* estimated performance significantly below *TBI Control* patients ($F(1,39) = 4.25, p < .05$).

Results when data from the *Non-compliant Simulators* were removed from the analysis revealed a significant Trial effect ($F(1,30) = 6.29, p < .005$). The Group effect was not significant ($p > .05$). In contrast the Group x Trial interaction was significant ($F(2,60) = 3.94, p < .05$). One-way ANOVAs revealed no significant differences between groups on Trial 1 and Trial 2 ($p > .05$). On the Retention Trial, *TBI Compliant Simulators* scored significantly below *TBI Control* subjects ($F(1,30) = 7.79, p < .01$).

As with the results of Study 1 (University Simulators), estimate scores are equivalent between the groups initially, indicating that the perceived difficulty of the test outweighs the actual difficulty. Although estimate scores on Trial 2 were not significantly different between the two groups, examining the mean performance revealed that scores did increase relative to Trial 1. *Control Group* subjects also increased their estimate performance at a higher level than those in the *Simulation*

Group. The lack of statistical significance may be an artifact of the large variability within each score.

In summary, performance on the TOMM revealed that it can accurately differentiate between groups. Using a cutoff score of 45 on Trial 2 results in 100% sensitivity and 96% specificity when data from the *Non-Compliant* subjects are removed from the analysis. Results also showed that it is important to include a manipulation check to ensure that subjects have complied with task demands. As shown in the present study, there were two groups of *TBI Simulators*: those that could lower their performance (*Compliant Simulators*) and those that could not intentionally lower their performance (*Non-compliant Simulators*). Results also revealed that the perceived difficulty outweighs the actual difficulty, as reflected in the estimation scores in relation to the obtained scores.

The Abbreviated Hiscock: Examining mean scores on the abbreviated Hiscock (see Table 11) reveals that *TBI Simulators*' performance was markedly lower than the *TBI Control Group*. The *TBI Control* subjects maintained errorless performance across all three intervals. Subsequent analysis confirmed this observation.

A repeated measures multivariate analysis of variance on all trials of the abbreviated Hiscock revealed a significant Group effect ($F(1,28) = 11.93, p < .002$). The Trial effect was non-significant, as was the Group x Trial interaction ($p > .05$). However, if data from the *Non-compliant Simulators* are removed from the analyses all effects are significant (Group: $F(1,20) = 32.08, p < .0001$; Trial: $F(2,60) = 6.29, p < .005$; Group x Trial Interaction: $F(2,60) = 3.94, p < .05$).

 Insert Table 11 about here

Subsequent one-way ANOVA's revealed that the *Simulation Group* scored significantly lower than the *Control Group* on all three trials of the abbreviated Hiscock (abbreviated Hiscock 5: $F(1,28) = 8.69, p < .01$; abbreviated Hiscock 10: $F(1,28) = 11.45, p < .005$; abbreviated Hiscock 15: $F(1,28) = 13.07, p < .005$). Similar results occurred when data from the *Non-complaint Simulators* were removed.

For all trials combined, the recommended cutoff score of 33 (Guilmette et al., 1994) results in 56% sensitivity and 100% specificity. Removing the data of the *Non-compliant Simulators* results in 88% sensitivity and 100% specificity.

The distribution of scores was also analyzed according to chance levels of responding. Chance performance on the abbreviated Hiscock is a score of 18. Based on a 95% confidence interval, a score of 13 is significantly below chance performance. No subject in the *Control Group* received a score below chance performance. As seen in Table 2, all *Control* subjects received a perfect score across all trials. In contrast, *Compliant Simulators*' performance was more variable. The majority of subjects scored between the cutoff score of 33 and chance level performance of 18 (75%). Only one subject (13%) received a score below chance.

In summary, performance on the abbreviated Hiscock differentiates between *Simulators* and *Controls* (as observed in Study 1 with University Simulators). Again, findings showed that using chance level performance to identify possible malingerers

Table 11
Group Means and Standard Deviations on the
Abbreviated Hiscock

Measure	Group		
	Simulation TBI	Compliant TBI	Control TBI
A-Hiscock--5	9.4 (3.2)	8.1 (3.6)	12.0 (0.0)
A-Hiscock--10	9.2 (3.1)	7.3 (3.0)	12.0 (0.0)
A-Hiscock--15	8.5 (3.6)	6.6 (3.3)	12.0 (0.0)
A-Hiscock--Total	27.1 (9.6)	22.0 (9.4)	36.0 (0.0)

results in a low classification rate.

Neuropsychological Tests

Mean percentile scores for each test are presented in Table 12. Examination of Table 12 reveals that *Simulators'* performance on neuropsychological tests is remarkably similar to the *Control Groups'* performance. A MANOVA procedure on the entire data set (*Compliant* and *Non-compliant Simulators*) did not reach significance ($p > .05$). Reviewing the subsequent univariate analyses revealed three significant comparisons. *Simulators* scored significantly lower than the *Controls* on acquisition of the Visual Reproductions subtest of the WMS-R, and on the Finger Tapping Test (both preferred and non-preferred hand). The MANOVA was also not significant when data from only the *Compliant Simulators* were used. However, analysis revealed an additional significant univariate comparison (*Simulators* scored significantly lower than *Controls* on the Mental Control subtest of the WMS-R).

Insert Table 12 about here

Examining scores in Table 12 reveals that the *TBI Simulators* and the *TBI Controls* were roughly equivalent on both learning and retention scores of the CVLT. Both groups showed substantially reduced scores indicating that it was a difficult test. The largest discrepancy between the two groups occurred on measures of General Ability, Attention, and Executive Functioning, with *Simulators* performing at reduced levels compared to the *Controls*.

Table 12
Group Means for Percentile Scores on
Neuropsychological Tests

Measure	Group		
	Simulation TBI	Compliant TBI	Control TBI
Learning and Retention			
CVLT			
Learn	1st	1st	5th
Retention	1st	1st	8th
Visual Reproduction			
Learn	49th	42nd	69th
Retention	36th	36th	48th
Attention			
Mental Control	33rd	16th	50th
Trail Making-A	31st	32nd	51st
Digit Span Forward	29th	16th	41st
General Ability			
Finger Tapping			
Preferred Hand	16th	2nd	84th
Nonpreferred Hand	8th	1st	84th
Information	25th	9th	50th
Executive Functioning			
Trail Making-B	39th	30th	50th
Digit Span Backward	30th	25th	52nd
FAS	< 1st	<1st	16th
Animal Naming	16th	16th	50th

In summary, neuropsychological test performance revealed that *Compliant Simulators* typically received lower test scores than the *Control* subjects. However, differences between these group scores did not reach statistical significance. Comparing scores of *Compliant Simulators* to the performance of University Simulators in Study 1, reveals that University Simulators suppressed their performance to a greater extent than the *TBI Compliant Simulators*. That is, University Simulators had lower scores on the visual learning and retention task, and on measures of attention and executive functioning than *TBI Compliant Simulators*. Such a finding suggests caution when extrapolating results of simulation studies that have only used university students.

Rating Scales

Rating scales were given only to those in the *TBI Simulation Group*, and the following analyses were conducted only on the *Compliant Simulators*.

Subjects were asked to endorse how convincing they thought they were in their attempt at deception on each test (Rating Scale 4). The scale ranged from not convincing at all (0%) to very convincing (100%). Examining the frequency distribution of this scale revealed that, on average, subjects did not feel strongly one way or the other about how convincing they were in their attempts at deception, although there was wide variability among the scores ($M=52.0\%$, $SD=30.0\%$).

Scale 6 attempted to address whether subjects could identify the true intent of the tests (i.e., whether they could differentiate between malingering and non-malingering tests). Scores ranged from 1 to 10 (1 = test identified correctly, 10 =

misperception of nature of test). Overall, mean performance was 3.5 ($SD=3.6$). As observed during Study 1 with the University Simulators, wide variability within each scale precludes interpreting the results in a meaningful manner.

On the final scale given, subjects were asked to report what strategy they used to malingering for each test given (see Appendix D). Results were analyzed in a qualitative manner and revealed the use of similar strategies outlined in Study 1.

Strategies included:

1. Reduce speed of performance on timed tests (Finger Tapping Test, Trail Making Test, Mental Control).
2. Change sequence of items (Trail Making Test, Digit Span).
3. Respond with wrong answer or "I don't know" (Information subtest).
4. Randomly omit items (California Verbal Learning Test, Digit Span).
5. Change details of designs (Visual Reproduction subtest).

As with Study 1, strategies appear to be instruction-dependent. That is, subjects used as their main strategy that which was emphasized in the test instructions.

Descriptively, on tests of malingering (TOMM and abbreviated Hiscock) subjects commented that the abbreviated Hiscock was too easy to fake, although almost all *Compliant Simulators* reported doing so. One of the most interesting statements given was that the test was too easy to do poorly on and that if the numbers were more similar (between target and distractor) it would make the test more difficult (see Appendix D). Comments indicate that most individuals thought the test was too easy to fake, yet no one made the association between a test that was too easy to fail and a possible malingering test. In contrast, descriptive comments regarding the TOMM indicate that no one thought it was too easy to fake.

In summary, the present study demonstrated, that it is imperative to include a manipulation check in simulation studies. Fifty percent of the *TBI Simulation Group* did not comply with task demands. For various reasons (to be elaborated on in the General Discussion), they were unable to intentionally lower their performance and instead gave their best performance on the tests. It is also important to include some form of discussion or debriefing regarding how subjects approached the tests (i.e., what strategy did they employ). Only then can researchers have a better idea of how possible malingerers may alter their performance.

The importance of including a clinical simulation group in malingering research was also demonstrated. Inspection of mean percentile scores of neuropsychological tests showed that university simulators suppressed their performance below that of the *compliant TBI simulators*. Such a finding indicates that university simulators may not be the best type of subject for use in simulation research. Instead, TBI patients may produce results that are more representative of what actually occurs in clinical practice.

Study 3

Litigation Study

Study 3 was included as another way of providing converging evidence that the TOMM is a clinically useful measure of malingering. Both previous studies used simulation procedures to evaluate the effectiveness of the TOMM. This study used the more traditional approach used in the literature to test the effectiveness of malingering tests, and that has been reported to produce parallel effects to those found in simulation studies. This approach compares litigating and non-litigating subjects based on the assumption that litigating subjects are more likely to be representative of those suspected of malingering.

Study 3 included comparison groups (cognitively impaired, community and hospital controls, and non-litigating traumatically brain injured) and it was expected that all subjects, in all groups would perform substantially higher on the TOMM than those who were at-risk for malingering (i.e., litigating).

Method

Subjects

Sixty-five subjects were divided into five groups: two control groups (*Community Control* and *Hospital Control*) and three clinical groups (*Cognitively Impaired*, *At-Risk*, and *Not-At-Risk* for malingering TBI patients). All subjects were matched for age and education and subsequent analyses showed that groups were not

significantly different on these variables (age ($p > .05$), education ($p > .05$)).

The *Community Control* group consisted of 13 cognitively intact subjects (8 males and 5 females) recruited from the community. Mean age was 39.2 years ($SD=8.8$). Mean education was 12.5 years ($SD=1.5$). The *Hospital Control* group consisted of 13 male inpatients, who had received a complete neuropsychological evaluation and were judged to be without any significant cognitive impairment. The sample is heterogeneous with respect to diagnostic category (e.g., presymptomatic Huntington's Disease ($n=3$), alcohol and/or drug abuse ($n=3$), multiple chemical sensitivity ($n=1$), seizure disorder ($n=1$), myocardial infarction ($n=1$) or unknown etiology ($n=4$)). Mean age was 45.9 years ($SD=15.0$). Mean education was 13.5 years ($SD=2.1$). The *Cognitively Impaired* group ($n=13$, 8 males and 5 females) consisted of hospitalized neurological patients who had received a complete neuropsychological evaluation, and who were judged to have significant focal neuropsychological impairment. These individuals suffered from a variety of disorders, including psychiatric ($n=3$), polysubstance abuse ($n=4$), stroke ($n=3$), multiple sclerosis ($n=1$), Huntington's Disease ($n=1$), and toxin exposure ($n=1$). Mean age and education was 41.8 years ($SD=8.5$) and 12.1 years ($SD=1.8$), respectively³. The *At-Risk* ($n=13$) and the *Not-At-Risk* for malingering groups ($n=13$) were comprised of TBI patients from either a neurological unit, a neuropsychiatry unit or an outpatient neuropsychological assessment service. TBI

³The *Cognitively Impaired* group was not part of the original design initially but was added from a separate data base to provide a further comparison group.

patients were determined to be at risk or not at risk for malingering before the neuropsychological assessment. Subject assignment was determined in the clinical interview or before the clinical interview if the referral source raised the question of malingering. The *Nor-At-Risk* TBI patients (9 males and 4 females) were included if their cranial CT or MRI scans were interpreted as compatible with traumatic brain injury by staff radiologists. Mean age was 37.4 years ($SD=13.1$). Mean education was 13.3 years ($SD=2.1$). The *At-Risk* group, comprised of 12 males and 1 female, had a mean age of 39.2 years ($SD=8.0$) and mean education of 11.8 years ($SD=3.1$). Risk of malingering was defined by the presence of either pursuit of a personal liability suit or of a disability petition.

