

**A Critical Review of the Validity of the Credibility Assessment Tool (CAT) and its
Application to the Screening of Suspected Malingering**

by

Karen Sunette Theunissen

*Thesis presented in fulfilment of the requirements for the degree
Master of Occupational Therapy at the University of Stellenbosch*



Supervisor: Ms René Kemp
Co-supervisor: Ms Blanche Pretorius
Faculty of Health Sciences
Division of Occupational Therapy

March 2011

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

A handwritten signature in blue ink, consisting of several loops and a long horizontal stroke extending to the right.

Signature:

Date: 11/02/2011

Our patients repose in us a sacred trust, and rely upon us not only to guard them from and to alleviate the results of real suffering; but by an unspoken compact, they also look to us to stimulate them to activity when disease has abdicated its throne, but may have left behind morbid disinclination to meet the daily routine of business and the renewed struggle for existence. It is for us to regulate these returning powers; to even forcibly dispel the clouds which retard them, and often delay the recuperative result of a return to the battlefield of life, which is itself the best tonic; and in so doing we are, in one more sense, combating what, if not dispelled, may degenerate into a something which might become first cousin to malingering, that is, fanciful incompetence for duty.

Dr Tennyson Patmore (1894)

ACKNOWLEDGEMENTS

I would like to acknowledge the following people for their contribution to my research:

The members of the LOA who supported the use of their data for this research, specifically Dr Coetzer for his endorsement;

My study leaders, Ms René Kemp and Ms Blanche Pretorius, for their continued support despite its challenges;

Prof Martin Kidd for his assistance with the statistical analysis;

Ms Elana Human for her selfless assistance with the peer review;

Colleagues and friends who supported me with humour and countless cups of tea;

My husband, for his unwavering belief in me and patience throughout the research process.

ABSTRACT

Malingering, the intentional simulation or exaggeration of symptoms for secondary gain, has a significant financial impact on disability insurance given its prevalence. Multidisciplinary professionals involved in disability determination therefore require a tool which would assist in the screening of suspected malingerers.

AIM: The Credibility Assessment Tool (CAT), a tool which was developed as part of the Performance APGAR, was reviewed in terms of its validity and application to the screening of malingering. Research objectives included the review of face and content validity through a literature review and concept analysis, as well as the review of construct and concurrent validity by comparing the results with the operationalised malingering construct and available malingering protocols. The adapted Slick criteria as proposed by Aronoff, applicable to chronic pain, neurocognitive, neurological and psychiatric symptoms, was identified as the most suitable criterion standard for use of comparison.

DESIGN: The research design was a descriptive analytical design, which was performed retrospectively with a report review from insurance referrals to the researcher. Informed consent was obtained from insurers who legally own the reports. A saturated sample of convenience of 184 cases with depression and pain as predominant symptoms were analysed. Recall bias were minimised through omission of personal identifiers and the use of a peer check of 20 random cases. Results in the peer check were suggestive of poor inter-rater reliability, rather than recall bias.

METHOD: Cases were analysed according to the guidelines from the respective authors of the CAT and adapted Slick criteria, however this was further defined to ensure that the study could be replicated.

RESULTS: Face validity was adequate in terms of purpose, item selection and association between consistency criteria, however require improvement in terms of standardised instruction and weighting of the scale. Content validity was rated as adequate to excellent, given that it supports criteria linked to the malingering construct. Construct validity was adequate as demonstrated by association between concepts obtained through concept analysis. Correlation between the CAT and adapted Slick was strong ($r > 0.5$) however caution is expressed that this requires further research.

CONCLUSION: Recommendations for further research included the review of content validity with subject experts, criterion and predictive valid through a case-control study of known-groups, as

well as the reliability of the CAT, and the use of specialised ADL indices for malingering detection. Adaptation to the CAT was depicted in the proposed Consistency Assessment Tool.

Key words: Malingering, credibility, Credibility Assessment Tool, symptom exaggeration, disability insurance, consistency, multimodal assessment

OPSOMMING

Malingering, die opsetlike nabootsing of oordrywing van simptome vir sekondêre gewin, het 'n beduidende finansiële impak op ongeskiktheidsversekering as gevolg van die prevalensie daarvan. Multidissiplinêre professionele persone betrokke by ongeskiktheidsevaluasies het daarom 'n meetinstrument nodig om moontlike malingeerders te identifiseer.

DOEL: Die Credibility Assessment Tool (CAT), wat ontwikkel was as deel van die Performance APGAR, was ondersoek in terme van geldigheid en toepassing op malingering. Navorsingsdoelwitte het die ondersoek van voorkoms- en inhoudsgeldigheid deur 'n literatuurstudie en konsep analise behels, sowel as konstruk- en korrelasie geldigheid deur die vergelyking van die resultate met beskikbare malingering protokolle en operasionele konstrunkte. Die aangepaste Slick kriteria soos voorgestel deur Aronoff, wat toepaslik is op kroniese pyn, neurokognitiewe, neurologiese en psigiatriese simptome, was ge-identifiseer as die meeste gepaste kriterium standaard vir vergelyking.

ONTWERP: Die studieontwerp was 'n beskrywende analitiese studie wat retrospektief uitgevoer was deur 'n ondersoek van verslae van versekeraars. Ingeligte toestemming was verkry van versekeraars wat die wetlike eienaars van die verslae is. 'n Gerieflikheidssteekproef van 184 gevalle met depressie en pyn as hoof simptome was geanaliseer. Sydighede was verminder deur persoonlike inligting te verwyder en die gebruik van 'n eweknie evaluasie van 20 ewekansige getrekte gevalle. Voorlopige resultate dui onbevredigende betroubaarheid aan, eerder as sydighede.

METODE: Gevalle was ge-evalueer volgens die riglyne van die verskeie outeure van die CAT en aangepaste Slick kriteria, en was sodanig verder gedefinieer om te verseker dat die studie herhaal kan word.

RESULTATE: Voorkomsgeldigheid was voldoende, maar verbetering is aanbeveel in terme van gestandaardiseerde instruksie en skaal verdeling. Inhoudsgeldigheid was beduidend in vergelyking met die wetenskaplike literatuur en die geoperasionaliseerde konstrunkte. Konstruktgeldigheid was bevestig deur die positiewe verhoudings tussen die aangepaste Slick en CAT kriteria. 'n Sterk korrelasie was gevind tussen die aangepaste Slick en CAT, maar hierdie moet versigtig ge-interpretier word aangesien verdere navorsing verlang word.

GEVOLGTREKKING: Aanbevelings vir verdere navorsing sluit in die ondersoek van die inhoudsgeldigheid met eksperte, kriterium- en voorspellingsgeldigheid, sowel as die betroubaarheid van die CAT en die gebruik van gespesialiseerde ADL indekse vir uitkenning van malingering. Aanpassing vir die CAT word ook voorgestel.

LIST OF OPERATIONAL DEFINITIONS

Construct

The term construct refers to the concept, attribute, or variable that is the target of measurement and is usually not directly observable ⁽¹⁰⁵⁾. For the purpose of this study, the malingering construct will be reviewed in terms of underlying concepts.

Concept analysis

Concept analysis is used to clarify phenomenon, or to examine the characteristics of a concept, for the derivation of operational definitions. The concept analysis is often graphically presented to define boundaries and interrelationships ⁽¹⁰²⁾.

Criterion standard

A criterion standard is a measure accepted by consensus of content experts as the best available for determining the presence or absence of a particular phenomenon. When there is no perfect criterion standard, then pragmatic criteria can be used as a criterion standard ⁽⁹¹⁾.

Depression

Major depression includes a constellation of symptoms such as depressed mood, diminished interest or pleasure, change in neurovegetative functioning, feelings of worthlessness, cognitive difficulties and suicidal thoughts ⁽²⁾. For the purpose of this study, depression refers to the predominant claim cause and may be linked to another psychiatric diagnosis. Given that there is often more than one diagnosis in practice, it does not only refer to Major Depression, but also to Bipolar Mood Disorder, Post Traumatic Stress Disorder and Generalised Anxiety Disorder.

Disability

Disability is a complex phenomenon as the definition thereof depends on the context. Disability is often used synonymously with impairment, which refers to the alteration of functional capacity whether this is physical, cognitive or emotional. For the purpose of this study, disability is defined as the impact of the impairment on personal, social or occupational demands ^{(14) (16) (48)}. The definition is the same definition used in the insurance industry, which therefore contextualises it.

Functional Capacity Evaluation

Functional Capacity Evaluations are comprehensive and performance-based evaluations conducted by rehabilitation practitioners to determine the safe functional ability of a person with a work-related impairment ^{(9) (23) (24)}.

Malingering

The intentional production of false or exaggerated symptoms motivated by external incentives, such as obtaining compensation or drugs, avoiding work or military duty, or evading criminal prosecution ⁽¹⁾.

Secondary Gain

Secondary gain refers both to external factors, such as financial awards, or internally motivated factors, such as the adoption of the sick role, which provide advantages attained by the patient as a consequence of illness ⁽⁵⁶⁾.

Pain

The aetiology and dynamics of pain is considered complex even though it is common. For the purpose of this study, pain is defined as chronic pain which has not responded to usual treatment or within usual treatment duration parameters. Pain could be a result of surgery, injury or illness, although the predominant cause in this study is linked to musculoskeletal spinal pain.

Validity

The validity of an instrument is a determination of the extent to which the instrument actually reflects the abstract construct being examined⁽¹⁰²⁾. In this study, the validity therefore refers to the extent to which the CAT reflects the construct of malingering.

Validity was traditionally categorised into three or four specific types: face and content validity, criterion-related validity (which included concurrent and predictive validities), and construct validity. However this has been considered problematic given that types are often interrelated and therefore not mutually exclusive. The latest APA standards⁽¹⁰³⁾ have therefore indicated that validity is a unitary concept that considers the appropriateness, meaningfulness and usefulness of the specific inferences made from instrument scores. It therefore considers the degree to which both evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Of specific note is that construct validity is considered the key and unifying type of validity.

It is therefore recommended that validity should be reviewed in terms of the sources of evidence rather than distinct types of validity.

- Evidence based on test content:

This type of validity evidence is based on logical analyses and experts' evaluations of the content of the measure, including items, tasks, formats, wording and processes. It addresses questions about the extent to which the content of a measure represents a specific content domain or construct.

- Evidence based on internal structure:

This is considered part of construct-related evidence and examines the extent to which the internal components of a test match the defined construct.

- Evidence based on relations to other variables:

This encompasses many of the old specific types of validity such as criterion and construct validity (including convergent and discriminant validity). The most common approaches to the collection of this type of evidence are correlational, criterion-group or known-group and experimental studies.

Having said that, there is ongoing difficulty employing this in practice as articles and test manuals still present the former method. For this reason, for the purpose of the study, reference will be made to the following:

Face validity

The verification that the CAT looks like it is valid, or gives the appearance that it is measuring the concepts of malingering ⁽¹⁰²⁾. This refers to evidence based on test content.

Content validity

Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose. In this instance, it refers to the degree to which elements of the CAT are relevant to and representative of the targeted construct of malingering for the screening thereof. This refers to evidence based on test content.

Construct validity

Construct validity refers to the degree to which inferences can legitimately be made from the operationalisations in the study to the theoretical constructs on which those operationalisations were based. This refers to evidence based on internal structures and the relations to other variables.

Concurrent validity

Concurrent validity is a measure of how well a particular test correlates with a previously validated measure when administered at the same time. In this study, concurrent validity will be measured by analysing how well the CAT correlates to other available malingering protocols. It refers to evidence based on relations to other variables.