Measures and Procedures

For all subjects, except the *Community Control* group, the TOMM was administered in the context of a complete neuropsychological evaluation. Due to the diversity of etiologies for different patients, a standard neuropsychological battery was not administered. The TOMM was administered approximately 30 minutes after the beginning of each session. For the *Community Control* group, a similar procedure was used except that the participants were informed that they were participating in a memory experiment.

Administration of the TOMM and other neuropsychological tests followed standard administration procedures.

Results and Discussion

Mean scores for all trials of the TOMM for each group are presented in Table

13. Inspection of the table shows that performance on each trial of the TOMM is similar for all groups, except those in the *At-Risk* group. Performance by this latter group is markedly lower compared with other groups. Subsequent analyses confirmed this observation.

 Insert Table 13 about here

A repeated measures multivariate analysis of variance was conducted on all three trials of the TOMM. A significant effect of Group ($F(4, 43) = 27.46$, $p < .001$), Trial ($F(2, 106) = 46.84$, $p < .0001$) and Group x Trial interaction was found ($F(8, 106) = 2.95$, $p < .005$).

A one-way analysis of variance (ANOVA) was conducted for each trial. Results revealed a significant difference between groups on each trial of the TOMM (TOMM 1: $F(4, 64) = 30.71$, $p < .0001$; TOMM 2: $F(4, 64) = 21.86$, $p < .0001$; Retention Trial: $F(4, 57) = 21.79$, $p < .0001$). To examine which groups differed on the TOMM, multiple comparisons using Tukey's HSD test were used. Findings revealed that the *At-Risk* group scored significantly lower than all other groups on each trial of the TOMM. No other significant group differences emerged.

The distribution of scores on each trial of the TOMM is shown in Table 14. Comparison among groups show that the distribution was substantially different for the *At-Risk* group. This is most evident on Trial 2 and the Retention Trial. Examining the distribution of scores reveals that there is virtually no overlap between

Table 13
Group Means and Standard Deviations on the TOMM

Measure	Group				
	Community Control	Hospital Control	Cogitively Impaired	Not-At Risk	At-Risk
TOMM 1	47.8 (2.7)	47.9 (2.1)	42.5 (4.6)	45.3 (5.0)	26.6 (10.5)
TOMM 2	50.0 (0.0)	50.0 (0.0)	48.7 (2.8)	49.5 (1.0)	32.9 (12.5)
Retention	49.9 (0.8)	50.0 (0.0)	49.1 (2.4)	50.0 (0.0)	35.4 (10.3)

scores for those *At-Risk* and scores from other groups by Trial 2. Using a score of 49 on Trial 2 results in 100% specificity (i.e., non-malingering subjects correctly identified) for the two control groups (*Community and Hospital*) and 85% specificity for the *Cognitively Impaired* and *Not-At-Risk* groups. Sensitivity (i.e., malingering subjects correctly identified) was 85% for the *At-Risk* group. That is, 85% of the *At-Risk* group scored below 49. Using the cutoff score of 45 suggested by Tombaugh (1996), maintains specificity values for the *Community and Hospital Control and Cognitively Impaired* groups, increases specificity to 100% for the *Not-At-Risk* group, and decreases sensitivity to 77% for the *At-Risk* group.

 Insert Table 14 about here

Examining performance of the *Cognitively Impaired* group revealed that two subjects scored below 45 (one received a score of 41, the other a score of 44). *Cognitively Impaired* subjects were selected from a larger group ($n=42$) (Tombaugh, 1996), and of that larger group, only four subjects scored lower than 45. When specificity is computed on the entire group, it increases from 85% to 91.4%. Examining the neuropsychological profiles of these two subjects shows that they were severely cognitively impaired (retention scores on verbal and visual measures are less than the 1st and 5th percentile, respectively). Both individuals also had identifiable focal deficits, as well as additional medical and psychiatric disorders (e.g., seizure disorder, post-traumatic stress disorder, history of alcohol abuse etc.). In fact, one

Table 14
Percent of Subjects Obtaining Different
Scores on the TOMM

Groups	Number of Correct Responses							
	50	49	48	47	46	45	44-40	<40
Community Control								
Trial 1	31%	31%	8%	6%			15%	
Trial 2	100%							
Retention	92%	8%						
Hospital Control								
Trial 1	23%	23%	23%	15%		8%	8%	
Trial 2	100%							
Retention	100%							
Cognitively Impaired								
Trial 1		8%		15%	15%		31%	31%
Trial 2	69%	15%					15%	
Retention	82%	9%					9%	
Not at Risk								
Trial 1	23%	8%	8%	15%	8%	8%	15%	15%
Trial 2	77%	8%	8%	8%				
Retention	100%							
At Risk								
Trial 1						15%	85%	
Trial 2	15%				8%	8%	69%	
Retention	9%	9%			9%	9%	64%	

subject was admitted to hospital for respite care, suggesting that functionally he was severely impaired. Thus, although these two subjects scored below a cutoff score of 45, their overall neuropsychological profile and medical history is significantly different from the *At-Risk* group, and it is highly unlikely that they would be suspected of malingering.

The performance of the *Cognitively Impaired* group revealed that accurate performance on the TOMM was not substantially influenced by cognitive impairment, particularly when subjects have an opportunity to learn the stimuli (e.g., performance improved on Trial 2 relative to Trial 1 and was maintained over time). These findings also suggest that the TOMM is a relatively "easy" test. That is, cognitively impaired subjects had comparable scores to the control subjects, even when their performance on other measures of learning and retention showed significant cognitive impairment. Thus, the TOMM appears to meet one of the criteria for a good test of malingering: it is not substantially influenced by cognitive impairment.

Seventy-seven percent of the *At Risk* group were correctly identified using a cutoff score of 45. Examining the data revealed that three subjects in the *At-Risk* group achieved a score at or above the cutoff score. One subject received a score of 45, while two subjects received a score of 50 on Trial 2. Examining the neuropsychological profile for each subject reveals that each profile was consistent with TBI, each had documented periods of unconsciousness, and corroborating radiological reports. Recall that risk of malingering was defined, before the assessment, by the presence of a pursuit of a personal suit or a disability petition.

Thus, although these criteria are necessary in identifying possible malingering, they are not sufficient for a diagnosis of malingering. These three subjects highlight the observation that not all subjects seeking compensation are necessarily malingering and convergent evidence is necessary before one can make a diagnosis of malingering.

The distribution of scores for groups was also analyzed according to chance levels of responding. Chance performance on the TOMM is a score of 25. Except for those in the *At-Risk* group, no subject in any group scored at or below chance on any trial of the TOMM. In marked contrast, on Trial 1, 54% of the *At-Risk* group was at or below chance (score ≤ 25). Of those scoring below chance, 39% scored significantly below chance levels (score ≤ 19). On Trial 2, 23% of the *At-Risk* group were at or below chance (including 15% that scored significantly below chance). Performance was not substantially better on the Retention Trial (18.2% scored at or below chance).

Although only subjects in the *At-Risk* for malingering group scored below chance levels, thereby violating the laws of probability, using a below chance performance as a criterion for malingering instead of cutoff scores would have resulted in a high misclassification rate (e.g., 23% of the *At-Risk* group were identified using chance level performance compared to 77% when a conservative cutoff score was used). Although lower than chance performance suggests a definite lack of effort, it is too restrictive and needs to be used in conjunction with cutoff scores derived from appropriate patient groups.

In summary, by using various clinical groups (*Cognitively Impaired, TBI Nor-*

At-Risk) a conservative cutoff score of 45 derived from Tombaugh (1996), resulted in 100% specificity and 77% sensitivity. As previously stated, corroborating evidence (e.g., neuropsychological test performance) is needed to ensure correct classification, but performance on the TOMM provides an objective measure of possible malingering or lack of motivation.

Study 4

Clinical Data

Prior studies have shown that below chance responding is not an adequate criterion for determining malingering. As shown in the previous studies, a more adequate criterion of differentiating malingerers from non-malingerers is based on cutoff scores. The purpose of Study 4 is to provide (1) a criterion based score for TBI patients and (2) to provide data that shows that the TOMM is not sensitive to cognitive impairment. In order to demonstrate this, data from diverse groups of patients are needed. Study 4 highlights performance of individuals with cognitive impairment, aphasia, traumatic brain injury and dementia. Converging evidence demonstrates that the TOMM is a useful measure of malingering.

Method

Subjects

Data was collected from 138 consecutive inpatient and outpatient neuropsychological assessments to the Veterans Administration Medical Centre, Boston, MA. Although some patients had multiple diagnoses, the primary diagnoses were divided into five major groupings: *No Cognitive Impairment* ($n=13$), *Cognitive Impairment* ($n=42$), *Aphasia* ($n=21$), *Traumatic Brain Injury* ($n=45$), and *Dementia* ($n=40$). Three subjects in the demented group were so severely impaired that they could not be tested and were not included in the analyses described below.

The distribution of patients within each category is as follows: *Dementia*: AD

(17), vascular (11), frontal (2), other (7); *Aphasia*: fluent (7), non-fluent (6), global (2), mixed (3), other (3); *TBI*: MVA (9), assault (4), gunshot (1), falls (2), falling object (1), other (6); *Cognitive Impairment*: amnesics (6), multiple sclerosis (2), Parkinson's Disease (3), subdural haematoma (1), carbon monoxide exposure (1), arteriovenous malformation rupture (1), colloid cyst resection in the 3rd ventricle (1), and other (4). All diagnoses were made on the basis of DSM-III-R and DSM-IV criteria.

Mean age of the *No Cognitive Impairment* group was 45.9 years ($SD=15.0$). Mean education was 13.0 years ($SD=3.6$). Mean age of the *Cognitively Impaired* group was 57.2 years ($SD=15.1$). Mean education was 12.4 ($SD=3.0$). The *Aphasia* group had a mean age of 66.2 years ($SD=10.9$) and a mean education of 13.0 ($SD=3.5$). Mean age and education of the *TBI* group was 40.0 ($SD=15.3$) and 13.0 ($SD=2.4$), respectively. The final group, the *Dementia* group, had a mean age of 72.1 years ($SD=7.6$) and a mean education of 11.9 years ($SD=3.4$).

Materials and Procedure

The TOMM was administered in the context of a complete neuropsychological evaluation. Due to the diversity of etiologies, a standard battery was not given. However, scores from a verbal learning task (CVLT or Word-List subtest of the Learning and Memory Battery, LAMB; Schmidt & Tombaugh, 1995) and a visual learning task (Visual Reproduction subtest of the WMS-R) are available for 82% of the non-aphasic patients.

The TOMM was administered according to the instructions outlined in Study

1. Particular care was taken to ensure that severely impaired patients attended to the stimuli. For example, on learning trials it was often necessary to direct the attention of demented patients to the stimuli by pointing to each target as it was presented. On test trials, the following instructions were repeated with each pair of test stimuli.

"Point to the picture I showed you before. Was it this one (examiner points to top picture) or this one (examiner points to bottom picture)?" To avoid perseverative responding (e.g., always selecting either the upper or lower picture), a patient's response was not accepted until the examiner had pointed to both pictures. It was also important to ensure that demented patients understood that they were to select 1 of the 50 pictures they had been shown rather than merely picking a familiar picture. In the latter case, the patient would usually say something to the effect of "Oh that's a mouse." When this occurred, the examiner would say "Yes, that is correct, but which picture did I show you earlier? Was it this one (points) or this one (points)?"

Results and Discussion

Group means for performance on the TOMM are presented in Table 15. Inspection of the table reveals that the performance of the first four groups (*No Cognitive Impairment*, *Cognitive Impairment*, *Aphasia*, *TBI*) was comparable with the performance of the *Dementia* group being substantially lower. The robustness of the TOMM is particularly evident in Trial 2 in which the dementia group obtained 92% correct, and the remaining four groups responded at greater than 97% accuracy.

 Insert Table 15 about here

These observations were verified by statistical analyses. ANOVAs performed on the scores from each trial revealed significant differences among the groups (Trials 1, 2, and Retention; $F(4,153) = 7.22, 9.02, \text{ and } 7.10$, respectively; $p < .001$). Tukey-HSD multiple comparisons, using the harmonic mean to adjust for uneven number of patients in the different groups, showed that the performance of the *Dementia* group on Trial 1 was significantly lower than that for all groups except the *Cognitively Impaired* group. On the Trial 2 and the Retention Trial, the *Dementia* group was significantly lower than all groups which did not differ significantly from each other.

Scores from a visual learning task (Visual Reproduction subtest from the WMS-R) and a verbal learning task (CVLT or Word-List subtest from the LAMB) were available for 82% of the non-aphasic patients. Table 16 shows the mean percentile scores for the verbal and visual learning tests. Comparison of the learning and retention scores to those obtained on the TOMM shows that recognition performance (TOMM) was relatively unrelated to measures of visual and verbal learning. This observation was supported by correlational analyses showing that the correlation coefficients between TOMM and the other learning measures for all groups ranged from .20 to .35 across the three trials. The lack of substantial correlation between the TOMM and visual and verbal learning scores is important

Table 15
Group Means and Standard Deviations Number
of Correct Responses on the TOMM for the
Clinical Sample

Measure	Trials		
	Trial 1	Trial 2	Retention
No Cognitive Impairment	47.9 (2.1)	50.0 (0.0)	50.0 (0.0)
Cognitive Impairment	43.9 (5.3)	48.6 (3.1)	49.5 (1.5)
Aphasia	46.3 (4.2)	49.3 (1.9)	49.8 (0.2)
Traumatic Brain Injury ¹	45.9 (4.7)	49.4 (1.3)	49.6 (1.1)
Dementia ²	41.0 (6.6)	45.7 (5.3)	47.0 (4.4)

1. No patient of the Traumatic Brain Injury group was involved in compensation hearing or litigation.

2. 3 of the original 40 patients were too severely demented to be tested and are not included in the table.

because it demonstrates that the TOMM is relatively insensitive to cognitive impairment that may be associated with various neurological impairments.

 Insert Table 16 about here

Table 17 shows the distribution of scores for the clinical sample. The scores from Trial 2 are particularly noteworthy. The *Dementia* group was the only group that had a relatively high percentage of scores (27%) that fell below 45. Only four patients (9.6%) in the *Cognitive Impairment* group, one patient (4.8%) in the *Aphasia* group and one patient (2%) in the *TBI* group scored below 45. Examining the neuropsychological profiles and medical history of these individuals indicates that these individuals would not be suspected of malingering. No patient in the *No Cognitive Impairment* group obtained a score lower than 45. It is also important to note that only four dementia patients (10.8%) scored less than 40 on Trial 2 (scores = 29, 33, 33, and 37, respectively). All of these patients were judged to be moderately-to-severely demented (MMSE scores = 16, 19, 15 and 7, respectively). Thus, based on these clinical data, a cutting score of 45 produces a high level of specificity. That is, it correctly classifies 95% of all non-demented patients (91% of all patients) as not malingering.

Table 16
Mean and Standard Deviations for Percentile Scores
on Verbal and Visual Tests of Learning and
Retention for the Clinical Sample

Group	Verbal ¹		Visual ²	
	Learning	Retention	Learning	Retention
No Cognitive Impairment	50.0 (29.2)	58.8 (29.6)	70.6 (32.9)	68.8 (35.2)
Cognitive Impairment	20.2 (22.0)	18.4 (21.6)	52.4 (29.1)	28.7 (33.8)
Traumatic Brain Injury	26.5 (31.6)	21.7 (23.6)	52.3 (34.1)	38.4 (34.4)
Dementia	5.1 (8.3)	3.1 (5.9)	21.5 (26.7)	9.5 (17.5)

1. Verbal test scores are from the CVLT or the Word List from the Learning and Memory Battery (LAMB).

2. Visual test scores are from Visual Reproduction of the WMS-R.

Insert Table 17 about here

In summary, data from this study clearly demonstrates that the TOMM is not significantly influenced by cognitive impairment. Thus, a low score on the TOMM indicates a lack of effort or motivation disorder, of which malingering would be considered a subtype. Recall that the definition of malingering implies a conscious intent to deceive. Low scores on the TOMM reflect a lack of effort or motivation but do not address the issue of conscious intent to deceive.

Table 17
Percent of Correct Responses for Different Clinical Groups
on the TOMM

Groups	Number of Correct Responses							
	50	49	48	47	46	45	44-40	<40
No Cognitive Impairment								
Trial 1	23%	23%	23%	15%	0%	8%	8%	0%
Trial 2	100%	0%	0%	0%	0%	0%	0%	0%
Retention	100%	0%	0%	0%	0%	0%	0%	0%
Cognitive Impairment								
Trial 1	10%	12%	10%	7%	12%	2%	29%	19%
Trial 2	67%	12%	5%	5%	2%	0%	7%	2%
Retention	85%	6%	3%	0%	0%	3%	3%	0%
Aphasia								
Trial 1	14%	19%	10%	19%	10%	14%	10%	5%
Trial 2	76%	14%	5%	0%	0%	0%	5%	0%
Retention	82%	12%	6%	0%	0%	0%	0%	0%
Traumatic Brain Injury								
Trial 1	31%	16%	2%	11%	4%	2%	12%	15%
Trial 2	78%	7%	7%	4%	0%	2%	2%	0%
Retention	81%	5%	7%	5%	0%	2%	0%	0%
Dementia								
Trial 1	8%	5%	8%	5%	5%	5%	22%	43%
Trial 2	27%	14%	11%	8%	3%	11%	16%	11%
Retention	32%	18%	18%	7%	4%	4%	11%	7%

General Discussion

The four studies presented provide compelling evidence that the Test of Memory Malingering (TOMM) is an excellent indicator of lack of effort or motivation. As previously outlined, there are numerous tests available that putatively measure malingering, yet none have a comparable normative data base and none meet all of the criteria for a clinically useful measure of malingering.

For a test of malingering to be considered clinically useful, four criteria must be met. These are: (1) the test must differentiate malingerers from non-malingerers, (2) the perceived difficulty of the test must outweigh the actual difficulty, (3) the test must not be influenced by non-motivational factors, such as cognitive ability, age, or education, and (4) the test must be perceived as a legitimate measure of cognitive function (i.e., have face validity). The data from all four studies, summarized in Table 18, clearly illustrate that performance on the TOMM accurately and consistently differentiates between those who put forth effort from those who do not.

Results of both simulation studies clearly demonstrated that performance on the TOMM can differentiate between simulators and non-simulators. Using a cutoff score of 45 resulted in 84% sensitivity and 100% specificity in Study 1. In Study 2, 100% sensitivity and specificity was observed with the *Compliant Simulators*. High sensitivity and specificity were also observed in Study 3 between those patients who were *At-Risk* for malingering and all other patient and control groups (77% sensitivity and 100% specificity).

Insert Table 18 about here

The second criterion of perceived difficulty outweighing actual difficulty was evaluated by comparing estimated and obtained scores. The results showed that the TOMM was perceived to be more difficult than it actually was. This was particularly evident in the estimation scores for Trial 1 in both simulation studies. The simulation groups did not differ significantly from their respective control groups on their estimated performance on Trial 1 of the TOMM. By Trial 2, control groups increased their estimations, while simulators maintained performance at relatively low levels. Given that both groups were initially equivalent (simulation and control groups), they must have perceived the difficulty of the test in a similar manner. However, once exposed to the first trial, control subjects realized that it is a relatively easy test and subsequently increased their estimate scores.

In terms of empirical research, it was important to include TOMM estimation scores as an indication of perceived difficulty of the test. However, asking for estimate scores in a clinical population proved to be too cumbersome (i.e., patients were found reluctant to estimate their performance and would often state ranges of performance). Given that estimation scores provided little clinical utility, they were subsequently dropped from data collection with clinical groups.

The third criterion of a useful test of malingering is that it must not be sensitive to non-motivational factors. As previously reviewed, several malingering

Table 18
Summary Table for All Experiments

Group	TOMM Trials		
	TOMM 1	TOMM 2	TOMM Retention
Simulators			
University (Exp. 1)	32.6 (7.8)	35.7 (8.9)	34.3 (8.2)
Compliant TBI (Exp. 2)	28.1 (17.8)	32.1 (7.3)	31.6 (8.0)
At-Risk Patients (Exp. 3)	26.5 (10.5)	32.9 (12.5)	35.4 (10.3)
Controls			
University (Exp.1)	48.8 (1.7)	49.9 (0.2)	50.0 (0.0)
TBI (Exp. 2)	47.1 (3.6)	49.6 (1.3)	49.6 (1.0)
Hospital (Exp. 3)	47.9 (2.1)	50.0 (0.0)	50.0 (0.0)
Community (Exp. 3)	47.8 (2.7)	50.0 (0.0)	49.9 (0.3)
Not at risk (Exp. 3)	45.3 (5.0)	49.5 (1.0)	50.0 (0.0)
Patient Groups			
Cognitively Impaired (Exp. 4)	43.9 (5.3)	48.6 (3.1)	49.5 (1.5)
Aphasia (Exp. 4)	46.3 (4.2)	49.3 (1.9)	49.8 (0.6)
TBI (Exp. 4)	45.9 (4.7)	49.4 (1.3)	49.6 (1.1)
Dementia (Exp. 4)	41.0 (6.6)	45.7 (5.3)	47.0 (4.4)

tests have been found to be susceptible to cognitive impairment (e.g., Rey Fifteen Item Test, Portland Digit Recognition Test, Victoria Revision of the Hiscock Digits, Recognition Memory Test). Thus, lowered performance on these tests may be the result of malingering or the result of actual cognitive impairment. Performance on these tests have also been affected by age and education (e.g., Goldberg & Miller, 1986; Martin et al., 1994; Philpott & Boone, 1994; Schretlen, 1991). With such confounding variables, it is difficult to determine if performance is related to motivation or other factors, including neurological impairment.

In previous research, with a more difficult four-choice version of the TOMM, age and education did not influence performance (Hubley, McIntyre & Tombaugh, 1995). Consistent with this, performance on the two-choice TOMM was not influenced by age, education, or cognitive impairment. This was most strikingly illustrated with demented patients. Results presented in Study 4 demonstrated that by Trial 2 the demented patients obtained an accuracy rate of 92% correct. This evidence is particularly compelling since they were severely impaired on the more traditional learning and retention neuropsychological measures. Of all groups considered, including the *Cognitively Impaired*, *Aphasic TBI* and *Dementia* groups ($n=108$), there were only five individuals (4.6%) who scored below the established cutoff score of 45 on the TOMM. This shows that the performance on the TOMM is robust and not susceptible to the affects of cognitive impairment.

The fourth and final criterion for a test of malingering involves face validity. That is, a malingering test must appear to be a legitimate measure of cognitive

function, otherwise malingerers may suspect the validity of the test. In the case of the TOMM, it must appear to be a legitimate measure of memory. In the present study, validity of the tests was queried in the debriefing through use of the rating scales. As previously stated, quantitative analyses of the rating scales were not particularly helpful due to the large variability in the answer distributions. However, on an open-ended question, no subject stated that the TOMM was easy to fake or identified it as a test of malingering. Behavioral observations support these findings. That is, subjects attended to the stimuli (TOMM pictures) in an effortful manner and were often surprised by how well they did (control group). Comments by simulators at the end of testing also revealed how surprised they were that they could have actually done well. In an interesting case, one *TBI Simulator* reported trying her best on the TOMM because she was challenged by the test and wanted to see if she could perform well on it. This was a woman who had been severely impaired, had undergone extensive rehabilitation and reportedly takes every opportunity to demonstrate that she is recovering. Such statements reinforce the finding that the perceived difficulty of the test outweighs the actual difficulty.

Another advantage of the TOMM, over all other available malingering tests, is the large normative data base that has been compiled. Scores are available on the TOMM from a diverse group of clinical patients that make questions about suspected malingering easier to resolve. Individual scores can be compared to mean scores of specific diagnostic groups (e.g., TBI, Dementia). For example, if an individual undergoing litigation, living independently in the community and having no

complicating medical and/or psychiatric background, scores lower than the cutoff score, malingering would be strongly suspected after comparing his/her scores to those in the normative sample. To date, the TOMM is the only malingering test available that has such a comprehensive normative data base.

Based on a review of the literature, the abbreviated Hiscock appeared to be the best test available (prior to the development of the TOMM) because of high sensitivity and specificity and relative ease of administration (e.g., Guilmette et al., 1993; 1994; Hiscock & Hiscock, 1989; Prigatano & Amin, 1993). Consequently, the abbreviated Hiscock was used in the current simulation studies to provide a benchmark against which the TOMM was compared.

In terms of the criteria outlined above for a useful measure of malingering, it was demonstrated that performance on the abbreviated Hiscock can differentiate between simulators and non-simulators. Using the recommended cutoff score of 33 resulted in 76% sensitivity and 100% specificity in Study 1. In Study 2, 88% sensitivity and 100% specificity was observed for *Compliant Simulators*. Thus, as with the TOMM, the abbreviated Hiscock is able to differentiate between groups.

The perceived difficulty of the abbreviated Hiscock is manipulated by increasing the length of the interval between presentation of the stimuli and recall (5 sec., 10 sec., and 15 sec.) (Hiscock & Hiscock, 1989). As a result, performance is predicted to deteriorate as length of the interval increases. Empirically, as predicted, results suggest a progressive decline in performance as the length of the interval increases. However, comments made during testing, as well as behavioral

observations, did not indicate that subjects actually believed it was a difficult test. Several simulators made suggestions on how the test could be altered to increase its difficulty. Thus, some questions remain as to whether the perceived difficulty exceeds the actual difficulty of the test. This further questions the face validity of the abbreviated Hiscock.