LIST OF ABBREVIATIONS

ADL	Activities of Daily Living
AMA	American Medical Association
APA	American Psychological Association
CAT	Credibility Assessment Tool
COID	Compensation of Occupational Injury and Diseases (formerly known as Workman's Compensation Act)
DSM	Diagnostic and Statistical Manual of Mental Disorders
FAADEP	American Academy of Disability Evaluation Physicians
FCE	Functional Capacity Evaluation
IME	Independent Medical Examiners
LOA	Life Offices Association (now known as ASISA)
MND	Malingered Neurocognitive Dysfunction
MPRD	Malingered Pain Related Dysfunction
SSA	Social Security Administration

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2.1	Potential categories of response style based on a cross-classification of “intent” and “effort.” Source: Frederick, 2003	18
2.2	Conceptual and assessment overlay between exaggeration, poor effort and malingering Source: Iverson, 2006	19
2.3	Operationalised malingering construct	29
2.4	Framework for Analysing Functional Performance Source: Genovese and Galper, 2009	33
3.1	Correlation with Peer Check	49
3.2	Method Flowchart	50
4.1	Gender distribution	70
4.2	Age distribution	71
4.3	Overall diagnosis distribution	72
4.4	Prevalence of Malingering	73
4.5	Base Rates of Malingering	74
4.6	CAT overall score distribution	77
4.7	Comparison between CAT score and Slick classification	78
4.8	Association between Response Bias and CAT criterion E	80
4.9	Association between Test Data and Known Patterns with CAT criterion E	81
4.10	Association between Test Data and Observed Behaviour with CAT criterion E	82
4.11	Association between No test data and CAT criterion E	83
4.12	Correspondence Analysis of Test Data on CAT criterion E	84
4.13a	Association between Exaggeration and overall CAT classification	86

4.13b	Association between Exaggeration and overall Slick classification	86
4.14	Association between Self Report and Known Patterns with CAT classification	87
4.15	Association between Self Report and Observed Behaviour with CAT classification	88
4.16	Association between Self Report and Collateral Information with CAT classification	89
4.17	Association between Self Report and Documented History with CAT classification	90
4.18	Correspondence Analysis of Self Report	91
4.19	Association between ADL and Slick classification	92
4.20	Association between Medication and Slick classification	93
4.21	Association between Treatment and Slick classification	94
4.22	Association between Medical Collateral and Slick classification	95
4.23a	Association between Injury and Slick classification	97
4.23b	Association between Injury and CAT classification	97

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1.1	Five year overview of fraudulent and dishonest claims statistics Source: Asisa, 2009	3
2.1	Secondary gain factors Source: Dersh, Polatin, Leeman and Gatchel, 2004	16
2.2	Factors suggesting the Presence of Malingering Source: Samuel and Mittenberg, 2005	31
2.3	Assessment Guide for Amplification / Malingering in Head Injury Source: Sreenivasan, Eth, Kirkish and Garrick, 2003	32
2.4	Criteria for Malingered Neurocognitive Dysfunction (MND) Source: Slick, Sherman and Iverson, 1999	34
2.5	Criteria for Malingered Pain Related Dysfunction (MPRD) Source: Bianchini, Greve and Glynn, 2005	35
2.6	Criteria for Malingered Pain, Neurocognitive, Neurological, and Psychiatric Symptoms Source: Aronoff, Mandel, Genovese, Maitz, Dorto, Klimek and Staats, 2000	37
2.7	Comparative Analysis of Models	38
2.8	Performance APGAR Model Source: Colledge, Holmes, Randolph Soo Hoo, Johns, Kuhnlein, DeBerard, 2001.	39
2.9	Credibility Assessment Tool Source: Colledge et al., 2001.	40
4.1	Comparison between Content of Slick and CAT	66
4.2	Correlation of CAT and Slick classification per diagnosis	76
4.3	Descriptive statistics of CAT score and Slick classification	79
4.4	Association between Financial Incentive with Slick and CAT classification	96
5.1	Proposed Consistency Assessment Tool	111

LIST OF APPENDICES

<u>Appendix</u>		<u>Page</u>
Appendix A	Data capture sheet	131
Appendix B	Scoring criteria	133
Appendix C	Informed consent	136
Appendix D	Ethical approval	140
Appendix E	Originality report	143

TABLE OF CONTENTS

		Page
Declaration		ii
Acknowledgements		lv
Abstract		v
Opsomming		vii
List of operational definitions		viii
List of abbreviations		xii
List of figures		xiii
List of tables		xv
List of appendices		xvi
CHAPTER 1	INTRODUCTION AND STATEMENT OF PROBLEM	1
1.1	Introduction	2
1.2	Background and Significance	2
1.3	Financial Compensation for Disability	3
1.4	The Disability Determination Process	5
1.5	The Role of the Occupational Therapist in the South African Insurance Industry	6
1.6	Statement of the Problem	7
1.7	Research Objectives	8
1.8	Summary	9
CHAPTER 2	LITERATURE REVIEW	10
2.1	Introduction	11
2.2	The Construct of Malingering	11
2.2.1	Intent	12
2.2.2	Simulation or exaggeration	13
2.2.3	Diagnosis	13
2.2.3.1	Malingering and Pain	13
2.2.3.2	Malingering and Depression	14

2.2.3.3	Malingering and Fibromyalgia	14
2.2.4	External incentive or gain	15
2.3	The Clinical Assessment of Malingering	17
2.3.1	Effort, Intent and Motivation	17
2.3.2	Use of Reported Information for Malingering Detection	19
2.3.2.1	Self-reported Measures	20
2.3.2.2	Collateral Sources	21
2.3.2.3	Records	21
2.3.2.4	The Concept of Credibility	22
2.3.3	Use of Tests for Malingering Detection	23
2.3.3.1	Behavioural Observations	23
2.3.3.2	Testing Specific to Diagnosis	24
2.3.3.3	Specialised Testing	24
2.3.4	Underlying Concepts of Malingering Detection	26
2.3.5	Conceptual Challenges of Malingering Detection	27
2.4	Synthesis of Findings from Literature Review	28
2.5	Methodological Challenges in Malingering Research	29
2.6	Models Used in Malingering Detection	31
2.7	Comparative Analysis of Models	38
2.8	The Development of the Credibility Assessment Tool (CAT)	39
2.9	Summary	41
CHAPTER 3	METHODOLOGY	42
3.1	Introduction	43
3.2	Method of Inquiry with Literature Review	44
3.2.1	Face Validity	44
3.2.2	Content Validity	44
3.3	Study Design to Measure Construct and Concurrent Validity	45
3.4	Population and Sample	45
3.4.1	Inclusion Criteria	46
3.4.2	Exclusion Criteria	46
3.5	Sampling Method and Biases	47
3.6	Method	49
3.7	Measurement Instrument	51
3.8	Interpretation of Measurement Criteria	51

3.8.1	Slick criteria	52
3.8.1.1	Slick criterion A (Incentive)	52
3.8.1.2	Slick criterion B (Testing)	53
3.8.1.3	Slick criterion C (Self Reports)	54
3.8.1.4	Slick criterion D (Alternative Factors)	55
3.8.2	CAT criteria	55
3.8.2.1	Effects of symptoms or impairment on performance of ADL	55
3.8.2.2	Type, dosage, effectiveness and side effects of medication	56
3.8.2.3	Treatment sought and received	56
3.8.2.4	Opinions that have been recorded by professionals who have treated and/or examined the patient	56
3.8.2.5	Conflicts in the allegations, statements or medical evidence in the file	57
3.9	Data Analysis	57
3.10	Ethics	58
3.11	Summary	59
CHAPTER 4	DATA ANALYSIS AND RESULTS	61
4.1	Introduction	62
4.2	Results of Face Validity of CAT	62
4.2.1	Purpose	62
4.2.2	Instruction	63
4.2.3	Item selection	63
4.2.4	Scaling and weighting	65
4.2.5	Level of measurement	66
4.3	Results of Content Validity Measurement of the CAT	66
4.3.1	Concept of Credibility	67
4.3.2	Discrepancy Methods	67
4.3.3	Effort, Motivation, Compliance and Incentive	68
4.3.4	Expert Review of Content Validity	68
4.4	Statistical Analysis of Construct and Concurrent Validity	68
4.5	Demographics	70
4.5.1	Gender	70
4.5.2	Age	71
4.5.3	Diagnosis	72

4.6	Prevalence of Malingering	73
4.7	Comparison between CAT and Slick Classification of Malingering	75
4.7.1	Correlation per diagnosis	76
4.7.2	Analysis of CAT scoring	77
4.8	Analysis of Test Criteria on Overall Outcome	79
4.8.1	Impact of Testing (Slick criterion B) on CAT	79
4.8.1.1	Probable Response Bias (RB)	80
4.8.1.2	Test Data and Known Patterns of Functioning (TD ≠ KP)	81
4.8.1.3	Test Data and Observed Behaviour (TD ≠ OBS)	82
4.8.1.4	Test Data and Reliable Collateral Reports (TD ≠ COLL)	82
4.8.1.5	Test Data and Documented Background History (TD ≠ HIST)	83
4.8.1.6	No Significant Test Data	83
4.8.1.7	Overall Association between Test Data on CAT criterion E (Conflicts in the allegations, statements or medical evidence in the file)	84
4.8.2	Association between Self-report (Slick criterion C) and CAT	85
4.8.2.1	Self-report and Exaggeration or Fabricated Psychological Dysfunction (EXAG)	85
4.8.2.2	Self-report and Known Patterns of Functioning (SR ≠ KP)	87
4.8.2.3	Self-report and Observed Behaviour (SR ≠ OBS)	88
4.8.2.4	Self-report and Collateral Information (SR ≠ COLL)	89
4.8.2.5	Self-report and Documented History (SR ≠ HIST)	90
4.8.2.6	Overall Association of Self-report on CAT classification	91
4.8.3	Association between CAT criteria and Slick classification	92
4.8.3.1	Association between ADL and Slick classification	92
4.8.3.2	Association between Medication and Slick classification	93
4.8.3.3	Association between Treatment and Slick classification	94
4.8.3.4	Association between Medical Collateral and Slick classification	95
4.9	Alternative Factors	96
4.9.1	Financial Incentive	96
4.9.2	Injury	97
4.10	Summary	98
CHAPTER 5	CONCLUSIONS AND RECOMMENDATION	99
5.1	Introduction	100
5.2	Face Validity of the CAT	100
5.3	Content Validity of the CAT	102

5.4	Construct and Concurrent Validity of the CAT	104
5.5	Limitations of the Study	106
5.6	Recommendations	107
5.6.1	Recommendations for Further Research	107
5.6.2	Practical Applications for Further Use	110
5.7	Conclusion	112
REFERENCE LIST		114

CHAPTER 1

INTRODUCTION AND STATEMENT OF PROBLEM

1. INTRODUCTION AND STATEMENT OF PROBLEM

1.1 Introduction

The assessment of malingering is an important aspect within disability determination for insurance purposes. The process, tools and challenges of disability determination will therefore be described. This will serve to highlight the problem and delineate the research objectives of this research study.

1.2 Background and Significance

Malingering is most frequently defined as the “intentional simulation or exaggeration of psychological or physical symptoms for secondary gain” ⁽¹⁾. The DSM IV-R ⁽²⁾ similarly defined malingering as “the intentional production of false or exaggerated symptoms motivated by external incentives, such as obtaining compensation or drugs, avoiding work or military duty, or evading criminal prosecution”.

Malingering is a widely publicised and debated subject with regards to financial compensation for illness or injury, especially in the insurance industry. The significance thereof is best illustrated in terms of the financial implication as indicated in table 1.1 ⁽³⁾. This refers only to the number of fraudulent and dishonest claims detected. The implication is that insurance companies would have to recover these losses from clients, which makes insurance less affordable and provide less opportunity for payment of discretionary rehabilitation benefits.

Table 1.1 Five year overview of fraudulent and dishonest claims statistics**Source: Asisa, 2009**

	2004	2005	2006	2007	2008
Fraudulent documents	R33 million	R31.9 million	R21.1 million	R75.8 million	R60.1 million
Beneficiary and syndicate fraud	R6.6 million	R17.2 million	R4.2 million	R million	R12 million
Material non-disclosure	R127.2 million	R157.4 million	R143.5 million	R127 million	R244.6 million
Misrepresentation	R32 million	R138.5 million	R69.3 million	R60 million	R49 million
Intermediary involvement	R15.6 million	R2.1 million	R4.2 million	R6.1 million	R10.2 million
TOTAL	R214.4 million	R347.1 million	R242.3 million	R278.9 million	R375.9 million

Available prevalence rates linked to disability assessments are high, with international literature indicating rates up to 40% ⁽⁴⁾ ⁽⁵⁾. This is especially true for diagnoses linked to pain and depression ⁽⁶⁾. In the South African insurance industry, claim payouts for pain and depression-related causes are up to 50% ⁽⁷⁾. For these reasons, medical professionals involved with disability determination cannot ignore the impact of malingering when rendering opinions. This is also relevant in the South African context where referrals are often made to occupational therapists by the insurance industry to determine the functional capacity of suspected malingerers ⁽⁸⁾.

1.3 Financial Compensation for Disability

Internationally there are various systems to compensate for disability, which differ in the definition of disability according to the policy and assessment methodology. It can usually be divided into public or national insurance, workers' compensation, third party insurance and private disability insurance policies ⁽⁹⁾.

These systems are generically applicable as follows:

- Public or national insurance are usually relevant to all persons with disabilities, irrespective of age or employment status ^{(9) (10)}.
- Workers' compensation is usually linked to any injury or illness that arises out of the course of employment and includes cover for treatment, as well as temporary and/or permanent disability awards ^{(9) (11) (12)}.
- Third party liability usually refers to cover by a third party in an accident, which in South Africa, is usually linked to the Road Accident Fund. Compensation is awarded for loss of earnings and general damages for pain and suffering ⁽¹³⁾.
- Private disability insurance can be purchased by individuals or employers to cover events linked to illness and injury. It is usually written in terms of specific illness definitions, level of functional impairment or occupational disability ⁽⁹⁾. Within the field of private disability insurance in South Africa, cover is usually in terms of the type of work (i.e. ability to perform own, similar or any occupation) or according to the duration of disability (i.e. temporary or permanent) or degree of disability (i.e. total or partial) ^{(12) (14) (15) (16)}.