Research examining the effects of age and education on performance of the abbreviated Hiscock are lacking. Although, given the nature of the test, it is suspected that these variables would not influence performance in any significant way. Research also has shown that the Hiscock is relatively insensitive to the effects of severe brain damage and/or depression (e.g., Guilmette, et al., 1993; 1994; Prigatano & Amin, 1993). Thus, it is likely that the TOMM and the abbreviated Hiscock are comparable in their insensitivity to cognitive impairment and demographic variables. However, research needs to be conducted evaluating the abbreviated Hiscock across a wider range of clinical groups to substantiate this claim.

Overall, the abbreviated Hiscock has been found to have good sensitivity and specificity, and it was not influenced by cognitive impairment (at least not brain damage or depression). Empirically, it is an adequate measure of malingering in that it differentiates groups. However, based on comments regarding the questionable validity of the test, it is unclear if the criteria of face validity and perceived difficulty are met. In clinical practice, the abbreviated Hiscock also runs the risk of "insulting" or alienating those individuals who have put forth good effort throughout testing, and may result in them giving up or losing motivation to perform well (e.g., Lezak,

1995).

In summary, data from the four studies provide converging evidence that the TOMM is an excellent measure of poor effort or lack of motivation. Results clearly demonstrated that the TOMM meets all criteria of a clinically sound malingering test, as well as providing an extensive normative data base. The abbreviated Hiscock was shown to meet two of the four criteria, however uncertainty exists regarding the remaining two criteria. Further, the abbreviated Hiscock test has no established normative data base from which to compare test performance. As such, the TOMM surpasses all available measures and should become the gold standard by which to compare other measures of malingering.

Other Issues

The current series of research studies addressed not only the adequacy of the TOMM as a measure of malingering, but also set out to address three issues regarding simulation research. First, the criterion used to establish malingering was evaluated. This involved comparing below chance responding to the use of cutoff scores. Second, the criticisms of simulation research were addressed, and suggestions to overcome these limitations were incorporated into the study design. Third, the issue of compliance and strategy use were addressed and evaluated. Each will be discussed below.

Below chance responding vs. cutoff scores

As previously discussed, symptom validity testing (SVT) developed out of a need to decide, with greater certainty, if a person was malingering. SVT enables the

practitioner to empirically decide if the patient was malingering or not, with malingering decided by the laws of probability (Pankratz et al., 1975; Pankratz, 1979; 1988). The SVT is based on the assumption is that if an individual is exposed to many trials (e.g., 100 trials) using a forced choice procedure, she/he will overestimate how severely impaired they must appear, thereby performing lower than chance (e.g., less than 50 correct). The only way to interpret below chance performance is that the person correctly identified the stimulus but chose to ignore it during test trials (i.e., reported the wrong stimulus).

As shown in Study 1, only three *Simulators* scored below chance performance on Trial 2 of the TOMM (less than 25). No simulator scored below chance levels on the abbreviated Hiscock (less than 18). In Study 2, only one simulator scored below chance on Trial 2 of the TOMM. One simulator scored below chance on the abbreviated Hiscock. Findings were similar in Study 3 (three subjects *At-Risk* for malingering scored below chance on the TOMM). Thus, relying on below chance responding as a criterion for malingering results in a high misclassification rate.

Part of the difficulty of relying on chance level responding is that relatively large numbers of trials need to be presented, which is often not possible in a clinical setting. As well, those suspected of malingering may score above chance levels, but score below expected norms. Findings from the current research studies are consistent with results reported in the recent literature that suspected malingerers or simulators do not ordinarily score below chance (Binder & Willis, 1991; Guilmette, Hart et al., 1993 & 1994; Martin et al., 1992; Prigatano & Amin, 1993; Slick et al.,

1994). As a result, cutoff scores are typically used.

Cutoff or criterion scores are scores that have been shown previously to separate "normal" from "abnormal," or as in the present context, scores that separate "suspected malingerers" from "non-malingerers" (Lezak, 1995). Scores can be derived statistically, or more typically by inspection of the distribution of scores. When examining the distribution of scores, it is usually the lowest score of the "normal" or "non-malingering" group that is selected as the cutoff score to differentiate between groups. The majority of "non-malingerers" score above this established cutoff and those simulating or suspected of malingering score below it.

Typically neuropsychological test scores are compared to a representative group of "normal" individuals. That is, scores obtained during testing are compared to scores from a large sample of cognitively intact individuals in order to determine whether the individual under evaluation deviates from what is expected in the population. However, in malingering research, the best reference group is not necessarily a group of cognitively intact individuals. Suspected malingerers are made up of individuals who have suffered a specific neurological insult, and who are not motivated to perform to the best of their ability. The best reference group is therefore a group who have suffered from the same type of neurological insult, but who are motivated to perform to the best of their ability. Using such a group as the "norm" removes the confound of cognitive ability associated with the neurological damage. Thus, any differences between groups is likely due to motivation and not the result of cognitive ability.

As shown in the present research, a score of 45 was considered a "conservative" cutoff score relative to the lowest scores obtained by the control subjects in the three experiments. That is, based on the data in Study 1, using a cutoff score of 49 resulted in 92% sensitivity and 100% specificity. In Study 2, a cutoff score of 48 resulted in 63% sensitivity and 96% specificity. In Study 3, a cutoff score of 49 resulted in 85% sensitivity and 100% specificity for the control groups and 85% specificity for the non-malingering clinical groups. However, the lowest score in each study was not selected as the cutoff score because it was based on a limited sample and hence restricted generalization. Based on the distribution of scores of all groups, particularly the four clinical groups in Study 4, a score of 45 was selected as the cutoff score. As previously stated, this cutoff score yielded high levels of sensitivity and specificity across the four studies, and suggest that when the TOMM is used clinically,

any score lower than 45 on Trial 2 or the Retention Trial should raise concern that the individual is not putting forth maximum effort. Of course, as with any test, this score is viewed as a guideline with the likelihood of malingering increasing as the score deviates further from the normative baseline for each specific diagnostic group.

Simulation Research

The area of simulation research has often been criticized because of its reliance on university students asked to simulate performance. Often, performance is compared only between two groups. It is generally not known if simulators approximate performance of the group they are trying to mimic (e.g., TBI), let alone

if they approximate performance of suspected malingerers. Thus, the ability to generalize findings is hampered.

However, simulation studies with university students are useful for several reasons, particularly when developing a new measure of malingering. First, they provide evidence that the malingering test can accurately differentiate between groups (simulators and controls). Second, they offer the opportunity to obtain a greater degree of control over variables than occurs with many clinical groups. And third, given the nature of the disorder (to intentionally lower performance without detection) it is extremely difficult to locate bonafide malingerers.

Taken together, the current series of simulation studies offer convincing convergent evidence, that the TOMM is an extremely sensitive measure of malingering/motivation. Although the traditional simulation paradigm was employed in Study 1, a number of suggestions were incorporated into the design to increase generalization of the findings. These included offering a substantial differential reward, allowing students one week to prepare for the study, embedding malingering tests in a relatively comprehensive neuropsychological evaluation, and using extensive debriefing.

A non-traditional and highly unique design was employed in Study 2. Here, a clinical population of traumatically brain injured individuals was asked to simulate or exaggerate performance. I was unable to locate any other published study in the literature that has asked a TBI group to "fake-bad." Using TBI individuals approximated, to the best degree available while still maintaining a degree of

experimental vigour, what may occur with suspected malingerers in a clinical setting. It was possible to evaluate not only their performance on the tests, but also the strategies they used in their attempt at deception. As observed, their strategies were remarkably similar to those reported by the university simulators. Again, such findings increase the level of confidence one has in results obtained with university students. Study 3 highlighted realistic comparisons between different patient and control groups, including a group at-risk for malingering (litigating TBI patients). Here, too, the TOMM differentiated between patients at risk for malingering and control subjects. Thus, overall, all three studies showed that simulators and suspected malingerers perform at significantly lower levels on the TOMM compared to those trying their best. TOMM mean scores were also equivalent across both groups of simulators and probable ("at-risk") malingerers (see Table 18). Based on this finding, it may be that university simulators more adequately portray the performance of head injured malingerers than initially thought. Finally, incorporating data from the diverse clinical populations (Study 4) showed that even those with relatively severe levels of impairment performed at relatively high levels on the TOMM. This shows that even though the TOMM is sensitive to malingering, it is *not* sensitive to bonafide neurological impairment.

Taken together, results from all studies clearly demonstrated that the Test of Memory Malingering is highly sensitive in detecting when someone is not putting forth maximum effort. Findings also increased the confidence one has in the findings of simulation paradigms using university students. University students may not be

ideal simulators, but they are better than initially thought.

Compliance and Strategy Use

As observed in the current simulation studies and in findings reported in the literature, including a manipulation check has two important benefits. First, it allows the researcher some degree of control and assurance that the subjects attempted to comply with the instructions. That is, it provides evidence as to whether the subjects actually tried to mangle. Second, it provides an exclusionary criterion for the study. As previously stated, 4% of the subjects in Study 1 and 50% of the subjects in Study 2 did not comply with simulation instructions. Data from these subjects were subsequently removed from all analyses. Other studies have also reported that not all subjects have been compliant with instructions (e.g., Binder & Willis, 1991; Goebel, 1983; Heaton et al., 1978).

The issue of compliance was interesting and provided some unexpected results, particularly in Study 2. Based on patterns established in the literature, it was expected that a few individuals would not comply with the simulation instructions. It was not anticipated, however, that 50% of the subjects in Study 2 would not comply with the instructions after they had agreed to participate in the study. During the initial screening and information session, an additional five individuals decided against participating when they found out they would be asked to intentionally lower their performance. It should be noted, however, that these same individuals offered to participate in future studies if they were allowed to perform to the best of their ability. Thus, it appeared to be the requirement of "faking" that detracted potential

subjects from participating. The reason for this lack of compliance is open to speculation. Based on comments made during the testing, there were different reasons for non-compliance. One reason appeared to be related to the person's own value system. That is, some individuals found that in the "testing" environment they could not bring themselves to fail a test, no matter how hard they tried.

A second reason for non-compliance may be related to cognitive ability and pertains primarily those who had either a more severe type of head injury or an injury involving the frontal lobes and associated executive system. These individuals appeared unable to get into the "set" or "mode" of intentionally lowering their performance. Although they were able to correctly verbalize the instructions to the examiner, they were unable to behaviourally lower their performance. Many individuals had difficulty conceptualizing how to "pretend" and their behaviour tended to be rather concrete or literal. Such behaviour is thought to be related to the executive system/frontal lobe functioning (e.g., Lezak, 1995). In most traumatic brain injuries it is the frontal lobe that is injured, particularly in motor vehicle accidents. Therefore, the observed behaviour is consistent with the type of damage and is not totally unexpected. It would be interesting in future research to address this issue more directly. That is, what is the relationship between type of injury and ability to "fake" performance.

Intuitively, one would assume that the more severe the injury, the less likely that an individual would be able to lower his/her performance in a realistic manner. This was certainly true of some individuals. However, examining the data of the *non-*

compliant TBI simulators in Study 2 revealed that a number of individuals with more mild-to-moderate injuries did not comply with the simulation instructions. These individuals, both those who had initially tried to lower their performance and those who had refused to participate, commented that they had worked so hard to overcome the sequelae following their TBI that they could not comply with the instructional set. That is, they appeared to be extremely "ego-involved" with their performance and were unable to comply with the instructions to lower their performance. These individuals were often very active in various head injury associations, rehabilitation, community education regarding the sequelae of TBI and would often take every opportunity available to demonstrate that they had recovered or were well on the way to recovery.

In summary, the extent of non-compliance, particularly as it pertained to the TBI simulators, was not anticipated. A number of speculations were put forth as possible explanations of non-compliance and future research needs to address this issue. The degree of compliance in the current studies relied on subjects being truthful about their performance. Perhaps an alternative way to assess compliance in a more objective manner, would be to test subjects under two conditions. Subjects could be tested under a simulation condition and under a condition where they try their best. In this manner, performance could be compared across the two conditions and compliance with instructional set (or condition) could be determined by the difference between the two scores.

Just as important as checking compliance, is examining what strategies subjects

used in their attempt to malingering, and whether they could identify which test(s) was (were) specifically designed to detect malingering. Although we know that performance can be intentionally altered without detection (e.g., Franzen, 1989), we do not know precisely how one alters performance. A first step in addressing this issue is to assess how simulators approach the tests.

There has been only one study published (Beetar & Williams, 1995) that has addressed the issue of strategy use. Unfortunately, Beetar and Williams inferred the strategy employed by the type of errors their subjects made, rather than directly soliciting the information from their subjects. The present studies attempted to learn about strategies by using various types of questionnaires. However, the rating scales used in the current study were not as useful as initially anticipated. Variability observed in the distribution of responses resulted in quantitative analyses that were virtually meaningless.

A number of problems existed with the rating scales employed. Primarily, too many scales were used at the end of a relatively long experiment, and subjects probably did not invest the time or energy to answer the questions in a thoughtful manner. Although future research needs to address participants' opinion about the tests and to determine whether they can identify which test(s) is (are) designed to detect malingering, it should be done in a more efficient manner. For example, it might have been better to use a dichotomous scale (versus a continuum as used in Scale 6) addressing whether subjects thought a particular test was designed to detect malingering or not. Those items endorsed in a positive manner could then be

addressed in more detail. Given that a majority commented that the abbreviated Hiscock was too easy to fake, yet did not realize that a test that was too easy to perform poorly on was likely a test of malingering, it would be interesting to ask subjects what they consider to be a test of malingering. That is, how would they identify a malingering test?