The context of this study is on private disability insurance. It is worth noting that the private insurance category does not refer to socio-economic status, as this category also include employee benefits provided by employers and therefore include unionised business, parastatal organisations and municipalities.

1.4 The Disability Determination Process

The awarding of monetary payouts for disability varies widely based on the contractual underpinning of the claim, assessment methodology and rating scales, as well as the definition of disability. Comment is often required regarding impairment, which refers to the alteration of functional capacity on medical grounds, as opposed to disability, which refers to the impact of the impairment on personal, social or occupational demands, and is assessed by non-medical means^{(14) (15) (16)}. Disability determination is therefore not a medical decision, but a collective decision made by a panel of experts including medical advisors, legal advisors and claims consultants. It is therefore clear that specialised skills and techniques over and above medical speciality are required⁽¹⁷⁾.

The assessment of impairment is primarily the role of Independent Medical Examiners (IME)^{(14) (16)}. Internationally, it is advised that IME have additional training and/or certification in disability medicine⁽¹⁷⁾. Functional impairment is only assessed once the patient has received reasonable optimal treatment, the condition is medically stabilised and maximal medical improvement has been reached⁽¹⁶⁾. The assessment can be performed with a record review of available information or an actual clinical assessment of the reported impairment^{(18) (19)}. If indicated, referral is made to independent specialists for further opinion. Clinical reasoning within disability determination includes review of the medical evidence, treatment and rehabilitation against time perspective, as well as efforts to recover and resume work⁽²⁰⁾. Comment is often required about causation, financial gain and motivation^{(18) (21)}.

Within the South African insurance industry, the role of IME are often supplemented or fulfilled by the claims consultant in conjunction with information that includes:

- Treating medical practitioner report
- Independent specialist opinion
- Functional capacity evaluation
- Collateral information from the employer
- Information from the rehabilitation team regarding progress and compliance⁽¹⁴⁾
- Video surveillance by private investigators⁽¹⁸⁾.

1.5 The Role of the Occupational Therapist in the South African Insurance Industry

The role of the occupational therapist in the South African insurance industry is two-fold: Firstly, they play an integral part in disability management by being employed as claims consultants or case managers ⁽²²⁾. Secondly, they also provide independent opinion in terms of functional impairment by conducting Functional Capacity Evaluations (FCE).

FCE are used internationally to assess the safe functional ability of a person with a work-related impairment ⁽²³⁾. It is a comprehensive and performance-based evaluation conducted by rehabilitation disciplines ^{(9) (24)}, although predominantly remains in the scope of occupational therapists in South Africa. It is interchangeably called Functional Assessments, Functional Assessment Evaluations, Vocational Assessment, Physical Capacity Evaluations and Work Capacity Evaluations ^{(25) (26)}. Depending on whether work visits are performed, it is also referred to as “Workplace Assessments” ^{(27) (28)}. For the purpose of this research study, reference will be made to FCE throughout.

Insurers rely on the results of FCE to ascertain entitlement to disability benefits ^{(15) (29)}. FCE are also often used to determine liability of rehabilitation benefits, as conclusions about an individual’s ability to return to work, and the recommended rehabilitation or work modification to achieve this can be obtained from an FCE. At times, FCE are used to determine the individual’s effort and consistency of performance with suspected malingerers or as a final comment for adjudication purposes ^{(26) (30)}. FCE methodology based on scientific research therefore minimises the financial risk to the insurer, improves opinion in litigious cases and maintains professional credibility.

Specifically to insurers there are definite financial risks related to the payment of claims. Valid and reliable methods to enable accurate disability determination are therefore required, more so given the litigious background of claims assessment.

1.6 Statement of the Problem

Different systems for measuring impairment and disability exist, dependent on the type of cover and the specific policy conditions. This highlights important challenges within disability determination:

- Impairment criteria are not always consistent ⁽⁹⁾.
- More than one financial award can be obtained for a single health event ⁽¹²⁾.
- There are poor inter-rater reliability between IME ⁽¹⁹⁾.

There is a lack of standardised evaluation protocols across all specialities related to disability determination, even though attempts have been made to provide guidelines ^{(12) (14) (16) (19)}. This is especially relevant with suspected malingering, where terminologies are often inconsistent, and multidisciplinary test instruments and protocols have not been extensively reviewed in terms of scientific principles such as validity and reliability ⁽⁹⁾. There have not been any specific guidelines regarding malingering in South Africa. Although a number of methods have been proposed to ascertain the sincerity of effort, there has been limited focus on developing a tool which can be used for the purposes of multidisciplinary assessment.

The Guide to the Evaluation of Permanent Impairment ⁽³¹⁾ (often referred to as the AMA Guidelines) has been used internationally by IME and claims consultants to ascribe a numeric rating for impairment. It is also used in South Africa for disability determination, and certain insurance policies are written specifically bearing it in mind. Nevertheless, it does not provide a determinant of work disability or rate sincerity of effort as part of the assessment protocol ⁽⁹⁾.

Within the USA's Social Security Administration (SSA), they have developed assessment criteria which include factors such as age, education and vocational function ⁽⁹⁾. It therefore does not reflect on an award of percentage impairment only, which therefore provides greater applicability. Proposed tools used in conjunction within the SSA include the Performance APGAR and Credibility Assessment Tool (CAT) which allow IME to uniformly measure sincerity of effort ⁽³²⁾.

Even though there has been limited published review of the CAT, it is therefore considered relevant to the insurance industry in terms of the tool's:

- Ability to be used by various specialities involved with disability determination.
- Original development for use by insurers.
- Flexibility in applying it to other test instruments and protocols.
- Utility both in terms of a stand-alone tool and in combination with other tools.

The CAT is therefore a tool that requires further research regarding its application to malingering.

1.7 Research Objectives

The aim of the research project is to perform a critical review of the validity of the CAT and its application to the screening of suspected malingering.

The objectives therefore include the review of the validity of the CAT, and its application to other existing malingering tests and protocols. This will be achieved by:

- Performing an extensive literature review of the malingering construct to ascertain face and content validity by performing an analysis of the underlying concepts and scale construction;
- Measuring construct validity by determining the relationships between the variables of the CAT and concepts obtained by the concept analysis and emerging factors during the literature review.
- Measuring concurrent validity by comparing the results of the CAT with the best available criterion standard as obtained from malingering protocols obtained in the literature review.

1.8 Summary

Malingering, the intentional simulation or exaggeration of symptoms for secondary gain, has a significant financial impact on disability insurance given the relatively high prevalence thereof. Multidisciplinary professionals involved in disability determination are presented with a number of challenges when rendering opinion regarding functional impairment, causation and efforts to recover and resume work. Specifically, the underlying motivation and sincerity of effort when interpreting comprehensive assessments, such as FCE, are not always clear. Of note is that there have not been any specific guidelines for malingering detection in the South African insurance industry.

There is a lack of standardised evaluation protocols across all specialities involved with disability determination. Professionals involved with disability determination would therefore benefit from a well-researched tool which would assist in the screening of suspected malingerers. Unfortunately limited research has been done in this regard.

The CAT, a tool which was initially developed as part of the Performance APGAR for the purpose of disability determination, will therefore be critically reviewed in terms of its validity and application to malingering.

CHAPTER 2

LITERATURE REVIEW

2. LITERATURE REVIEW

2.1 Introduction

In order to critically review the Credibility Assessment Tool (CAT) for the purpose of malingering screening, it is vital that a clear understanding of malingering is obtained from literature. This includes the reviewing of theoretical concepts linked to malingering detections in terms of test methodology and models in order to operationalise the construct of malingering. The conceptual and methodological challenges as it relates to the validity of malingering detection tools and approaches will also be investigated.

The search strategy included both online and manual searches for appropriate literature in databases predominantly linked to disability medicine, psychiatry, neuropsychology and occupational rehabilitation. Key terms linked to malingering were used, such as malingering, symptom exaggeration, symptom magnification, faking, dissimulation, effort, credibility, consistency assessment.

2.2 The Construct of Malingering

The Diagnostic and Statistical Manual of Mental Disorders (DSM) ⁽²⁾ defines malingering as “the intentional production of false or exaggerated symptoms motivated by external incentives, such as obtaining compensation or drugs, avoiding work or military duty, or evading criminal prosecution”. Simply put, the definition often referred to in research articles is that malingering is the “intentional simulation or exaggeration of psychological or physical symptoms for secondary gain” ⁽¹⁾.

Concept analysis indicates four separate concepts within the malingering construct as:

- Intent
- Simulation or exaggeration
- Diagnosis (psychological or physical symptoms)
- External incentive or gain.

2.2.1 Intent

The first concept of intent refers to the judgement of the intention of the examinee, which is diagnostic of malingering, somatoform and factitious disorders in terms of the discrimination between conscious and unconscious behaviour. Intent and motive are vital components of malingering when determining secondary gain. It is used to differentiate malingering and factitious disorders that have conscious intent, from somatisation disorders which are motivated by unconscious or involuntary intent ⁽³³⁾.

Within disability determination, even though tests can focus on certain clinical aspects, it cannot ascertain the motive or intention behind an individual's test presentation. Contrary to the diagnostic requirement thereof, it has also been argued that clinicians have no special expertise in the assessment of veracity and that there is poor empirical basis for such judgements ^{(34) (35) (36)}. It is therefore often argued that the judgement of malingering is a legal determination given the allegation of fraud ^{(34) (35)}. A further argument is that, if consistencies have been reported, it should not be interpreted ⁽³⁷⁾ as often further research is required to comment on the significance of it ⁽³⁴⁾. This, coupled with ethical reasons for misclassifying a malingerer ⁽²⁵⁾, has led to clinicians using terminology other than malingering during reporting ⁽³⁸⁾, such as referring to invalid or inconsistent test results.

These factors have complicated the operationalisation and use of the malingering definition. Further research has therefore largely focused on the other concepts to improve the scientific basis of malingering determination.

2.2.2 Simulation or exaggeration

A common misconception exists that malingering only refers to deliberate fabrication of symptoms, whereas in fact, dual criteria exists in terms of either simulation (therefore false representation) or exaggeration. This concept has further been classified in terms of positive malingering, which refers to the feigning of symptoms that do not exist, and partial malingering, which refers to the conscious exaggeration of symptoms that do exist ⁽³⁴⁾. In practice, terminologies to describe the exaggeration have included “functional overlay”, “symptom magnification syndrome”, “submaximal effort” and “abnormal illness behaviour” ⁽³⁴⁾, which complicates the delineation of concepts. Caution has been expressed that symptom exaggeration does not necessarily constitute malingering as factors such as personality, genuine brain damage and over-familiarity during the protracted medico-legal process could contribute to seemingly exaggeration ^{(39) (40)}.

2.2.3 Diagnosis

Symptoms, and the severity thereof, should be linked to the injury or illness for which a claim has been submitted. The literature predominantly focuses on neurocognitive sequelae of injuries and illnesses given that malingering research has mostly been in the domain of neuropsychology, however the body of evidence is growing in terms of psychiatric conditions and pain. Specific to the insurance industry, the conditions commonly associated with malingering is related to pain and/or depression, which is especially relevant given that these symptoms remain the largest cause of claim payouts ⁽⁷⁾.

2.2.3.1 Malingering and Pain

Pain management is often affected by factors, such as financial incentives and medication-seeking behaviour ⁽⁴¹⁾. This therefore raises suspicion of potential malingering given that these factors directly tie in with the definition of malingering which includes “motivated by external incentives, such as obtaining compensation or drugs” ⁽²⁾.

Malingered cognitive impairment has been documented in patients whose primary complaint is pain ⁽⁴²⁾ ⁽⁴³⁾. The presence of financial incentive has also been known to influence symptom report and test performance with workers' compensation pain patients who, comparatively, reported more cognitive symptoms than non-litigating patients with head injury ⁽⁴⁴⁾. Chronic pain patients involved in disability litigation have been found to fail cognitive symptom validity indicators at higher rates than non-litigating pain patients and non-litigating traumatic brain injury patients, indicating symptoms exaggeration of some litigating pain patients ⁽⁴⁵⁾ ⁽⁴⁶⁾.

It is reasonable to assume that a substantial ratio of the symptom exaggeration is intentional given that the base rate of malingering in pain has been found to be between 20% and 40% ⁽⁶⁾ ⁽⁴⁶⁾ ⁽³⁴⁾. Evidence of malingering in the form of covert video surveillance was found in 20% of pain patients pursuing compensation ⁽⁴⁷⁾.

2.2.3.2 Malingering and Depression

Psychiatric diagnoses feature as one of the most common causes of disability claims in South Africa. Noteworthy is that it usually includes common conditions, such as depression, anxiety and post-traumatic stress disorder, rather than major psychotic illnesses ⁽¹²⁾ ⁽⁴⁸⁾.

The most common potential areas for deception on psychiatric grounds are that of psychotic symptoms and cognitive impairment. Even though research has been extensive, the assessment of malingering remains complex given that it is difficult to operationalise poor effort and motivation during testing, as this remains part of the symptoms of depression ⁽¹²⁾ ⁽⁴⁹⁾. Nevertheless, approximately 25 to 30% of patients claiming disability due to major depressive disorder may perform in the range that suggests probable malingering upon testing ⁽⁶⁾.

2.2.3.3 Malingering and Fibromyalgia

A separate note is made on fibromyalgia, given that it remains a contentious issue within the life insurance industry. Even though it can be considered a pain syndrome, many still argue that it is somatised depression. Fibromyalgia is often classified a functional somatic syndrome, as it is diagnosed with different symptom syndromes dependent on the medical specialist ⁽⁵⁰⁾. Given that

the nature and severity of the symptom is based on self-report and therefore often viewed as subjective, it remains a controversial syndrome despite a growing body of evidence challenging these perceptions ^{(51) (52)}.

Within the arena of disability insurance, it has been perceived as increasing in prevalence as a cause for disability ^{(53) (54)}. Juxtaposed with the need for disability determination, it was also found that the repeated attention to pain symptoms during disability assessments also amplifies the condition ⁽⁵⁵⁾.

The cost of disability linked to fibromyalgia is significant, especially measured against the fact that the syndrome is easily simulated, where approximately 25 to 30% claimants may perform in the range that suggests probable malingering on forced choice tests ⁽⁶⁾.

2.2.4 External incentive or gain

Secondary gain is often erroneously used as synonymous with financial compensation associated with disability ⁽⁵⁶⁾. The original construct was however coined by Freud when describing “interpersonal or social advantage attained by the patient as a consequence of...illness” ⁽⁵⁷⁾. The term is descriptive of both external factors, such as financial awards, as well as internally or psychologically motivated factors which are affected by conscious and unconscious motivation, personality, relationship dynamics and reinforcers ⁽⁵⁶⁾.

To illustrate, whereas the external factors refer to financial incentive, the internal factors include the adoption of the “sick role”. A comprehensive listing of internal and external secondary gains has been made by Dersh, Polatin, Leeman and Gatchel ⁽⁵⁶⁾ which has been summarised in table 2.1.

Table 2.1 Secondary gain factors**Source: Dersh, Polatin, Leeman and Gatchel, 2004.**

Internal	External
Gratification of pre-existing unresolved dependency striving or revengeful strivings	Obtaining financial awards associated with disability – Wage replacement – Settlement – Disability-based debt protection – Subsidised child and family care, housing and food.
An attempt to elicit care-giving, sympathy, and concern from family and friends.	
Family anger because of patients' disability may increase patient resentment and determination to get his/her due to prove entitlement	
Obtaining one's entitlement for years of struggling, dutiful attention to responsibilities, and a "much-earned" recompense	
Ability to withdraw from unpleasant or unsatisfactory life roles, activities, and responsibilities	
Adoption of "sick role" allow the patient to communicate and relate to others in a new, socially sanctioned manner	Protection from legal and others obligations (child support payments, court appearances, parole or probation demands)
Converting a socially unacceptable disability (psychological disorder) to a socially acceptable disability (injury or disease)	Job manipulation (promotion or transfers, handling work adjustment difficulties, prevention of termination)
Displacing the blame for one's failures form oneself to an apparently disabling illness beyond one's control	Vocational retraining and skills upgrade.
Maintenance of status in family, holding a spouse/partner in a marriage/relationship, avoiding sex, contraception	
Obtaining drugs.	

Even though secondary gain, by definition, is necessary for malingering to occur, it can also be present without malingering. Secondary gain by itself is rarely suggestive of pure malingering (i.e. feigning of disability when it does not exist) ⁽³⁴⁾. The concept of secondary gain is further complicated by whether psychodynamic processes are conscious or unconscious, as this would have a direct impact on discriminating between factitious and somatisation disorders as indicated in section 2.2.1.

Given these factors, recent malingering research has defined the gain concept as external incentive ^{(58) (41)} which therefore delineate it from underlying psychological factors.

2.3 The Clinical Assessment of Malingering

Even though a comprehensive review of specialised testing falls beyond the scope of this research study, the underlying concepts of assessments provide valuable information regarding the malingering construct as delineated in section 2.2. As such, assessment would consider aspects of effort, intent and motivation, reported symptoms, testing specific to diagnosis as well as specialised testing.

2.3.1 Effort, Intent and Motivation

Performance during assessment is often interpreted as motivated as opposed to malingered ⁽³³⁾, however this simplistic approach does not conceptualise the malingering construct or allow for the continuum of test behaviour.

Motivation includes both the effort expended, as well as the underlying intent for certain behaviours. The motivation behind malingering is crucial when determining secondary gain. Rogers ⁽⁵⁹⁾ classified this in three types:

- Pathogenic motivation includes symptom exaggeration in psychiatric illnesses in order to avoid managing one's own life.
- Criminological motivation is in keeping with the DSM. It was best described by Rogers ⁽⁵⁹⁾ as "a bad person" (with antisocial personality disorder) "in bad circumstances" (legal difficulty) is "performing badly" (uncooperative).
- Adaptational motivation is linked to the cost-benefit analysis involved when confronted with an adversarial situation when personal stakes are high.

Within the arena of private disability insurance, the adaptational model is usually most suitable when considering the motivation behind malingering.

Even though it has been recommended that intent is a legal determination as discussed in section 2.2.1, in practice it is often difficult to delineate intent from concepts such as motivation and effort as it is often interrelated. Inter-relationships have been described with regards to effort and intent. This has been modelled to define compliance by Frederick⁽⁶⁰⁾ as indicated in figure 2.1.

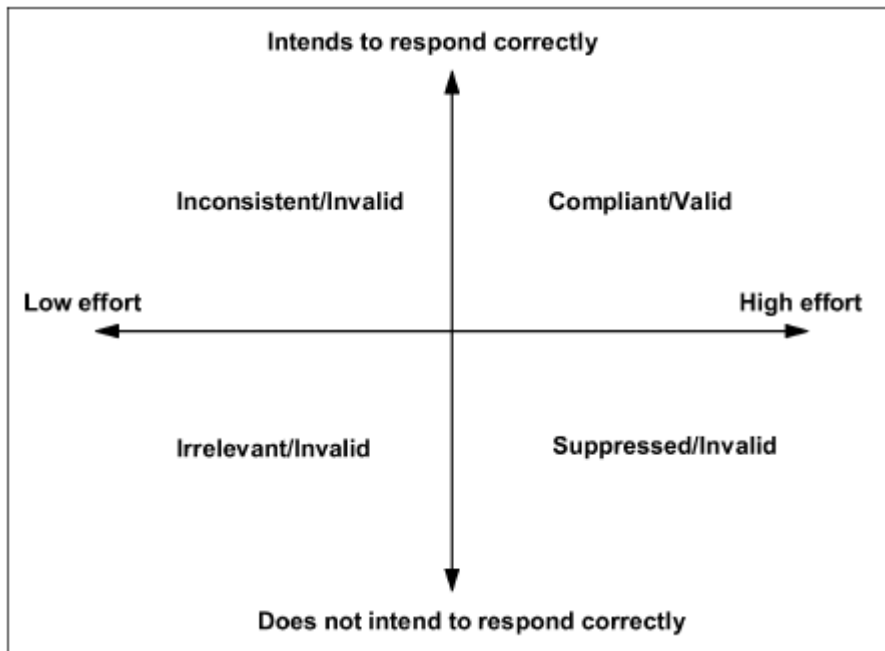


Figure 2.1 Potential categories of response style based on a cross-classification of “intent” and “effort.”
Source: Frederick, 2003

Categories include:

- Compliant (high effort, intent to respond correctly)
- Inconsistent (low effort, intent to respond correctly)
- Irrelevant (low effort, no intent), and
- Suppression (high effort, no intent).

Of note is that compliance is linked to high effort, which therefore indicates an inverse relationship with malingering. This also supports the DSM-IV definition of malingering, which includes criteria such as the lack of cooperation during the diagnostic evaluation and compliance with prescribed treatment. Of interest is that the DSM-IV does not consider malingering a psychiatric disorder, but rather a condition that may be a focus of clinical attention for reasons of non-compliance.⁽²⁾

Effort is usually linked to performance during testing, for example an individual may underperform during testing ⁽⁴⁰⁾. Even though it could indicate malingering, it may also point to misunderstanding of instructions, poor test administration technique, job dissatisfaction, learnt illness behaviour, test anxiety, fear-avoidance of activity, fatigue or side-effects of medication ^{(25) (34) (61) (62)}. It should therefore be interpreted with caution, and on a continuum from poor to outstanding.

Looking at test performance, the overlap between poor effort and symptom exaggeration has been graphically depicted as indicated in figure 2.2, which therefore also indicate that not all poor effort is necessarily considered malingering ⁽⁴⁰⁾. Even though this provides a clear picture, it does not include all the underlying concepts of the malingering construct as proposed in section 2.2.

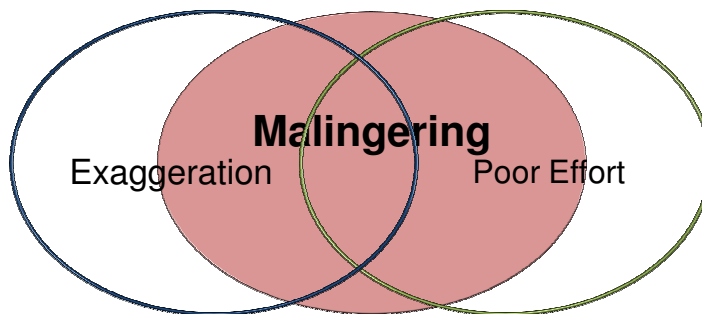


Figure 2.2 Conceptual and assessment overlay between exaggeration, poor effort and malingering

Source: Iverson, 2006

2.3.2 Use of Reported Information for Malingering Detection

The traditional clinical approach to assessment often relies heavily on reported information from the patient, significant others and treatment team. In an attempt to standardise responses and interpret information, various tools were developed to aid independent assessments. A synopsis of the approaches, strengths and weaknesses will follow:

2.3.2.1 Self-reported Measures

A detailed interview could reveal inconsistencies, but needs to be interpreted with caution ⁽³³⁾. It also provides a source of information which can be used as a basis for comparison with behavioural observations and test results. Pre-morbid clinical and socio-economic history provide opportunity for further corroboration (37), but response biases need to be taken into account as plaintiffs often rate their pre-injury functioning superior to non-plaintiffs ⁽⁶³⁾.

Often structured interviews, as opposed to the traditional interview, are used in the assessment of malingering. The Structured Interview of Reported Symptoms (SIRS) is a well-researched measure which is recommended for malingering detection ^{(59) (64)}. The SIRS was developed specifically to assess whether an examinee is malingering psychological symptoms based on strategies such as excessively endorsing rare and blatant symptoms, indiscriminately reporting symptoms, and claiming absurd or outrageous problems.

Self-reported complaints (SRC) may have a specific role in malingering detection through identification of inconsistent symptom-performance relationships. The use of SRC measures has been debated. The advantage is that it can be used to identify and analyse suspicious patterns of complaints. In addition, SRC data may be used in conjunction with objective tests data to corroborate test findings and identify discrepancies between reported symptoms and performance on objective tasks ⁽⁶⁵⁾.

The argument against the use of SRC includes the difficulty in establishing base rates. Studies by Sullivan and Richer ⁽⁶⁶⁾ have shown limited difference between the number and type of symptoms provided by personal injury claimant and head-injured patients, head-injured patients and controls, or simulator-malingers and head-injured patients. In other studies, less severely head-injured patients typically reported more symptoms than more severely injured patients ^{(65) (67)}, partly linked to poor insight of the severely injured ⁽⁶⁸⁾. Furthermore, SRC does not improve diagnostic accuracy as symptoms often overlay with those plausible in the context of additional stress induced by the process of undergoing investigation. Lastly, knowledge of symptoms among the general population is reasonably high ⁽⁶⁹⁾.