The most useful information came from the last scale where subjects were asked verbally, in an open ended format, how they approached each test. That is, what was their strategy in lowering their performance on each test? Although difficult to examine in a quantitative manner, insight was gained into strategies used. It was generally observed that participants' strategies stemmed from the instructions given for each test. Given that participants in Study 1 had never been exposed to previous neuropsychological testing, listening to the instructions about what was expected on various tests, presented an opportunity for them to figure out how to modify their performance. It would be interesting to retest these subjects now that they know what to expect, in order to examine whether their strategies remained the same or whether prior exposure altered their strategies. That is, whether their strategies would reflect a more sophisticated approach because of prior exposure. Such a study would have clinical relevance as patients suspected of malingering often go through repeated assessments. This last point was highlighted by the performance and comments of one *TBI simulator* who had undergone six neuropsychological evaluations since her motor vehicle accident five years ago. During testing, she could anticipate the demands of the tests thereby permitting her to alter her performance in

a plausible manner. For example, on the California Verbal Learning Test, a verbal list learning task, she omitted a few items, exhibited a mild degree of interference from List B, and recognition performance was higher than free recall. She reported that it was because she had previously been exposed to the tests that she was able to alter her performance in a realistic manner. She also stated that had she not had the prior exposure, she would not have been able to perform the simulation in a realistic manner.

In summary, although the scales did not produce the hoped for information regarding which test(s) was (were) designed to detect malingering, the results obtained can be used to guide future simulation research. Understanding what strategies simulators use and why they chose the ones they did, may provide insight into the strategies that suspected malingerers use in clinical practice. In the present series of studies, several pronounced strategies were used. Simulators in Study 1 generally reported lowering their performance on all tests employed. They did not try to adjust their performance depending on the nature of the task. This self-report is consistent with quantitative evidence showing that mean scores were lower for simulators on all tests compared to controls. This finding is consistent with the literature (e.g., Binder & Willis, 1991; Greiffenstein, et al., 1994; Mittenberg et al., 1993; Trueblood & Schmidt, 1993; Wasylw & Cavanaugh, 1989) and may reflect the fact that as a group they had little knowledge of the effects of TBI and that they approached the tests in a relatively unsophisticated manner. Comparing mean percentile scores of the university simulators to that of TBI control subjects supports this observation.

Although one frequently voiced criticism of using university students in simulation research, no published study has compared university simulators to TBI simulators. Examining mean percentile scores of the tests given between the two simulation groups (university and compliant TBI) revealed a remarkable similarity between groups. Performances of these two groups were comparable except for performance on the Visual Reproduction subtest of the WMS-R and the Trail Making Test, part A. Compliant TBI simulators received substantially higher scores on these latter measures. Performance on the malingering tests (TOMM and the abbreviated Hiscock) revealed that TBI simulators scored somewhat lower than the university simulators, although not substantially so. These comparisons suggest that although university simulators may exaggerate their performance compared to a motivated TBI group, their performance is not substantially different from a TBI simulating group. Thus, university students are perhaps more satisfactory in their attempts at malingering than what is generally reported in the literature. Of course it can be argued that both these simulation groups do not approximate performance of suspected malingerers. However, in Study 3, mean performance on the TOMM of those *At-Risk* for malingering are available for comparison and findings revealed remarkable similarities between groups, particularly on Trial 2. It would therefore appear that simulators approximate performance of what occurs in suspected malingerers (at least on the TOMM), indicating that perhaps the degree of generalizability is higher than anticipated.

In summary, university simulators exaggerate their performance in relation to

a motivated TBI control group. However, performance is roughly equivalent to a compliant TBI simulation group, indicating that university simulators may be more satisfactory in their attempt at simulation than initially thought.

Summary

In conclusion, the results of the four studies showed that the TOMM is an exceptional indicator of lack of effort or motivation. Data were presented from diverse groups of subjects, under different design paradigms, that provided converging evidence that a low score on the TOMM indicates that there was lowered motivation or a lack of effort during testing. A score below the recommended cutoff score does not, however, suggest that the individual is malingering. That is, the definition of malingering involves a conscious intent to deceive for external secondary gain. A low score on TOMM indicates a lack of effort or an exaggeration or false production of symptoms, but it does not address the volitional aspect of performance. As highlighted in Study 3, some individuals were found to have scores below cutoff level. However, given that there was no other supporting evidence of possible malingering these individuals were *NOT* suspected of malingering. Such findings highlight the importance of considering converging lines of evidence prior to the diagnosis of malingering. The TOMM provides objective evidence of lack of effort or motivation and will increase the examiner's confidence in the diagnosis of malingering once all evidence is considered.

References

- American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.
- Andrikopoulos, J. (1994, February). Disproportionate attention deficit in malingering. Poster presented at the annual meeting of the International Neuropsychological Society, Cincinnati, OH.
- Baker, W.J., Hanley, J.R., Jackson, H.F., Kimmance, S., & Slade, P. (1993). Detecting the faking of amnesia: Performance differences between simulators and patients with memory impairment. Journal of Clinical and Experimental Neuropsychology, 15, 668-684.
- Baker, W.J., Gola, T.J., & Greiffenstein, M.F. (1994, October). Discriminant analysis of neuropsychological tests with brain damaged, post concussive and malingering subjects. Poster presented at the 14th annual meeting of the National Academy of Neuropsychology, Orlando, FL.
- Beck, A.T. (1987). Beck Depression Inventory. San Antonio, TX: The Psychological Corporation.
- Beetar, J.T., & Williams, J.M. (1995). Malingering response styles on the memory assessment scales and symptom validity tests. Archives of Clinical Neuropsychology, 10(1), 57-72.
- Benton, A., & Spreen, O. (1961). Visual memory test: The simulation of mental competence. Archives of General Psychiatry, 4, 79-83.
- Bernard, L.C. (1990). Prospects for faking believable memory deficits on

neuropsychological tests and the use of incentives in simulation research.

Journal of Clinical and Experimental Neuropsychology, 12(5), 715-728.

Bernard, L.C. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: The effect of serial position. Archives of Clinical Neuropsychology, 6, 81-88.

Bernard, L.C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. Journal of Clinical Psychology, 46(4), 432-436.

Bernard, L.C., Houston, W., & Natoli, L. (1993). Malingering on neuropsychological memory tests: Potential objective indicators. Journal of Clinical Psychology, 49(1), 45-53.

Bernard, L.C., McGrath, M.J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler Memory Scale-Revised. Archives of Clinical Neuropsychology, 8, 539-551.

Binder, L. (1993a). Assessment of malingering after mild head trauma with the Portland Digit Recognition test. Journal of Clinical and Experimental Neuropsychology, 15(2), 170-182.

Binder, L. (1993b). An abbreviated form of the Portland digit recognition test. The Clinical Neuropsychologist, 7(1), 104-107.

Binder, L. (1992). Deception and malingering. In A. Puente, & R. McCaffrey (Eds.), Handbook of neuropsychological assessment (pp. 353-374). New York: Plenum.

- Binder, L. (1991). Malingering and deception. In A. Puente & R. McCaffrey (Eds.), Psychobiological factors in clinical neuropsychological assessment (pp. 353-374). New York: Plenum.
- Binder, L. (1990). Malingering following minor head trauma. The Clinical Neuropsychologist, 4(1), 25-36.
- Binder, L., & Pankratz, L. (1988). Neuropsychological evidence of a factitious memory complaint. Journal of Clinical and Experimental Neuropsychology, 9(2), 167-171.
- Binder, L., Villanueva, M.R., Howieson, D., & Moore, R.T. (1993). The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. Archives of Clinical Neuropsychology, 8, 137-147.
- Binder, L., & Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3(2), 175-181.
- Brandt, J. (1988). Malingered amnesia. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 65-83). New York: Guilford Press.
- Brandt, J. (1992). Detecting amnesia's imposters. In L. Squire and N. Butter (Eds.), Neuropsychology of memory (2nd ed) (pp. 156-65). New York: Guilford Press.
- Bruhn, A.R., & Reed, M.R. (1975). Simulation of brain damage on the Bender-Gestalt Test by college students. Journal of Personality Assessment, 39, 244-255.

- Cradock, M.M., Gfeller, J.D., & Falkenhain, M.A. (1994, October). Detecting feigned memory deficits with RAVLT and the Recognition Memory Test. Presentation at the 14th annual conference of the National Academy of Neuropsychology, Orlando, FL.
- Cullum, C.M., Heaton, R.K., & Grant, I. (1991). Psychogenic factors influencing neuropsychological performance: Somatoform disorders, factitious disorders, and malingering. In H.O. Doerr, & A.S. Carlin (Eds.), Forensic neuropsychology: Legal and scientific bases (pp. 141-171). New York: Guilford Press.
- Davidson, H., Suffield, B., Orenczuk, S., Nantau, K., & Mandel, A. (1991, February). Screening for malingering using the memory for fifteen items test (MFIT). Poster presented at the 20th annual meeting of the International Neuropsychological Society, San Antonio, TX.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (1987). The California Verbal Learning Test. San Antonio, TX: The Psychological Corporation.
- Faust, D., Hart, K., Guilmette, T.J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 56, 578-582.
- Faust, D., Hart, K., Guilmette, T.J., & Arkes, H.R. (1988). Neuropsychologist's capacity to detect adolescent malingerers. Professional Psychology: Research and Practice, 19(5), 508-515.
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-mental state.

- Journal of Psychiatric Research, 12, 189-198.
- Franzen, M.D., Iverson, G.L., & McCracken, L.M. (1990). The detection of malingering in neuropsychological assessment. Neuropsychology Review, 1(3), 247-279.
- Frederick, R.I., & Foster, H.G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3(4), 596-602.
- Gansler, D., Tombaugh, T., Peterson-Moczynski, N., & Rees, L. (1995, October). Test of Memory Malingering (TOMM): Initial validation in a traumatic brain injury cohort. Poster presented at the 15th annual meeting of the National Academy of Neuropsychology, San Francisco, CA.
- Gfeller, J.D., Cradock, M.M., & Falkenhain, M.A. (1994, October). Detecting feigned neuropsychological impairment with the Seashore Rhythm Test. Poster presented at the 14th annual meeting of the National Academy of Neuropsychology, Orlando FL.
- Goebel, R.A. (1983). Detection of faking on the Halstead-Reitan neuropsychological test battery. Journal of Clinical Psychology, 39, 731-742.
- Goldberg, J.O., & Miller, H.R. (1986). Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. Journal of Clinical Psychology, 42(5), 792-795.
- Greene, R.L., (1988). Assessment of malingering and defensiveness by objective personality inventories. In R. Rogers (Ed.), Clinical assessment of

malingering and deception. Guilford Press: New York

- Greiffenstein, M.F., Baker, W.J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. Psychological Assessment, 6(3), 218-224.
- Grosz, H., & Zimmerman, J. (1965). Experimental analysis of hysterical blindness. Archives of General Psychiatry, 13, 225-260.
- Guilmette, T.J., & Giuliano, A.J. (1991). Taking the stand: Issues and strategies in forensic neuropsychology. The Clinical Neuropsychologist, 5(3), 197-219.
- Guilmette, T.J., Hart, K.J., Giuliano, A.J. (1993). Malingering detection: The use of a forced-choice method in identifying organic versus simulated memory impairment. The Clinical Neuropsychologist, 7(1), 59-69.
- Guilmette, T.J., Hart, K.J., Giuliano, A.J., & Leininger, B.E. (1994). Detecting simulated memory impairment: Comparison of the Rey fifteen-item test and the Hiscock forced-choice presentation. The Clinical Neuropsychologist, 8(3), 283-294.
- Guilmette, T.J., Sparadeo, F.R., Whelihan, W., & Buongiorno, G. (1994). Validity of neuropsychological test results in disability evaluations. Perceptual and Motor Skills, 78, 1179-1186.
- Hawk, G.L., & Cornell, D.G. (1988). MMPf profiles of malingers diagnosed in pretrial forensic evaluations. Journal of Clinical Psychology, 45(4), 673-678.
- Hayes, J.R., Emmons, J., & Lawson, K.A. (1993). Psychiatric norms for the Rey 15-Item visual memory test. Perceptual and Motor Skills, 76, 1331-1334.