There is ongoing debate on the vulnerability of symptoms assessments using SRC measures compared to less structured approaches to complaint assessment ⁽⁶⁶⁾. Not only may self-report checklists teach malingerers how to simulate symptoms of traumatic brain injury, but may also inadvertently convince these patients that they have the symptoms listed on such checklists ⁽⁷⁰⁾.

2.3.2.2 Collateral Sources

Collateral interview data is often helpful in obtaining information the patient may be reluctant to self-report. Reservation about reporting symptoms could be due to personality, embarrassment, or the lack of self-awareness or insight ⁽³³⁾.

Given that family members make observations in the real world and unstructured setting, their observations would also be expected to have greater ecological validity than test scores. However the level of subjective complaints by malingerers was not always supported by their significant others ⁽⁶⁵⁾. This could however be linked to several other characteristics, such as level of psychological distress and negative spouse response ⁽⁷¹⁾.

2.3.2.3 Records

Review of prior collateral records is important to obtain information about premorbid functioning and previous medical conditions. Careful review of records can also assist in determining whether the complaints are consistent with the diagnosis and to determine if there have been alternative medical diagnoses which would contribute to his current complaints ⁽³³⁾.

However, it has been cautioned that the use of records only in determining an individual's condition are based upon different report writing styles, limited direct contact, focus on only part of a clinical evaluation, and provide subjective biased interpretations ⁽⁷²⁾.

2.3.2.4 The Concept of Credibility

Credibility is defined as the quality of being believable or trustworthy ⁽⁷³⁾. By its nature, it therefore implies a judgement about believability in terms of competence, moral character and trustworthiness ⁽⁷⁴⁾. This is typical of the criminological model of malingering such as in the DSM where malingerers are considered “bad persons” ⁽⁵⁹⁾.

Within the field of forensic science, credibility assessment tools include polygraph, brain fingerprinting and brain imaging ⁽⁷⁵⁾ ⁽⁷⁶⁾. The most crucial aspect of the assessment however revolves around using various data obtained through different data collection methods ⁽⁷⁷⁾ which are similar to clinical protocols when assessing malingering.

Even though the judgement of malingering has been argued to be a legal and not clinical decision, in practice, an expert may be expected to offer an impression of the credibility of the claimant ⁽³⁷⁾. The legal argument against this is that the jury may substitute the expert’s credibility assessment for its own common sense determination. Focus should rather be on the scientific validity of test instruments and results, and not just opposing views of the claimant’s credibility ⁽⁷⁸⁾. This has been strengthened in the Daubert case ruling ⁽⁷⁹⁾ which found that opinions should only be formed following interpretation of tests with proven scientific validity and reliability. This landmark case ruling has led to increased rigour in malingering research and provides strong motivation for similar practice guidelines and research initiatives in South Africa.

It is however interesting to note that the credibility of litigated cases, especially with diagnoses such as fibromyalgia, is often the deciding factor when awarding disability benefits. A recent study of judges’ perceptions of plaintiff credibility ⁽⁸⁰⁾ did not only support this, but also showed that the degree of credibility was in direct relation to the amount granted. Conceptual and legal arguments aside, the importance of assessment of credibility cannot be argued against given the impact on the outcome of disability claims.

2.3.3 Use of Tests for Malingering Detection

Malingering assessment tools and protocols vary significantly, and conclusions should not rest on any single finding ⁽³⁸⁾. It is therefore recommended practice that a variety of tools and data are used ⁽⁸¹⁾.

2.3.3.1 Behavioural Observations

Even though behavioural observation is included as part of usual assessment protocol during interviews and test administration, validity and reliability studies have been conflicting.

One of the concerns is the inability to distinguish between the truth and deception, with studies showing that people perform just above the level of chance ⁽⁸²⁾. Behavioural observations during research ⁽⁴⁹⁾ where effort, honesty and accuracy of performance was rated, indicated that experimenters rated malingerers lower in terms of level of effort and honesty than the control honest group. When forced to categorise them as malingerer or honest, they correctly classified significantly fewer participants in the malingerers groups. They therefore had a high degree of specificity but only a moderate level of sensitivity.

Other behavioural observations, such as the use of facial expressions as an indicator of pain, are considered inconsistent and unreliable as a method to identify malingerers ⁽³⁴⁾.

Research on the use of behavioural rating scales has been limited and it has been suggested that it is used for the determination of cooperation during assessment ⁽⁴⁹⁾.

2.3.3.2 Testing Specific to Diagnosis

One of the fundamental concepts of malingering detection is the correlation between test results and observations with the specific diagnosis. Diagnosis is therefore used to substantiate self-reported claims or clinical observations, as well as the development of specialised tests. The interpretation therefore considers what is expected anthropomorphically or according to the injury site ^{(18) (26) (83)}.

2.3.3.3 Specialised Testing

Several disciplines have researched the assessment of specific malingering tools, or manners in which existing tools could be adapted for that purpose.

Most notably, personality tests, such as the Minnesota Multiphasic Personality Inventory (MMPI) are considered one of the most thoroughly researched tools ⁽⁵⁹⁾, usually assessing malingering in patients who present with psychiatric symptoms ⁽⁶⁵⁾, and more recently pain symptoms ⁽³⁴⁾.

There are a number of common theoretically based deception strategies used in malingering test instruments, e.g. symptom validity/forced-choice procedures, learn and recall, floor effect, response bias/inconsistency, pattern of performance method, performance curve analysis, magnitude of error. Tests designs for the detection of malingering are intended to have a low true difficulty level, but a high face difficulty level, thereby tempting malingerers to perform poorly ^{(84) (85)}.

One of the most popular paradigms for assessment of malingering of intellectual and neuropsychological abilities has been symptom validity testing. Symptom validity tests often use a forced-choice paradigm and works on the assumption that malingerers perceive the task as more difficulty than it is, perform worse than severely impaired clinical groups, or perform at a level worse than chance ^{(86) (87)}. It is most widely used and researched, however the disadvantage is that it has led to reduced sensitivity due to coaching.

There has been conflicting opinion about the value of Functional Capacity Evaluations (FCE) with the malingering detection. FCE usually includes a multimodal assessment comprising of an interview, record review, collateral information and a comprehensive series of test activities which includes standardised and performance-based testing, checklists and rating scales, as well as real or simulated work tasks. Test measures are usually compared to job demands ^{(29) (23) (25)}. Even though the assessment process is generally protracted, it has been argued that individuals are aware that their performance will be measured and therefore do not necessarily reflect an accurate presentation ⁽⁹⁾.

Within the field of FCE, reference is often made to effort. When determining the level of effort produced, issues of motivation, consistency of performance and normality of responses are considered. According to a review of practices ⁽²⁶⁾, effort has been described as maximal, full physical effort, valid, submaximal, insincere, self-limiting, insufficient, optimal and inadequate. It has also been postulated that submaximal effort, or less than the best possible effort, equated to malingering effort ⁽³⁴⁾. The latter ties in with the reference to insincere effort as deliberate or conscious less than full effort during an evaluation ⁽⁸⁸⁾. Based on Kroemer and Marras ⁽²⁴⁾, maximal effort is considered more consistent and submaximal effort less consistent.

Research into the validity and reliability of methods used for determination of sincerity of effort in FCE has had conflicting results ^{(34) (27)}. Methods includes Waddell's nonorganic signs, documentation of pain behaviour, grip measures, coefficient of variation, correlation between musculoskeletal evaluation and FCE test, heart rate and pain intensity ⁽⁸⁹⁾. Further research to differentiate between levels of effort and for relevant patient groups has been recommended. Consistency of performance is the most common basis for determination of poor effort, with various measures of consistency used to examine this across repeated trials and similar activities, with what is expected anthropomorphically and according to diagnosis ⁽²⁶⁾.

2.3.4 Underlying Concepts of Malingering Detection

A fundamental supposition in malingering research is that the more inconsistencies an individual presents across multiple or relatively independent domains, the more likely the performance reflects deliberate efforts at misrepresentation ⁽⁴¹⁾. Therefore, the review of consistency of test presentation across different domains provides an indication of possible malingering.

Confidence in the detection of malingering is based on the pattern and magnitude of inconsistencies and whether alternative explanations for these inconsistent findings have been ruled out. This could include fatigue, medical illness, medications and emotional factors ⁽⁹⁰⁾.

Consistency assessment has therefore remained the cornerstone of malingering detection. The approach of detecting malingering by measuring deviation of presentation from reasonable expectations has been labelled discrepancy methods ⁽⁹¹⁾. This includes the following five types:

- Internal consistency/inconsistency which refers to grossly divergent performance on tests that should be highly correlated.
- Disease deficit comparability (concurrent validity) which refers to impairments which are not considered primary symptoms of a claimed disorder.
- Inconsistent with severity of injury which refers to the dose-response relations magnitude, timing and response of symptoms.
- Ecological validity discrepancy which refers to inconsistency between test scores and observed behaviours from the same domain.
- Violations of performance curves (violations of difficulty hierarchy) which refers to gross violations of difficulty hierarchy.

It is therefore important that these discrepancy methods, which include various types of validity of test performance, are considered during test development and review. It is important to distinguish between validity as a scientific concept and attempts to measure sincerity of effort during test performance. As such it has been noted that inappropriate use of the term validity occasionally occurs in disability assessment, such as “invalid effort” ⁽²⁴⁾.

Another important concept is the use of systematic multitrait-multimethod strategies ⁽⁹²⁾ during the assessment of the malingering construct ⁽⁶⁾. The use of a combination of behavioural observation, multiple sources of information together with various standardised assessment tools have been recommended by those involved with disability determination, including neuropsychologists and occupational therapists ^{(9) (49) (93)}. Participants in a study conducted by Innes and Straker ⁽⁹⁴⁾ indicated strategies analogous to those used in qualitative research, amongst other factors to ensure reliability or trustworthiness of results by the use of multiple data sources and methods of data collection, as well as structural coherence by triangulation of the results until no unexplained inconsistencies remain. Triangulation is the cornerstone of rigour in the assessment process involving the analysis of inter-relationships of both qualitative and quantitative data in order to make judgements regarding performance ⁽⁹⁵⁾.

Competency standards were defined in Australia in a study conducted by Travis ⁽⁹⁵⁾ as requiring at least three data sources and three or more data collection methods for functional capacity evaluations.

Similarly, guidelines by South African insurers also recommend that the consistency of the claimant's presentation be correlated with other medical opinions, and that clinical reasoning is employed to render an opinion following the use of various data sources and assessment tools ⁽¹⁵⁾. This provides opportunity for a multitrait-multimethod assessment strategy which conceptually underpins malingering detection.

2.3.5 Conceptual Challenges of Malingering Detection

Clinicians have been cautioned about the complexities surrounding the assessment of malingering in terms of the assessment of truthfulness, the use of malingering as a psychiatric diagnosis, and ethics-related problems associated with misclassification, as it risks impressionistic opinions and potentially negative moral evaluation of the person ⁽³⁶⁾.

The content validity of the malingering construct is problematic as malingering is context and time dependent. Put differently, it is not possible to identify all malingerers with one method as deception can occur in one or more functional domains, which can change over time ⁽⁴¹⁾. Malingerers may use different strategies to appear impaired and no indicators are capable of detecting all strategies ⁽³⁶⁾. There is therefore no instant test to ascertain whether someone is malingering despite many years of scientific research.

Therefore, the assessing clinician make inferences based on test results, which often varies dependent on personal preference and experience ⁽³⁸⁾. In addition, information obtained from interviews is dependent on clinician and interactional variables affected by personality traits, which therefore leads to argument that inferences are idiosyncratic ⁽³⁶⁾.

Another conceptual challenge unique to malingering is that malingering detection tools and techniques are subjected to the constant need for change or increased sophistication due to coaching by attorneys as well as the availability of information thereof on the internet ^{(96) (97)}.

2.4 Synthesis of Findings from Literature Review

The literature review therefore confirmed that it is difficult to clearly operationalise the construct given the interplay of underlying concepts. For the purpose of this study, it has been operationalised as indicated in figure 2.3.

- The blue-shaded row refers to the original concept analysis based on the DSM-IV diagnosis.
- The pink-shaded row refers to testing methodology and related concepts.
- The green-shaded row indicates the conceptual underpinning of malingering detection.
- Based on the cautionary approach that intent remains a legal and not clinical decision, this is not further commented on, other than ensuring that the patient is not misclassified in terms of related diagnoses (such as somatoform disorders).
- Incentive, effort, motivation and compliance is not as clearly delineated in practice.