- Heaton, R.K., Smith, H.H., Lehman, R.A., & Vogt, A.T. (1978). Prospects for faking believable deficits on neuropsychological testing. Journal of Consulting and Clinical Psychology, 46, 892-900.
- Hiscock, C., Layman, L., & Hiscock, M. (1994). Cross-validation of two measures for assessing feigned mental incompetence in male prison inmates. Criminal Justice and Behaviour, 21(4), 443-453.
- Hiscock, M., & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. Journal of Clinical and Experimental Neuropsychology, 11(6), 967-974.
- Hiscock, C.K., Rustemier, P.J., & Hiscock, M. (1993). Determination of criminal responsibility: Application of the two-alternative forced-choice stratagem. Criminal Justice and Behaviour, 20(4), 391-405.
- Hubley, A.M., McIntyre, N., & Tombaugh, T.N. (1995). Unpublished data. Carleton University, Ottawa.
- Knight, J.A. Basic Knowledge Questionnaire. Personal Communication, 1996.
- Knight, J.A., & Meyers, J.E. (1994, October). Malingered memory performance patterns on the Rey-Osterrieth Complex Figure Test. Paper presented at the 14th annual meeting of the National Academy of Neuropsychology, Orlando, FL.
- Larrabee, G.J. (1992). On modifying recognition memory tests for detection of malingering. Neuropsychology, 6(1), 23-27.

- Lezak, M. (1995). Neuropsychological assessment (3rd ed). New York: Oxford University Press.
- Martin, R.C., Haut, J.S., Stainbrook, T., & Franzen, M.D. (1994, October). Preliminary normative data for objective measures to detect malingered neuropsychological deficits in a population of adolescent patients. Poster presented at the 14th annual meeting of the National Academy of Neuropsychology, Orlando, FL.
- Mensch, A.J., & Woods, D.J. (1986). Patterns of feigning brain damage on the LNNB. International Journal of Clinical Neuropsychology, 8, 59-63.
- Millis, S.R. (1994). Assessment of motivation and memory with the Recognition Memory Test after financially compensable mild head injury. Journal of Clinical Psychology, 50(4), 601-605.
- Millis, S.R. (1992). The recognition memory test in the detection of malingered and exaggerated memory deficits. The Clinical Neuropsychologist, 6(4), 406-414.
- Millis, S.R., & Putnam, S.H. (1994). The Recognition Memory Test in the assessment of memory impairment after financially compensable mild head injury: A replication. Perceptual and Motor Skills, 79, 384-386.
- Millis, S.R., & Ricker, J.H. (1994, February). The California Verbal Learning Test in the Detection of Feigned Memory Impairment. Poster presented at the annual meeting of the International Neuropsychological Society, Cincinnati, OH.
- Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification

- of malingered head injury on the Wechsler Memory Scale-Revised. Psychological Assessment, 5(1), 34-40.
- Morgan, S.F. (1991). Effect of true memory impairment on a test of memory complaint validity. Archives of Clinical Neuropsychology, 6, 327-334.
- Nies, K.J., & Sweet, J.J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. Archives of Clinical Neuropsychology, 9(6), 501-552.
- Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. Journal of Consulting and Clinical Psychology, 47(2), 409-410.
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficit. Perceptual and Motor Skills, 57, 367-372.
- Pankratz, L. (1988). Malingering on intellectual and neuropsychological measures. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 169-192). Guilford Press: New York.
- Pankratz, L., & Erickson, R.D. (1990). Two views of malingering. The Clinical Neuropsychologist, 4(4), 379-389.
- Pankratz, L., Fausti, S.A., & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. Journal of Consulting and Clinical Psychology, 43(3), 421-422.
- Paul, D.S., Franzen, M.D., Cohen, S.H., & Fremouw, W. (1992). An investigation into the reliability and validity of two tests used in the detection

- of dissimulation. International Journal of Clinical Neuropsychology, 14, 1-9.
- Philpott, L.M., & Boone, K.B. (1993, February). The effects of cognitive impairment and age on two malingering tests: An investigation of the Rey memory test (RMT) and the Rey dot counting test (RDC) in Alzheimer's patients and normal middle aged/older adults. Poster presented at the 22nd annual meeting of the International Neuropsychological Society, Cincinnati, OH.
- Prigatano, G.P., & Amin, K. (1993). Digit memory test: Unequivocal cerebral dysfunction and suspected malingering. Journal of Clinical and Experimental Neuropsychology, 15(4), 537-546.
- Reber, A.S. (1985). Dictionary of Psychology. New York: Penguin Books Ltd.
- Resnick, P.J. (1988). Malingering of post-traumatic disorders. In R.Rogers (Ed.), Clinical assessment of malingering and deception (pp. 84-103). New York: Guilford Press.
- Rogers, R. (Ed.). (1988). Clinical Assessment of Malingering and Deception. New York: Guilford Press.
- Rose, F.E. (1993). Malingered performance on four measures of malingering, the California Verbal Learning test and the effects of coaching. Unpublished doctoral dissertation, University of Montana.
- Rose, F.E., Hall, S., & Szalda-Petree, A.D. (1995). Portland Digit Recognition Test - computerized: Measuring response latency improves the detection of malingering. The Clinical Neuropsychologist, 9(2), 124-130.

- Rose, F.E., Petree, A.D., & Hall, S. (1993, February). The detection of malingering using a computerized version of the Portland Digit Recognition Test: The utility of response latency. Poster presented at the annual meeting of the International Neuropsychological Society, Cincinnati, OH.
- Ruff, R.M. (1988). The Ruff Figural Fluency Test. San Diego, CA: Neuropsychological Resources.
- Schmidt, J.P., & Tombaugh, T.N. (1995). Learning and Memory Battery (LAMB). Toronto: Multi-Health Systems Inc.
- Schretlen, D.J. (1988). The use of psychological tests to identify malingered symptoms of mental disorder. Clinical Psychology Review, 8, 451-476.
- Schretlen, D.J., & Arkowitz, H. (1990). A psychological test battery to detect prison inmates who fake insanity or mental retardation. Behavioural Sciences and the Law, 8, 75-84.
- Schretlen, D.J., Brandt, J., Krafft, L., & Van Gorp, W.G. (1991). Some caveats in using the Rey 15-Item Memory Test to detect malingered amnesia. Psychological Assessment, 3(4), 667-672.
- Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.
- Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. Journal of Clinical and Experimental

Neuropsychology, 16(3), 472-481.

Slick, D., Hopp, G., Strauss, E., Spellacy, F., & Hunter, M. (1995, February).

Use of the MMPI-2 and the Victoria Revision of the Hiscock-Digit Memory

Test for evaluating validity of closed-head injury symptoms. Poster presented

at the 24th annual meeting of the International Neuropsychological Society,

Seattle, Wa.

Standing, J., Conezio, J., & Haber, R.N. (1970). Perception and memory for

pictures: Single-trial learning of 2500 visual stimuli. Psychon. Sci., 19(2), 73-

74.

Strauss, E., Spellacy, F., Hunter, M., & Berry, T. (1994). Assessing believable

deficits on measures of attention and information processing capacity.

Archives of Clinical Neuropsychology, 9(6), 483-490.

Theodor, L.H., & Mandelcorn, M.S. (1973). Hysterical blindness: A case report

and study using a modern psychophysical technique. Journal of Abnormal

Psychology, 82, 552-553.

Tombaugh, T.N. (1996). Test of Memory Malingering. Toronto: Multi-Health

Systems.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity

considerations in the neuropsychological evaluation of mild head injury.

Journal of Clinical and Experimental Neuropsychology, 15(4), 578-590.

Warrington, E.K. (1984). Recognition Memory Test: Manual. Berkshire, U.K.:

NFER-Nelson.

- Wasyliw, O.E., & Cavanaugh, J.L. (1989). Simulation of brain damage: Assessment and decision rules. Bulletin of the American Academy of Psychiatry and Law, 17(4), 373-386.
- Wechsler, D. (1981). The Wechsler Adult Intelligence Scale-Revised. San Antonio, Tx: The Psychological Corporation.
- Wechsler, D. (1987). The Wechsler Memory Scale-Revised. San Antonio, Tx: The Psychological Corporation.
- Wiggins, E.C., & Brandt, J. (1988). The detection of simulated amnesia. Law and Human Behaviour, 12(1), 57-78.
- Zielinski, J.J. (1994). Malingering and defensiveness in the neuropsychological assessment of mild traumatic brain injury. Clinical Psychology: Science and Practice, 1(2), 169-183.

Appendix A

Tests of Malingering/Simulation:

1. Test of Memory Malingering (TOMM):

The TOMM consists of two learning trials and a retention trial. On each learning trial, 50 line drawings of common objects (e.g., iron, boat, hat) are shown to the person at a 4 second exposure rate. 50 multiple choice recognition tests are then administered (1 target, 1 distractor). The person is asked to point to the picture that was previously shown to them. Feedback is provided and if an incorrect response is made, the examiner will identify the correct one. Following a 15 minute delay the retention trial is administered; subjects are asked to point to the picture that they were previously shown. The number of items correctly remembered for each trial is recorded (range 0 to 50). Administration time is approximately 20 minutes.

2. Abbreviated-Hiscock Forced Choice Procedure (A-HFCP):

The A-HFCP is a recognition task. Subjects are shown a card with a 5 digit number on it. They are asked to read the number out loud and to try to remember it. The card is removed and following a 5 second delay a second card is presented. This second card has two, 5 digit numbers printed on it and the subject is asked to point to the one s/he had just seen. Three sets of 12 numbers are presented to the subject. As stated, during the first set the subject waits 5 seconds between presentation and recall. For the second and third sets, the interval increases to 10 & 15 seconds, respectively. The number of items correctly recalled for each set is recorded (range 0 to 12). Administration time is approximately 10 to 15 minutes.

3. Knight's Basic Knowledge Questionnaire (BKQ):

The questionnaire consists of 16 questions pertaining to everyday knowledge that all but the most severely impaired should be able to answer correctly (e.g., what colour is a stop sign?). Participant's effort is also rated by the observer. Administration time is approximately 5 minutes.

Neuropsychological Tests:

1. Mental Control Subtest of the Wechsler Memory Scale-Revised (WMS-R):

Participants are asked to count backward from 20 to 1, repeat the alphabet and to count by 3s beginning with 1. Time taken to complete each part and errors made are recorded. Administration time takes approximately 3 minutes.

2. Digit Span Subtest of the Wechsler Memory Scale-Revised (WMS-R):

Participants are asked to repeat a series of digit sequences increasing in number from three to eight (digit span forward). Subjects are then asked to repeat a string of digits in the reverse order (digit span backward). Administration time takes approximately 5 minutes.

3. Visual Reproduction subtest of the Wechsler Memory Scale - Revised (VRI & II):

Participants are shown four designs, one at a time for 10 seconds. Following each exposure, subjects are asked to draw the design. Following a 30 minute delay subjects are asked to draw all four designs from memory. Scoring follows standardized criteria. Administration time takes approximately 10 minutes.

4. California Verbal Learning Test (CVLT):

The CVLT is an auditory verbal learning test. A list of 16 semantically related shopping items is presented to the subject and the subject is asked to recall the items. The list is presented for 5 trials. A second list of items is presented as a means of interference. Immediate and delayed cued and free recall is assessed, as is recognition. Administration takes approximately 30 minutes.

5. Trail Making Test

Participants are asked to connect 25 encircled numbers, in sequence, randomly distributed throughout a page as fast as they can (Trails A). Subjects are then instructed to connect 25 encircled numbers and letters in alternating sequence, as fast as they can (Trails B). Time to complete each part is recorded. Administration takes approximately 5 minutes.

6. Finger Tapping Test (FTT):

Participants are asked to tap as fast as they can for 10 seconds using the index finger of their preferred hand. Five trials are given with a brief break in between each trial. The same procedure is followed for the non-preferred hand. Mean number of taps is recorded over the 5 trials for each hand. Administration time takes approximately 5 minutes.

7. Information subtest from the WAIS-R:

Participants are asked a series of general questions that reflect a breadth of knowledge and academic achievement. Total correct responses is recorded. Administration time takes approximately 5 minutes.

8. Ruff Figural Fluency Test:

Participants are presented with a sheet of paper with a specific dot pattern printed onto it. They are asked to draw as many different design patterns, by connecting the dots, as quickly as they can within a one minute period. They are presented with five separate patterns. Total correct for each of the five patterns is recorded. Administration time takes approximately 10 minutes.

9. Word Fluency (FAS & Animal Naming):

Participants are asked to produce as many words as quickly as they can that begin with a particular letter (F, A & S). They have one minute to complete the task for each letter. Responses are recorded and summed over the three letters. Following this task, subjects are then asked to name as many different animals as quickly as they can for one minute. Number of responses are recorded. Administration time takes approximately 5 minutes.

Appendix B

Subject Number: _____

RATING SCALE

1. Indicate how hard you tried to follow the instructions you were given at the beginning of the experiment by circling a number that best describes your effort.

1	2	3	4	5
didn't try at all		tried moderately hard		tried very hard

2. Indicate how successful you think you were in deceiving the examiner by circling a number that best describes your belief.

1	2	3	4	5
not successful at all		somewhat successful		very successful

3. Do you have any personal knowledge of, or personal experience with, brain-damage?

_____ No
 _____ Yes, family, friends
 _____ Yes, myself
 _____ Yes, courses, books, TV
 _____ Yes, other

4. In your estimation, how convincing do you think you were in your attempts to respond like someone with brain-damage on each of the tests listed below. Place a check between the hashmarks on the scale for each test.