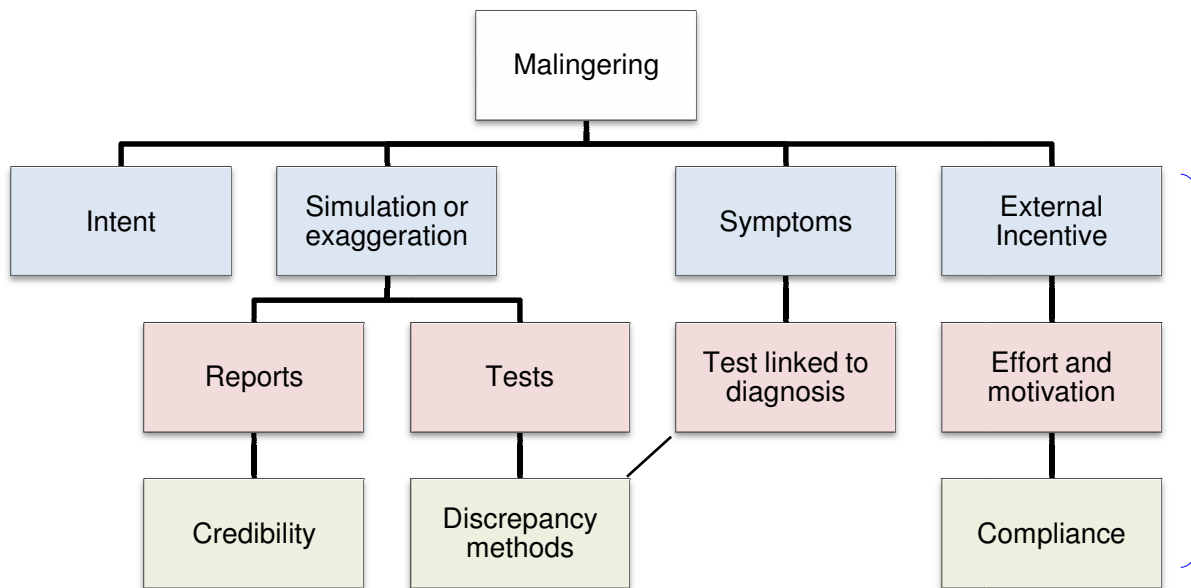


Figure 2.3 Operationalised malingering construct

2.5 Methodological Challenges in Malingering Research

The development of tests for malingering detection has presented with various challenges in terms of research design and application.

Three different research designs are generally used in the study of malingering: differential prevalence, simulation and known groups ⁽⁹⁸⁾.

- The differential prevalence design compares two groups which are expected to have different rates of malingering. Even though this can provide estimates of base rates, it cannot provide meaningful classification data.
- The simulator design requires uninjured subjects to intentionally feign impairment. This is often necessary in the early stages of test development but often has limitations with generalisability.
- The known-groups design is the best design as it requires strict operationalisation of malingering and allows for comparison between clinical patients and malingers. ⁽⁵⁹⁾

When establishing malingering groups, differentiation need to be based on external criteria derived from a systematic analysis, as well as integration of multiple sources of clinical information of behaviour in multiple domains ^{(41) (98) (86) (58)}. In addition, base rates should be established for malingering in different clinical settings ⁽³⁶⁾.

Data regarding the accuracy of detection techniques are essential for test development, clinical application and legal admissibility ⁽⁹⁶⁾. This means that the test instrument should indicate the sensitivity, specificity, and predictive values. Sensitivity and specificity are indices of overall accuracy, while predictive value is an index of confidence in a classification. Specificity and positive predictive value are especially critical in the detection of malingering. The sensitivity of all individual indicators of malingering detection will always be less than perfect to guard against excessive false positive errors. The specificity is set at a high value to minimise the occurrence of false-positive errors, that is, misidentifying someone as a malingerer who is not truly malingering. The focus of malingering research should therefore be on maximising specificity, given the importance of specificity and the inherent limitations of sensitivity ^{(96) (41)}.

2.6 Models Used in Malingering Detection

Various models and checklists have been developed to provide a framework for the detection of malingering.

Samuel and Wittenberg ⁽⁹⁹⁾ recommended that the investigation of malingering should include a search for features in four domains: motivation/circumstances, symptoms, claimant interview presentation and activity/behaviour outside the interview as indicated in table 2.2. The degree of correlation with malingering is proportionate to the number of these factors and is then reported as highly, moderately or minimally consistent with malingering. Specifically within the symptom domain, criteria such as symptom exaggeration, unusual symptoms, or those incongruent with the usual course of illness.

Table 2.4 Factors suggesting the Presence of Malingering
Source: Samuel and Mittenberg, 2005

A. Motivation / Circumstances	<ol style="list-style-type: none"> 1. Financial incentive 2. Solution to socioeconomic problems 3. Antisocial acts / behavior 4. Career dissatisfaction 5. Work conflict 6. End of career (retirement) 7. In treatment for documentation purposes 8. History of lying, malingering or dishonesty 9. Change in diagnosis to fit policy requirements
B. Symptoms	<ol style="list-style-type: none"> 1. Unusual or atypical symptoms 2. Currently asymptomatic with claim of future decompensation 3. Exaggeration or symptoms / impairment 4. Psychological test results 5. Symptoms incongruent with usual course of illness
C. Claimant Interview Presentation	<ol style="list-style-type: none"> 1. Admission of malingering 2. Uncooperative with evaluator or with divulging information 3. Discrepancies between interview reports and history / documentation
D. Activity / Behaviour Outside Interview	<ol style="list-style-type: none"> 1. Working during period of claim 2. Capacity for recreation, non-work activity 3. Functioning well except in particular line of work 4. Noncompliance with treatment 5. Surveillance

Sreenivasan, Eth, Kirkish and Garrick ⁽³⁹⁾ provided a sample checklist for the assessment of malingering head injury claimants as indicated in table 2.3, which combined specific neuropsychological testing and various consistency assessments.

Table 2.3 Assessment Guide for Amplification / Malingering in Head Injury
Source: Sreenivasan Eth, Kirkish and Garrick, 2003

	Genuine Injury	Symptoms Amplification
I. Neuropsychological testing issues a. Base rates of brain damage b. Testing comports with severity of injury c. Motivational tests abnormally positive		
II. Congruence of testing and behaviour a. Data consistent with observed behaviour in testing session b. Serial testing consistent with CNS process c. Testing data comports with medical reports d. Testing data comports with occupational or school functioning		
III. Congruence of symptoms or signs with clinical data a. Symptoms/signs comport with clinical interview b. Symptoms/signs consistent with clinical course c. Symptoms/signs consistent with past records d. Symptoms/signs consistent with physical exam e. Symptoms/signs consistent with objective labs f. Symptoms/signs consistent with collateral or surveillance data g. Medication response consistent with natural history of CNS disease h. Symptoms/signs consistent with social, occupation, or school functioning		
IV. Nonclinical factors a. No decline in income/business pre-injury b. No pending lawsuits pre-injury c. No burn-out, job actions, conflicts with co-workers, skills problems pre-injury d. Compensation less than pre-injury income e. Evaluated several times with same tests f. Context of evaluation impacting presentation g. Expectations for recovery reasonable		
Presence of Condition	No	Yes
V. Presence of psychiatric and other conditions that may contributed to amplified or atypical symptoms a. Depression/anxiety b. Personality disorder c. Conversion/somatisation d. Substance abuse e. Cumulative concussion f. Impact of chronic pain g. Impact of medications h. Impact of medical comorbidities		
VI. Miscellaneous a. Prior history of litigation b. Prior history of lying, malingering c. Prior criminal activity d. Prior job track record e. Prior responses to injury		

Conclusions are then made in terms of:

- Genuine disorder – no amplification
- Genuine disorder – with atypical symptoms or not consistent with severity
- Atypical presentation – amplification
- Atypical presentation – malingering.

This is similar to findings by FCE practitioners ⁽²⁴⁾ who consider performance consistency, commonalities of functional limitations, patterns of inconsistency and clinical substantiation when patterns of performance is assessed in view of test effort as indicated in figure 2.4. Categories of performance outcomes are then identified as: Consistent – Substantiated; Inconsistent – Unsubstantiated; Consistent – Unsubstantiated; and Inconsistent – Substantiated.

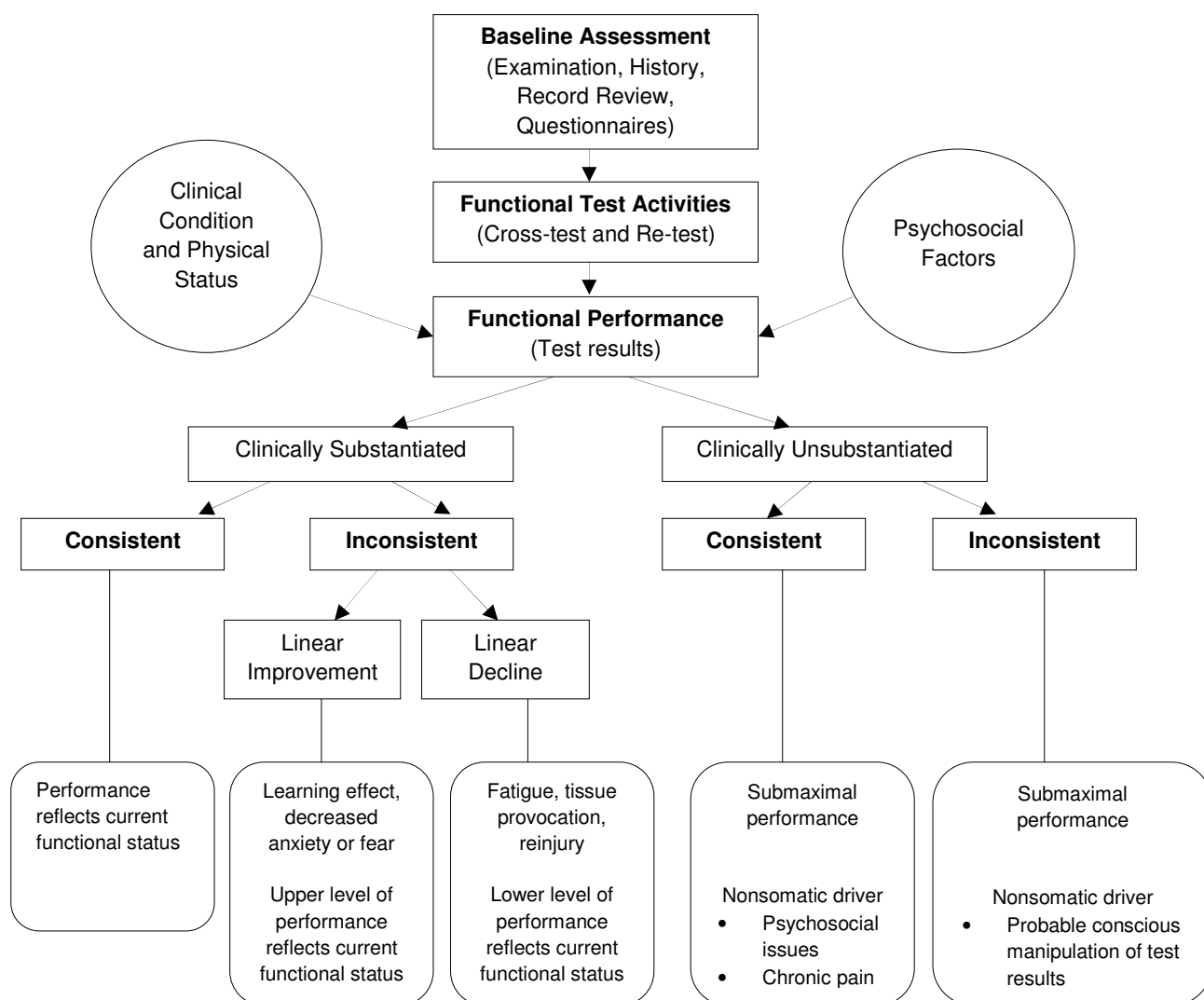


Figure 2.4 Framework for Analysing Functional Performance
Source: Genovese and Galper, 2009.

Even though the above models are helpful in establishing a framework, it does not provide quantified criteria. Slick, Sherman and Iverson⁽⁵⁸⁾ provided criteria for the diagnosis of Malingered Neurocognitive Dysfunction (MND) which has served as the basis for ongoing research studies of malingering in the field of neuropsychology. The criteria included consideration based on evidence from neuropsychological testing and self-report as indicated in table 2.4. Specific rules were assigned for a general confidence level to a diagnosis (e.g. definite, probable, possible malingering). It was considered a landmark paper as it employed a multidimensional-multimethod approach, correlated with the DSM-IV definition and allowed for the creation of “known groups” for ongoing research⁽⁸³⁾.