	Not Convincing 0%	50%	Very Convincing 100%
1. Counting 20-1, alphabet, counting by 3s, spelling words backwards.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
2. Learning a grocery shopping list (drill, plums, vest, parsley etc).	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
3. Drawing designs from memory after looking at them for 10 sec.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
4. Connecting circles containing numbers and letters.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
5. Rapidly pressing a key.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
6. Repeating a series of numbers forward and backward.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
7. Recognizing pictures of 50 different objects (2-item multiple choice).	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
8. Answering Information questions (Where does sun rise? Four men who were Presidents of the US?).	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
9. Connecting dots to make designs.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
10. Saying words beginning with the letters F, A, S. and naming animals.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
11. Selecting 5-digit numbers after 5, 10 and 15 sec.	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>
12. Answering a set of questions about everyday basic knowledge (Color of stop signs?).	<div><div></div></div>	<div><div></div></div>	<div><div></div></div>

5. How well do you think each of the following tests would work in catching someone trying to fake brain-damage? In other words, rate each test on how easy it would be to fake.

	No obvious way to fake this test	Test was too simple to fake	Test too difficult to fake	Obvious way to fake this test.
1. Counting 20-1, alphabet, counting by 3s, spelling words backwards.	_____	_____	_____	_____
2. Learning a grocery shopping list (drill, plums, vest, parsley etc).	_____	_____	_____	_____
3. Drawing designs from memory after looking at them for 10 sec.	_____	_____	_____	_____
4. Connecting circles containing numbers and letters.	_____	_____	_____	_____
5. Rapidly pressing a key.	_____	_____	_____	_____
6. Repeating a series of numbers forward and backward.	_____	_____	_____	_____
7. Recognizing pictures of 50 different objects (2-item multiple choice).	_____	_____	_____	_____
8. Answering Information questions (Where does sun rise? Four men who were Presidents of the US?).	_____	_____	_____	_____
9. Connecting dots to make designs.	_____	_____	_____	_____
10. Saying words beginning with the letters F, A, S, and naming animals.	_____	_____	_____	_____
11. Selecting 5-digit numbers after 5, 10 and 15 sec.	_____	_____	_____	_____
12. Answering set of questions about everyday basic knowledge (Color of stop signs?).	_____	_____	_____	_____

6. The purpose of some tests you took was to catch people who might try to fake brain damage. Rate each test on whether you thought it was trying to do this by placing a check between the hashmarks on the scale for each test.

Definitely NOT trying to catch fakers - 5	I'm not sure 0	Definitely trying to catch fakers +5
--	-------------------------	---

1. Counting 20-1, alphabet, counting by 3s, spelling words backwards.	
2. Learning a grocery shopping list (drill, plums, vest, parsley etc).	
3. Drawing designs from memory after looking at them for 10 sec.	
4. Connecting circles containing numbers and letters.	
5. Rapidly pressing a key.	
6. Repeating a series of numbers forward and backward.	
7. Recognizing pictures of 50 different objects (2-item multiple choice).	
8. Answering Information questions (Where does sun rise? Four men who were Presidents of the US?).	
9. Connecting dots to make designs.	
10. Saying words beginning with the letters F, A, S. and naming animals.	
11. Selecting 5-digit numbers after 5, 10 and 15 sec.	
12. Answering set of questions about everyday basic knowledge (Color of stop signs?).	

ORAL DEBRIEFING

1. What did you do, if anything, to prepare for taking the tests? That is, how did you go about preparing to fake the tests -- to fool the professor?

2. Did you use the same strategy for each test: Yes or No.
 - a. If yes, what strategy did you use.

 - b. If no, please tell me the different strategies for each of the following tests.

1. Counting 20-1, alphabet, counting by 3s, spelling words backwards.
2. Learning a grocery shopping list (drill, plums, vest, parsley etc).
3. Drawing designs from memory after looking at them for 10 sec.
4. Connecting circles containing numbers and letters.
5. Rapidly pressing a key.
6. Repeating a series of numbers forward and backward.
7. Recognizing pictures of 50 different objects (2-item multiple choice).
8. Answering Information questions (Where does sun rise? Four men who were Presidents of the US).
9. Connecting dots to make designs.
10. Saying words beginning with the letters F, A, S, and naming animals.
11. Selecting 5-digit numbers after 5, 10 and 15 sec.
12. Answering set of questions about everyday basic knowledge (Color of stop signs?).

Pagination Error
Text Complete

Erreur de pagination
Le texte est complet

National Library of Canada
Canadian Theses Service

Bibliothèque nationale du Canada
Service des thèses canadiennes

Appendix C

**Study 1
Oral Debriefing
University Simulators**

Mental Control:

mix myself up but some were too basic to screw up
 tried to have a little confusion
 just did it; paused at times
 slowed down
 go slower
 made mistakes on counting by 3s because it was a tougher task than other two
 too similar to fake; tried to go more slowly
 slowed down
 took a little longer; went slower
 slowed down
 left off some numbers; didn't know what to do
 slowed down; answer wrong
 little slower
 slower
 go slower in middle of sequence
 numbers in teens thought would be harder therefore when slower; 1-10 okay; slower
 on alphabet
 trying to go for more mathematical & weaker on arts side (right side of brain)
 I did it normally, didn't think they'd have a problem with that
 nothing there - it's hard to fake general knowledge; too well learned
 slowed down - told myself it was hard
 made it look like I stumbled (during sequencing)
 tried to fake it - pretend I didn't know the numbers; go slower
 slower, hesitation
 I did okay - slower than normal
 tried to be confused; reversing some of the numbers, letters; go slowly
 just to go slow because you can't really fake that one
 I went slower, left some out
 slow
 some I muffled (the one's without teens); left out some; slower

California Verbal Learning Test:

tried to keep the last ones you said in my memory and forget the initial ones
 confuse the items and mix it up, especially during recognition - if similar or sounded

the same I would try to say yes
 was difficult because it was repeated - I confused a few items, threw in a few items
 that may have sounded similar, mixing the lists
 After a while I realized that there were categories & I figured that a BD person
 wouldn't notice them so I tried to block them (from memory)
 that was hard; tried to forget
 tried to get 1st and last words of list because those are generally remembered whereas
 middle ones aren't; start grouping them, decreased memory overall
 tried to stick with 1 syllable words; didn't listen would let my mind go blank/wander
 to act slow, little memory capacity & function
 associated obvious fruits; remembered last two you said
 picked one or two words always said those; ignored others
 tried to remember same thing each time; mix the lists; added some in
 I would mix the tues list w/ mon.; I didn't fully memorize the list; left some off
 tried to fumble around in the middle; I'd remember the beginning & the end
 first two times tried to remember the items - then I categorized it; plums, pliers
 always together
 just tried to remember certain things - like I didn't absorb anymore
 tried to learn it pretty well and mix up a few or forget a few (mixed up days -
 mon/tues)
 its easy to forget things; added items in; tried to categorize items
 at first it was just blanks; I remembered the first and last, didn't connect to themes;
 didn't concentrate
 added hammer (relates to tools); items that had salience to me (eg tools) I kept in and
 if I liked it (eg apricots)
 didn't mention all the items; randomly left off items; tried to pick the same items to
 make it consistent
 slower; randomly left off items
 I pretty much did my best; I didn't want to feign memory deficits; attention type task
 I did poorly; no categories
 I remembered what I could then I threw in some odd things like I just remembered
 the tools etc
 I really didn't fake on it; it was difficult I tried to do well
 basically tried to keep fruits and spices together; I kept saying word ignored the two
 others; I would only say 1 item
 I left out things & tried to keep it consistent; tried to forget more overtime (poor
 retention)
 stuck with words with "p" or food; left off others; stay with a theme or consistent

Visual Reproduction Subtest:

I did it like I didn't know space very well or I reversed it
 Confused shapes and lines, omit details, didn't differentiate size different on
 rectangles on Card C

as difficulty increased I made minor errors, leaving out details, switching sides
 I wasn't sure, I made 1 mistake in each drawing, changing position of things
 tried to remember as little as possible; omit details, on Copy A didn't know what to
 do - switched left and right on Card D
 tried to change some elements - exaggerate some of elements (eg Card D); leave off
 some other elements
 tried to exaggerate them, particularly card D didn't focus on detail; left it off
 to show little memory capacity and function
 I had no idea what to do; drew it poorly; changing some of them
 left off detail; for Card C, filled in the detail - no pattern
 tried to make them similar - something small I would change - nothing too obvious; I
 wasn't sure myself
 tried to forget how they look; what elements were assoc. with each one
 to be as honest as possible tried as hard as I could; if I failed I'd get burned
 changed details - direction of flags; left off some details
 not really grouping shapes and designs; get basics but not details
 drew them a little different - mixed up some of details (changing direction of flags)
 VR I too simple to forget; changed shape a little; honestly forgot one at end
 need to pick up a few basic things - tried to change some details; turn them around
 tried to change triangle; left off some details, changed shape etc
 tried not to make them accurate reproductions; changed details
 I left some things off; unattached circles etc (leaving off details)
 only thing I tried to do I didn't the lines poorly (dexterity decreased)
 tried to see an object in the fig - then drew it - eg baseball diamond for card A
 I had trouble with that one too; last two were hard. Sort of tried
 left off some details
 reversed some details; left off some others
 tried to; left off some details; changed them

Trail Making Test:

trying to mix myself up, be slower, made errors
 purposely go in correct order instead of alternating between sets
 tried to go slower; it was difficult to alternate, therefore thought I took alot longer
 to do
 I don't know, just made mistakes, went slower
 just tried to do it very slowly, its something you'd know how to do
 slowed down
 slowed it down; then wouldn't chunk (1A,2B) would start at beginning of sequence at
 various points in test
 slowed it down
 tried to go for a pattern, what looked like obvious next choice I would make it
 winged it
 I wasn't sure, picking up my pencil alot; tried not to think about where next one was

trying to be slower, numbers and letters got confusing
 honest about that
 just did it, no problem with numbers (A), Part B, made mistakes
 randomly went to any circle, whatever was closer
 I did it normally, maybe a little slower, it was easy
 tried to do it; didn't fake it, did it normally
 tried to confuse my mind, block myself, go slow
 tried to go within the sequence (number to number, letter to letter)
 went slower
 because I was timed, I took my time, showed difficulty finding the next
 number etc
 problems with attention span, I tried to do it poorly, more on B than A, went more
 slowly
 kind of easy to fake it - go from 1 to 2 instead of 1 to A etc
 that was easy - you couldn't fake it - took my time
 I'd go slow, then mess up
 did it slow

Finger Tapping Test:

tried to go slower with both hands and particularly with left
 slower, became more frustrated
 not much, slower than I would normally do it; I wasn't trying
 I did my best on that one; I didn't think that would be affected
 went slower - didn't know what to do
 slowed down & discontinued in places
 tried to do left hand faster than right, purposely did right hand slower
 to act slow
 not pressing all the way down, slower; missing with Lh on purpose
 slowed it down; conflicting results, R slower than L
 no idea; tried to keep it around the same number
 that's my normal rate, didn't fake on it; wasn't sure if BD person would do poorly on
 that
 tried to do that as well as I could
 tried to keep it constant
 tried to tap more with right hand (connected to left hemisphere) - left hand (right
 hemisphere damage) therefore wouldn't do it
 just did it; tried to keep it consistent
 didn't fake it; it was hard, did it best of my ability
 that was hard; my finger didn't work; I didn't really try
 slowly and consistent - keep numbers approx. equal
 slower
 tried to go slower, a consistent decrease between L & Rh; you got me on that one
 tried to do poor on tests of dexterity

tried to change my time so it would be all over the place
 no real strategy there; could have gone faster on my Lh
 went slow then fast - inconsistent
 did it slow
 slowed down on my Lh; kept Rh consistent

Digit Span Subtest:

I reversed the numbers
 on the longer strings, I'd reverse the numbers
 didn't do too much, I allowed myself to get confused
 I was trying to think of chunking & that you can remember up to 7 digits, so a TBI
 person might remember only 6 digits
 omit #'s or change the order of #'s, saying it slowly
 changed some of the #'s, didn't want to look too stupid, I thought it might catch me
 purposely messed up the order
 didn't have a plan
 arbitrary/randomly changed numbers
 mixed up the order
 I have a problem with trying to repeat #'s, my memory works better with
 images/objects - #'s don't stick with me
 I did that sincerely, a couple of times I'd mix order up
 on DSB, just didn't say it in right sequence etc
 don't know, just changed it
 not paying attention when you were saying them
 easy to fake, changed the #'s, tried not to be alert
 tried to confuse my brain (go blank, distracted)
 slip in 5s instead of 9s, pretended I had problems remembering 9s
 took longer to repeat the numbers: change some of the numbers
 I scrambled the #'s, to show I was disorganized
 that's cognitive, so I didn't do too poorly (didn't believe him when I told him it was
 attention task)
 just reversed a couple of #'s
 tried to remember them but I couldn't
 I don't remember
 did it slow, then forgot some #'s
 put them in the wrong order

Test of Memory Malinger:

trying to pick out the ones with straight lines instead of detail; remember the ones at
 the end not the beginning
 tried to remember interesting or amusing pictures, if pictures had same shape I'd pick
 the wrong one

deliberate mistakes, it seemed easy, if looked similar I'd pick the wrong one, I found it very easy
 I was intentionally making mistakes; it was too easy, I don't know if a BD person would forget the items
 I just randomly lied
 I would try to pick the simpler pictures; picture the one from the CVLT
 purposely choose the wrong one, I messed up I should have picked the less complex one (even if it was right), sometimes I picked the one that was similar to the correct one
 tried to create a pattern, anything that would cut (knife etc), tried to develop little patterns (vest from CVLT)
 a lot of linear shapes - picked that no matter what it was; not random, more of a pattern in my choices
 tried to remember the latter pics - premise that LTM was shot
 tried to find pictures that looked similar to other ones; tried to mix them up
 I couldn't remember them - you were turning them too fast
 at beginning deliberate pointing to wrong item; pointed to any item that was on CVLT, second time I saw it I got more right and same with the last time I saw it
 random - If I said no to picture B I'd point to A etc
 intentionally picked wrong ones; was really hard to fake, simple everyday pictures; seeing where things are alike or similar & picked that; same family
 concentrate on a few pictures to really remember; others I just ignored
 tried to be consistent; if recognized 1st time I try to recognize it again
 tried to leave out a few; guess the wrong ones
 I found it easy - randomly pointed to wrong ones
 was that a test to ferret out the fakers? it was easy - picked similar, items (still faked on it)
 I counted the # wrong I was getting & kept it fairly close on each trial - random
 I tried to pick out a few key ones - couldn't remember all of them
 trying to do the first set (trial 1) a certain number wrong & progressively get worse
 left some out; tried to be consistent; increased for Trial 2 then decrease for delay
 randomly leaving off some; only picking top item (a)

Information subtests:

tried to give simple answers or I don't remember answers
 tried to make it vague, non-specific
 I just answer them almost to my abilities; I left out somethings (I don't know answers)
 I stared answering them and then I realized that some questions come from books and some from experience - I tried to answer the experience questions correctly and the ones from books incorrectly
 I just lied; I don't know answers

didn't really fake this one; thought it was too easy
 I purposely choose the wrong answer; randomly choose the wrong one
 tried to show little memory capacity and function
 I figured I'd just block out memory; stuff you wouldn't hear every day
 I didn't think of any strategy
 don't know responses, answered incorrectly
 history was never my strongest topic, listed more recent ones; if question was
 something to do with science I'd answer slowly or give wrong answer
 I did that as well as I could; I found it fun; if you went overboard it would be obvious
 I would assume that you'd take age into consideration; last or more recent
 questions...alot of them were obvious
 tried to answer the factual info correctly; wouldn't answer opinion ones
 a few I answered wrong; the ones that warrant too common
 can't fake general knowledge (very difficult to)
 obvious questions I answered, not so obvious questions I didn't answer
 trying to go for older memory; left off Clinton; I don't know answers
 I didn't fake that - too easy or basic
 lying on them; I don't know
 that's more .. a different part of the brain so I didn't really do too poorly on it
 just picked an answer out of the air
 I could have tried my best and not got half of them; tried to do well
 basically straight forward - anyone could mess it up; if someone really trying to fake
 they would really screw up
 just said I don't know
 couldn't think of a way to screw it up; sometimes said I don't know

Ruff Visual Fluency Test:

I did repetitions on that, went slower than I could have gone
 went very slow, made it look tedious; repeated some designs
 did it slowly, I started to repeat designs
 I did my best on this one; I figured it was a test of creativity
 didn't know what I was suppose to do
 tried to keep the continuity from 1 design to the next; my own rule to follow
 I was fairly systematic started at one dot etc and did all possibilities of that design
 to act slow
 just connected dots, tried not to pay attention
 I just did whatever; some tried to make it different, and some made it the same
 repeating designs
 just to make sure I wasn't copying the designs (repeating) which slowed me down
 repeated 1/2 designs, basically tried to do well
 trying to make different designs but 2-3 were the same (at the beginning and the end)
 just repeated the designs (repeat 3/4 patterns)
 did it slower than I normally would; not consistent

repeated some designs
 I really concentrated on this one; tried to do it fast so I'd make mistakes
 tried to screw it up; sometimes repeated designs
 I didn't do as many as a normal person would have
 I went slow
 that one had to do with attention; when you started confusing the dots (adding
 interference) my performance decreased
 random, tried to repeat some
 didn't try to do well, doing the most simple patterns first, going slow
 kept it as easy as possible, the patterns, going slow
 was simple, going slow
 kept it simple, slow, repeat designs

FAS/Animal Naming:

go slow, repeat myself
 that was recall so after 4/5 I'd try not to do anything for along time
 went a little bit slower; not too slow though
 did my best; I didn't think it would effect a BD person
 simple, said it slowly; its everyday words, how could you forget?
 tried to stick to the same routine if started with sa would continue
 that was more difficult; I tried to pick 1/2 syllable words didn't want to use larger/
 more complex words
 to go slow
 tried to repeat the words, go slow
 make like I couldn't differentiate between silent p and t sound (eg phone)
 saying words that were phonemically similar (eg celery = s)
 purposely being slow for animal name; with letters I was a bit slower
 inserted accident, automobile, tired to think of words that my bring back some
 memories of the person
 just said everyday type of words; no real strategy
 name some and then repeat them
 I tried on that one
 was easy to fake; went slower
 blanked out my brain
 didn't get this; named things that were familiar to me, went slower
 just named a few; going slower
 very slow
 tried not to follow instructions; new learning is down; think of more phonetic words
 (eg ph)
 said whatever came to mind
 tried to remember as many as I could; hard when your put under the gun; if I had a
 few days I'd do okay

tried not to get words that were too elaborate; slow
 tried to be slow
 not much faking; slow

Abbreviated Hiscock Forced Choice Procedure:

random; I'd pick the wrong one
 the longer the interval the worse my performance
 no matter how hard I tried to forget I couldn't; it was really easy - the #s were very
 easy to pick up; assumed that my performance would be good
 started off doing it right, iconic memory is immediate but then I realized I could
 verbally repeat the #s; randomly got some wrong; as interval increased I
 thought I should do worse but then I thought I'd have more time to rehearse;
 wasn't sure
 randomly lied; the longer the interval the more I thought I should forget
 thought it was really simple, effort to fake; too simple to fake
 no mistakes with 5s interval, increased errors as interval increased; the distractor #s
 were not similar; random mistakes (numbers too dissimilar)
 too easy; I don't think you could fake it
 always picked string with 6/9; missed more on last one (15s) than other two intervals
 for first set did semi decent job (a perfect job); as interval increased I did
 worse but for last set I did it arbitrarily to throw off the results (50-50 chance)
 mixed up a couple; tried to do it more at end with longer wait I increased the errors
 with 5s interval answered correctly, for longer duration I was purposely trying to
 forget; looking around room etc
 tried to act like my attention span was short; looking around room
 I didn't think anybody could get those wrong; did best
 I just picked the one on the left side of card
 that was easy; I figured no one would get it wrong
 too easy to fake; tried to fake but the numbers are too different from each other
 problem faking that; that's STM
 it was too easy to fake; just randomly choose wrong one; you'd have to be a dolt not
 to recognize them; they start with different numbers
 didn't try to change much; felt the interval wouldn't make a difference
 I didn't do much; I made mistake because I was tired; didn't try to fake on it; its too
 easy
 very simple; I did worse than I should have; almost anyone could have done it right
 first set right; second set 30% wrong and last one 50% wrong; STM okay; worse
 memory as interval increased
 tried to remember but had difficulty after 10-15 sec.
 got progressively worse
 tried to pick less and less the right ones as interval increased
 didn't fake it; remembering 1st # in each one; too obvious to fake

Basic Knowledge Questionnaire:

did basic
 with obvious ones I tried to answer them quick and correct; other ones I tried to be
 vague
 decided to go with correct items; I wasn't sure what the questions were getting at; I
 thought someone with BD would know these things
 I started answering them correctly. Couldn't think of an incorrect answer to give; too
 obvious
 too simple to fake; its too obvious
 too simple; obvious questions; tried to keep it more in general terms; not too specific
 trying to get partial answers (eg color of stop sign = R instead of R & W). Where
 you can get different meanings or interpret differently I went with simpler
 response
 someone may have said R instead of R & W (for color of stop sign); kept it simple
 hard its so everyday; too basic
 I screwed up; I went overboard, they were just too simple not know that
 I just answered them
 no strategy; those were basic questions; they couldn't determine whether person had
 BD or not
 answered honestly
 basic knowledge question; I thought I should answer right
 colour questions tried to answer correctly; same with stuff like walking on your feet.
 common knowledge
 too obvious
 can't fake general knowledge
 I was taken by it; didn't fake it; too obvious
 it was too easy
 didn't change answers; too obvious
 too easy to fake
 tried to be creative; I didn't really think of anything I could say that was result of my
 damage
 at that point I had had enough, did well
 answered those normally, no way you could get one of those wrong.
 too obvious to fake
 too simple to fake
 kept sun and moon same color; nothing obviously wrong

Appendix D

Study 2
Oral Debriefing
Compliant TBI Simulators

Mental Control subtest:

slowed it down
 forgot to perform poorly on the first 2; went slower, purposefully acted like I got
 screwed up around 34 or 35, stalled a bit
 deliberately slowed performance, made self think about it (even though it was
 automatic)
 slowed self down
 basically test shows that numerical background intact; did best with perhaps odd little
 bit in there - nothing major in trying to distort it
 omitted numbers, pauses, slow
 slow performance down, tried to get numbers/letters mixed up or left off
 tried to be slower than usual

California Verbal Learning Test:

tried to do poorly; just by being myself I found it confusing to know which item came
 from what list
 by hardly remembering any in first list; with 2nd list I tried to put those items into
 first list - corresponding to the category
 tried to break up the connections (categories); tried to remember the first and last
 ones, sounds, mixing lists up (A&B), not to tell you all the ones I
 remembered.
 tried to remember just a few key items; forget all the rest; not to repeat all items
 known
 tried to do best because I can't do it
 tried not to listen; slow; omitted
 didn't listen to you too closely; remembered 1/2 categories (fruits & tools)
 tried to do well on that one; can't do it

Visual Reproductions subtest:

mixed it up (the details)
 different to be convincing, reverse a couple of things, subtle differences
 changed details -Card A changed flags, circles detached from top; moved boxes up
 and down; left off inner boxes
 just tried to be a little off; subtle - details were dropped

tried to do best
 changed details
 tried to think of a logical/acceptable variation- reversing the side of the elements,
 details, etc
 changed direction of elements

Trail Making Test

mixed up order
 broke sequence of alternating
 tried not to remember order or sequence, found it very hard, tried to confuse myself
 took longer to do it; tried to show a little confusion on which one came next
 didn't try to lower; may have gone a little slower
 did best
 went slower
 slower

Finger Tapping Test:

go slower
 tried to keep Rhand constant and slow down L hand; wanted a bigger variation in L.
 hand
 went slower, getting my finger to fall off
 tried to get a little shaky after a while; slow it down
 did best
 did best
 went slower
 tried to do best

Digit Span subtest:

tried to slow performance, change sequence
 tried to reverse a few numbers
 drop off numbers, or change the order
 as digits increase I made it like I didn't remember the numbers
 on forward I didn't do anything; on backward tried to throw in a few transpositions of
 numbers
 just left off numbers or stated the wrong number
 tried not to listen closely, changed order of numbers
 changed order

Test of Memory Malingering:

picked out wrong answer

tried to pick wrong answer that looked like other right answer
 hard to screw up, tried to remember grocery list and word association; tried to
 exaggerate
 remembered a few key items, or tried to choose just top or bottom
 random - every now and again choose wrong one
 picked the wrong one
 counting how many I missed etc, choosing wrong one
 tried to do best, my memory is bad

Information subtest:

gave wrong answers
 gave close approximate answer- plausible one, silly
 obvious you'd have to be a dense not to know that the sun rises in the east; changed
 answers
 hard to fake, gage how difficult it might be for someone; gave a few different
 sort of answers
 did well
 just said I don't know
 answered obvious ones, "forgot" some others
 didn't really try (to lower performance)

FAS/Animal Naming Test:

by not saying as many as I knew
 I didn't try to do poorly, I get a mental block when given that one
 stuck to 3 letter words, withheld words, thought of animals around house
 tried to keep it simple; common words; slowed performance down
 taking time; slowing it down; deliberately withheld words
 did it slow
 going slow, leaving off some words
 didn't say some

Abbreviated Hiscock Forced Choice Test:

point to opposite one
 a joke
 found it hard not to say right number, unless I tried to mix numbers up; its a good
 one to catch fakers; worse at end than beginning; there's a way to fake on it - must
 recognize pattern
 I recognized all of them; tried to make it look like I was having a hard time;
 difficult to fake
 didn't try - too easy to do poorly on; you should have made numbers closer
 together/transposed to make it more difficult

picked the wrong one
selected wrong one: tried to get next choice right after a wrong choice
tried my best