Table 2.4 Criteria for Malingered Neurocognitive Dysfunction (MND)
Source: Slick, Sherman and Iverson, 1999

Criteria for MND	
A	Presence of substantial external incentive
B	Evidence from neuropsychological testing 1. Definite negative response bias 2. Probable response bias 3. Discrepancy between test data and known patterns of brain functioning 4. Discrepancy between test data and observed behaviour 5. Discrepancy between test data and reliable collateral reports 6. Discrepancy between test data and documented background history
C	Evidence from self-report 1. Self-reported history is discrepant with documented history 2. Self-reported symptoms are discrepant with known patterns of brain functioning 3. Self-reported symptoms are discrepant with behavioural observations 4. Self-reported symptoms are discrepant with information obtained from collateral informants 5. Evidence from exaggerated or fabricated psychological dysfunction
D	Behaviours meeting necessary criteria from groups B and C are not fully accounted for by psychiatric, neurologic, or developmental factors
Diagnostic Categories for MND	
I. Definite	1. Presence of substantial external incentive (criterion A) 2. Definite negative response bias (criterion B1) - (i.e. worse-than-chance performance on forced-choice testing) 3. Behaviours meeting necessary criteria from groups B and C are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion D)
II. Probable	1. Presence of substantial external incentive (criterion A) 2. Two or more types of probable evidence of intent from B criteria (B2–B6) or one B criterion (B2–B6) and one or more C criteria 3. Behaviours meeting necessary criteria from groups B and C are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion D)
III. Possible	1. Presence of substantial external incentive (criterion A) 2. Evidence does not rise to the level sufficient for a diagnosis of Probable MND <ul style="list-style-type: none"> • Meets only one B criterion (B2–B6); or • Meets one or more C criteria; or • Evidence sufficient for a diagnosis of MND is present but criterion E is not met

Following the Slick criteria, a system was also developed by Bianchini, Greve and Glynn⁽⁴¹⁾ to include malingered pain-related physical and psychological complaints and deficits, over and above the cognitive deficits. Evidence from all behavioural domains is used as illustrated in table 2.5. It has been advocated for the assessment of invalid effort and exaggerated symptomatology in future research in pain malingering⁽⁵⁾⁽⁹¹⁾.

Table 2.5 Criteria for Malingered Pain Related Dysfunction (MPRD)
Source: Bianchini, Greve and Glynn, 2005

Criteria for MPRD	
A	Evidence of significant external incentive
B	Evidence from physical evaluation <ol style="list-style-type: none"> 1. Probable effort bias 2. Discrepancy between subjective report of pain and physiologic reactivity 3. Nonorganic findings 4. Discrepancy between the patient's physical presentation during formal evaluation and physical capacities documented when they are not aware of being observed
C	Evidence from cognitive/perceptual (neuropsychological) testing <ol style="list-style-type: none"> 1. Definite negative response bias 2. Probable response bias 3. Discrepancy between cognitive/neuropsychological test data and known patterns of brain functioning 4. Discrepancy between test data and observed behaviour
D	Evidence from self-report <ol style="list-style-type: none"> 1. Compelling inconsistency 2. Self-reported history is discrepant with documented history 3. Self-reported symptoms are discrepant with known patterns of physiologic or neurologic functioning 4. Self-reported symptoms are discrepant with observations of behaviour 5. Evidence from formal psychological evaluation that the person has significantly misrepresented current status
E	Behaviour meeting necessary criteria from groups B, C, and D are not fully accounted for by psychiatric, neurologic, or developmental factors
Diagnostic Categories for MPRD	
I. Definite	<ol style="list-style-type: none"> 1. Presence of substantial external incentive (criterion A) 2. Definitive evidence of intent (criterion C1 or D1) 3. Behaviours meeting the criteria for definitive intent (C1 or D1) are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion E)
II. Probable	<ol style="list-style-type: none"> 1. Evidence of significant external incentive (criterion A) 2. Two or more types of probable evidence of intent from criterion B (B1–B5), criterion C (C2–C5), and/or criterion D (D2–D6). This evidence must be well-validated and have a known error rate 3. Behaviour meeting necessary criteria from groups B, C, and D are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion E)
III. Possible	<ol style="list-style-type: none"> 1. Evidence of significant external incentive (criterion A) 2. Evidence does not rise to the level sufficient for a diagnosis of probable MPRD <ul style="list-style-type: none"> • Only one type of quantitative probable evidence of intent from criterion B (B1–B5), criterion C (C2–C5), and/or criterion D (D2–D6); or • One or more forms of qualitative evidence of intent from criterion B (B1–B5), criterion C (C2–C5), and/or criterion D (D2–D6); or • Evidence sufficient for a diagnosis of MPRD is present, but criterion E is not met

Four kinds of inconsistencies ⁽¹⁰⁰⁾ were considered as part of the Malingered Pain Related Dysfunction (MPRD) classification as indicated in table 2.5:

- nonorganic or functional findings on physical examination (exclusive of FCE),
- an inconsistency between the patients' behaviour during examination and their behaviour when they did not believe they were being observed,
- inconsistencies between the patients' subjective report of symptoms or history and their documented history, and
- evidence of submaximal effort, symptom magnification, or nonorganic/functional findings on a formal FCE.

To account for their qualitative nature, it is recommended that at least two documented inconsistencies be present to meet criteria and contribute to a diagnosis of MPRD, unless in the case of a "compelling inconsistency" ⁽⁴¹⁾.

MPRD differs from the MND criteria only in terms of how test findings can be combined to reach a diagnosis of probable malingering ⁽¹⁰⁰⁾. The weight of self-reported evidence (criterion C) has been weighted less by Slick in that, at best, self-reported evidence in the absence of test data (criterion B) can also achieve probable malingering. Bianchini, Greve and Glynn ⁽¹⁰¹⁾ have argued that the Slick criteria may not be sensitive enough and specific to cognitive dysfunction, but not necessarily relevant to the many other aspects that could be malingered. In addition he indicated that objective criteria for the evaluation of inconsistencies between the behavioural clinical presentation and aspects outside the clinical setting (such as self-report), should be developed ⁽⁴¹⁾.

Aronoff, Mandel, Genovese, Maitz, Dorto, Klimek and Staats ⁽¹⁸⁾ proposed a combined operational definition for the MND and MPRD criteria as indicated in table 2.6. This included a combined approach for chronic pain, neurocognitive, neurological and psychiatric symptoms.

Table 2.6 Criteria for Malingered Pain, Neurocognitive, Neurological, and Psychiatric Symptoms
Source: Aronoff et al, 2000

Proposed Operational Definition and Criteria for Possible, Probable, and Definite Malingered of Chronic Pain, Neurocognitive, Neurological, and Psychiatric Symptoms	
A	Presence of substantial external incentive
B	Evidence from testing <ol style="list-style-type: none"> 1. Definite negative response bias 2. Probable response bias 3. Discrepancy between test data and known patterns of pain, brain function, or neurological functioning 4. Marked discrepancy between test data and observed behaviour 5. Discrepancy between test data and reliable collateral reports 6. Discrepancy between test data and documented background history
C	Evidence from self-report <ol style="list-style-type: none"> 1. Self-reported history is discrepant with documented history 2. Self-reported symptoms are discrepant with known patterns of brain or neurological functioning or psychiatric syndrome symptoms. 3. Self-reported symptoms are discrepant with behavioural observations 4. Self-reported symptoms are discrepant with information obtained from collateral informants 5. Evidence from exaggerated or fabricated psychological dysfunction
D	Behaviour meeting necessary criteria from groups B, or D are not fully accounted for by psychiatric, neurologic, or developmental factors
Diagnostic Categories for Malingered	
I. Definite	<ol style="list-style-type: none"> 1. Presence of substantial external incentive (criterion A) 2. Definite negative response bias (criterion B1) 3. Behaviours meeting necessary criteria from groups B are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion D)
II. Probable	<ol style="list-style-type: none"> 1. Presence of substantial external incentive (criterion A) 2. Two or more types of probable evidence of intent from B criteria (B2–B6) or one B criterion (B2–B6) and one or more C criteria 3. Behaviours meeting necessary criteria from groups B and C are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion D)
III. Possible	<ol style="list-style-type: none"> 1. Presence of substantial external incentive (criterion A) 2. Evidence from self-report (one or more of criteria C1-C5). 3. Behaviours meeting necessary criteria from groups C are not fully accounted for by psychiatric, neurologic, or developmental factors (criterion D); or Criteria for Definite or Probably Malingered are met except for criterion D

2.7 Comparative Analysis of Models

The different malingering detection models in section 2.6 was therefore analysed according to the criteria within the malingering construct as operationalised in section 2.4. The results are indicated in table 2.7. This shows that the Slick criteria fulfil most of the criteria of the malingering construct compared with the rest of the models. It also provides diagnostic categories. Within literature, it has been considered the criterion standard for diagnostic classification of malingering when applied to research ⁽⁹¹⁾. Specifically, the adapted Slick criteria as proposed by Aronoff et al. ⁽¹⁸⁾ is applicable to chronic pain, neurocognitive, neurological and psychiatric symptoms and therefore provides opportunity for the multidisciplinary application. This is therefore most suitable for comparative purpose when considering the malingering construct.

Table 2.7 Comparative Analysis of Models

Concepts of malingering construct	Samuel and Mittenberg ⁽⁹⁹⁾	Sreenivasan et al. ⁽³⁹⁾	Genovese and Galper ⁽²⁴⁾	Slick et al. ⁽⁵⁸⁾
Origin	Psychiatry	Neuropsychology	Functional capacity evaluations	Neuropsychology
Intent	+	+	-	-
Exclude other causes (differential diagnoses)	-	+	-	+
Credibility				
- Self reports	+	+	+	+
- Significant others	+	+	-	+
- Medical collateral	+	+	+	+
- ADL	+	+	/	/
Discrepancy methods				
- Internal consistency	/	/	-	+
- Disease deficit comparability	+	+	+	+
- Inconsistent with severity of injury	+	+	/	+
- Inconsistency between test scores and observed behaviours	/	+	+	+
- Violations of performance curves	/	-	-	+
Effort, motivation, compliance and incentive	+(motivation, incentive and compliance)	+(motivation)	+(effort)	+(incentive and effort)
Rating	Highly, moderately or minimally consistent with malingering	Genuine or atypical with or without amplification	Clinically substantiated and consistency	Diagnostic categories

Key: + present - absent / implied

2.8 The Development of the Credibility Assessment Tool (CAT)

The Performance APGAR was developed after various experts in the field of disability evaluation performed an extensive literature review and developed a method to evaluate motivation, effort and its' role in determining residual functional capacity. The Performance APGAR has a score between one and ten, where ten is consistent with outstanding motivation to recover. Several factors such as acceptance, pain, gut intuition, acting, reimbursement are considered and scored as can be seen in table 2.8. Preliminary studies reportedly validated the inter-rater reliability and construct validity, but this has not been published and the authors recommended further more definitive studies ⁽³²⁾. Contact with the authors was attempted without success.

Table 2.8 Performance APGAR Model
Source: Colledge, Holmes, Randolph Soo Hoo, Johns, Kuhnlein, DeBerard, 2001

			Scoring Options		
			0	1	2
A	Acceptance (choose best test or average)	If this just does not get any better, what will you do?	I can't live like this	I am going to have some problems	I will live with it
		Are you satisfied with your job?	Not satisfied	Partially satisfied	Satisfied
P	Pain (choose best test or average)	Pain drawing	Nonphysiologic	Some of it is physiologic	Physiologic
		Pain behaviours score (AMA Guides table 18-5)	Exaggerated or nonphysiologic	Mixed or ambiguous	Appropriate and confirm clinical findings
G	Gut (intuition) (choose best test or average)	Credibility tool (see CAT)	Not credible	Partially credible	Credible
		Intuition of effort	Poor effort	Partial effort	Excellent effort
		Duration	Much longer than expected	Longer than expected	As expected
A	Acting (choose best test or average)	Consistency with distractions	Poor consistency	Partial consistency	Excellent consistency
		Waddell signs	More than 2 Waddell signs	2 Waddell signs	0-1 Waddell sign
		Grip-strength testing	Unreliable grip strength (high variance, etc.)	Partial validity	Reliable grip strength
R	Reimbursement	Compensation/litigation	Someone else liable WC, PI, Disability Application Attorney Representing	Someone else liable WC, PI, Disability Application	No one Liable
		Total Performance APGAR Score = _____ (Add A, P, G, A, R sections for a maximum of 10)			

One of the key aspects of the Performance APGAR is the CAT ⁽³²⁾. As displayed in table 2.9, it assesses the credibility of the functional limitations and their effect on residual functional capacity by awarding a determination of credible, partially credible and not credible based on:

- Effects of symptoms or impairment on performance of ADL
- Type, dosage, effectiveness and side effects of medications
- Treatment sought and received
- Opinions that have been recorded by professionals who have treated and/or examined the patient
- Inconsistencies or conflicts in the allegations, statements, or medical evidence in the file.

Table 2.9 **Credibility Assessment Tool**
Source: Colledge et al., 2001

		Not consistent with the objective evidence and/or expected outcome/severity (0 points)	Partially consistent with the objective evidence and/or expected outcome/severity (1 point)	Fully consistent with the objective evidence and/or expected outcome/severity (2 points)
A	Impact of symptoms or condition on ADL			
B	Type, dosage, effectiveness, and side effects of medications			
C	Treatment sought and received			
D	Opinions about function given by other treating and examining sources in the file			
E	Inconsistencies or conflicts in the allegations, statements, or medical evidence in the file (<i>sic</i>)			
			Total credibility score= _____ (0-10)	
	Credibility determination: Total credibility score of 0-3= Not credible Total credibility score of 4-7= Partially credible Total credibility score of 8-10= Fully credible			Result of credibility determination to be used in the APGAR table

2.9 Summary

The malingering construct is challenging given the clinical and ethical complexities linked to the assessment of credibility and potential misclassification. The ability to develop a clearly defined tool is challenged by the context and time dependent nature of assessments, as well as the impact of coaching by attorneys and information available on the internet.

Operationalising the malingering construct was therefore based on diagnostic criteria of the DSM-IV diagnosis, testing methodology and conceptual underpinning. In summary, it considers intent insofar as it other diagnoses and psychological causes are excluded, credibility of reported information, discrepancy method to consider inconsistencies with the diagnosis and test results, as well as incentive, effort and compliance.

Given that the CAT will be used as a screening tool by multidisciplinary disability evaluating professionals, various models used in malingering detection was reviewed. It was then compared with the underlying concepts of the malingering construct. The results indicated that the Slick criteria for MND fulfilled the most criteria and was therefore most suitable for comparison with the CAT. The adapted Slick criteria as proposed by Aronoff et al. ⁽¹⁸⁾ is applicable to chronic pain, neurocognitive, neurological and psychiatric symptoms and therefore provides opportunity for the multidisciplinary application to determine the content and construct validity of the CAT.

The methodology in which this will be achieved will be discussed in the following chapter.

CHAPTER 3

METHODOLOGY

3. METHODOLOGY

3.1 Introduction

The validity of an instrument is the determination of the extent to which it actually reflects the abstract construct being examined. Whereas it has traditionally been subdivided into three categories of content, criterion and construct validity, this has changed in that all three types are considered evidence of construct validity by the American Psychological Association (APA) ⁽¹⁰²⁾. According to the APA, validity is a unitary concept that considers the appropriateness, meaningfulness and usefulness of the specific inferences made from instrument scores. It therefore considers the degree to which both evidence and theory support the interpretations of test scores entailed by proposed uses of tests ⁽¹⁰³⁾.

Therefore, the initial step in reviewing the validity of the CAT was to consider the theoretical underpinning of the various concepts as highlighted in the literature review. Although often viewed as a similar concept ⁽¹⁰³⁾, the face and content validity will be delineated to review scale construction and evidence based on the content of the CAT respectively.

Through the extensive literature review, the Slick criteria ⁽⁵⁸⁾ were identified as most suitable for comparison with the CAT. As the objective of the Credibility Assessment Tool (CAT) included a multidisciplinary screening tool for symptoms of pain and depression, the adapted Slick criteria as proposed by Aronoff et al. ⁽¹⁸⁾ was used as it can be applied to pain, neurocognitive, neurological and psychiatric symptoms. The adapted Slick criteria were therefore used to measure the construct and concurrent validity of the CAT. The methodology employed to explore the relationships between these tools will therefore be described.

(For ease of reference adapted Slick criteria as proposed by Aronoff et al. ⁽¹⁸⁾ will be referred to as Slick throughout.)

3.2 Method of Inquiry with Literature Review

The method in which the theoretical concepts linked to the malingering construct were measured is discussed in the following section.

3.2.1 Face Validity

There are no specific guidelines for the review of face validity as it refers to a superficial opinion about whether the tool is a valid measure of content “on the face of it”⁽¹⁰²⁾. For the purpose of this study, face validity focused on scale construction⁽¹⁰⁴⁾ in terms of:

- Purpose
- Instruction
- Item selection
- Weighting
- Level of measurement.

3.2.2 Content Validity

Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose⁽¹⁰⁵⁾. Therefore, the content validity was measured against the malingering construct as defined in the extensive literature review in section 2.4. In order to measure this, the same concept analysis as applied to the various models was used and compared with the Slick criteria, which was identified as the criterion standard of diagnostic categories. This is considered relevant given that the CAT is also considered a screening tool for diagnostic categorisation.

Specific to malingering, it should be noted that content validity is difficult to determine given that malingering is often situation dependent, i.e. malingerers do not feign or exaggerate in the same manner for the same diagnosis and test every time. Therefore, the utility of the CAT should be

relevant to different situations, whilst allowing for flexibility of clinicians and available test equipment.

3.3 Study Design to Measure Construct and Concurrent Validity

The study design was a descriptive analytical design which was performed retrospectively with a report review from 2009 to 2005 where the results of the CAT and adapted Slick criteria were compared. Given that the results are compared between a new and a well-reputed tool, concurrent validity can be established.

The research design is considered descriptive analytical as it observes and describes malingering detection with the use of the CAT, but also analyses the underlying concepts. The analysis includes quantifying and describing the relationship between concepts, but does not include formal hypothesis testing or complete analytical designs, such as for example case-control design. It was considered suitable given the exploratory nature of the research question.

3.4 Population and Sample

Due consideration was given to the different research designs which are usually used in the study of malingering (see section 2.5). Malingering research in South Africa thus far has been limited to simulator designs, where subjects were asked to intentionally feign impairment^{(106) (107)}. One of the limitations thereof is generalisability, which was considered a significant limitation in this research study, given that the objective of the CAT would be to screen for multiple diagnoses. A known-group design was therefore attempted, given that this is generally considered better in terms of operationalising various levels of malingering⁽⁵⁹⁾. This approach therefore complies with international malingering research, which often uses the Slick criteria to define malingering groups based on external criteria and multiple sources of information. Even though the strict definition of a known-group design was not followed in that specific groups were not delineated based on, for example, case and control, this research design was still considered stronger than simulator designs.

The sample consisted of claim correspondence linked to referrals for functional capacity evaluations (FCE) by the researcher. Claim correspondence included information on their job and medical condition, including medical reports and test results. The results of the FCE were used to determine the level of disability in accordance with the insurer's policy requirements. These policies were limited to private insurance companies, whether through personal insurance or those organised by employers.

Claims correspondence were archival and collected from 2009 to 2005 during the course of the occupational therapy practice of the researcher who performed assessments on a national basis for all of the South African insurance companies.

3.4.1 Inclusion Criteria

Data was included of those claimants that reported depression and pain-related disability regardless of diagnosis or aetiology to allow for variation in subjective experiences and generalising. As it was often difficult to delineate the diagnosis or cause, especially with pain-related depression, all the cases with depression and pain-related symptoms were included, even if a final diagnosis was not reached. The focus was therefore on reported symptoms, rather than diagnoses.

Apart from the fact that the above-mentioned symptom constellations are the largest cause of claim payouts in South Africa ⁽⁷⁾, it has also been rated as areas with high prevalence of malingering as indicated in section 2.2.3.

3.4.2 Exclusion Criteria

The exclusion criteria considered factors, such as the availability of data for analysis and prevalence of malingering as reported in literature. As an example, neurological conditions were excluded by the limited use of neuropsychological assessment for insurance claims, which would make the interpretation of the Slick criteria difficult.

The exclusion criteria also prevented that the variability of data would minimise statistical interpretation. It was not due to expected difficulty or contraindication of its use with the CAT. It was therefore not considered to lead to sampling bias, as it strengthened statistical analysis. The results can also be used comparatively with other diagnoses, should cut-off scores be developed for different diagnoses.

Exclusion criteria for the study included the following:

- Neurological conditions, unless pain was a significant contributing factor, such as neuropathic pain or headaches.
- Brain injury including cerebrovascular accidents and traumatic brain injury.
- Hand injury, unless pain developed subsequently, such as complex regional pain syndrome.
- Psychiatric disorders without any depressive features, such as anxiety disorders without depressive symptoms, substance abuse or impulse disorders.

3.5 Sampling Method and Biases

For reasons of availability and access, the frame from which the sample was drawn included referrals of claimants with pain and/or depression for FCE, performed by the researcher from 2009 to 2005.

Even though the sampling method of using all available cases led to sampling bias, this was negated by the fact that data was available from most insurance companies for claimants throughout South Africa. Research was also with a known-group design, i.e. where actual cases were used as opposed to subjects simulating symptoms, which strengthened its scientific relevance in South Africa. Prior attempts to use known malingerers on the grounds of surveillance information was prevented by concerns regarding the ethical access of information and availability of data in these instances.

A saturated sample of convenience was used, given that available data of suspected malingerers were limited. As the researcher had a high referral rate of suspected malingerers due to her professional interest, there was concern regarding diagnostic suspicion bias and cognitive

dissonance bias. This was addressed by sifting statistics kept for private practice administration on a Microsoft Excel™ worksheet to a general population worksheet where 218 cases with symptoms of depression and pain were isolated. Information was captured on the data capture sheet (see section 3.6 for further detail) using the corresponding number on the general population worksheet. Following omission of data, which did not comply with the inclusion criteria or was incorrectly titled on the general population sheet, the unused numbers were deleted upon completion. The data capture sheet therefore included data reference numbers which could not be traced back to the original general population worksheet. A total number of 184 cases met the inclusion criteria and were identified for analysis.

In addition, the study method of using familiar cases could have led to recall bias. In an attempt to address this measurement bias, a colleague with similar professional experience performed a peer check of a random sample of completed analyses. The peer check included the review of 20 cases according to the CAT and adapted Slick criteria by using a similar data capture sheet. Given that there are no standardised wording or training requirements, a session was arranged where the literature was provided and a case example was performed together.

The intra-class correlation (ICC) was calculated between the researcher and peer to establish the inter-rater agreement. The ICC agreement score was relatively low (ICC=0.516) in terms of the agreement between CAT scores as indicated in figure 3.1. However, even though the correlation was low between specific CAT scores, the level of consistency between ratings was higher (ICC=0.654). This, in effect, meant that scoring was not consistent per numerical rating, but that there was correlation of scoring of the overall CAT classification between researcher and peer. The interrater reliability was found to be Kappa = 0.68 (p=0.09836), which is interpreted as fair agreement according to Landis & Koch ⁽¹⁰⁸⁾.

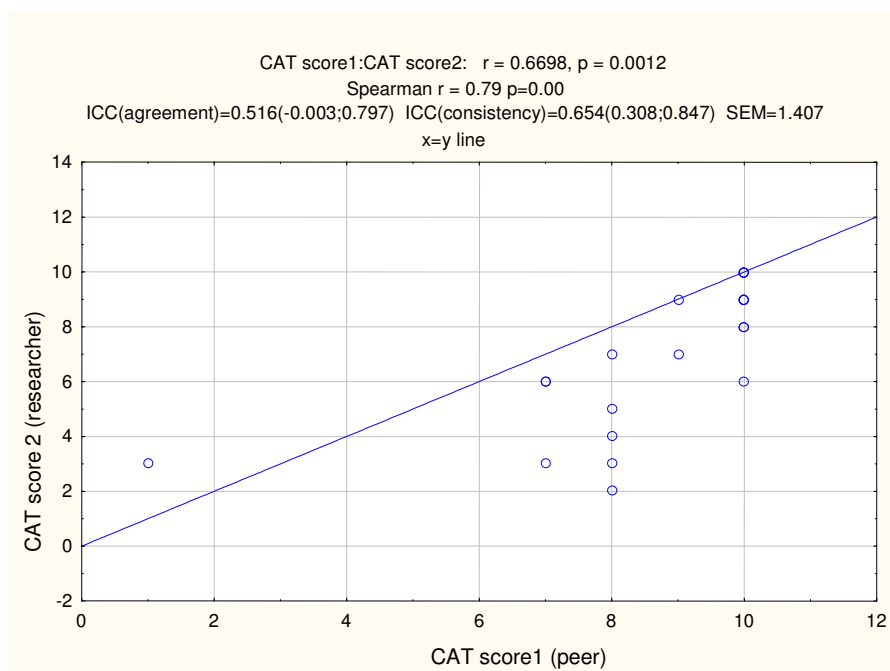


Figure 3.1 Correlation with Peer Check

The poor correlation in terms of the numerical rating requires further investigation, however it is postulated that the limited test instructions had an impact on inter-rater reliability. Owing to the fact that the ICC consistency score indicated overall correlation, the peer check was not found to conclusively indicate recall bias, however highlighted concerns with regards to reliability.

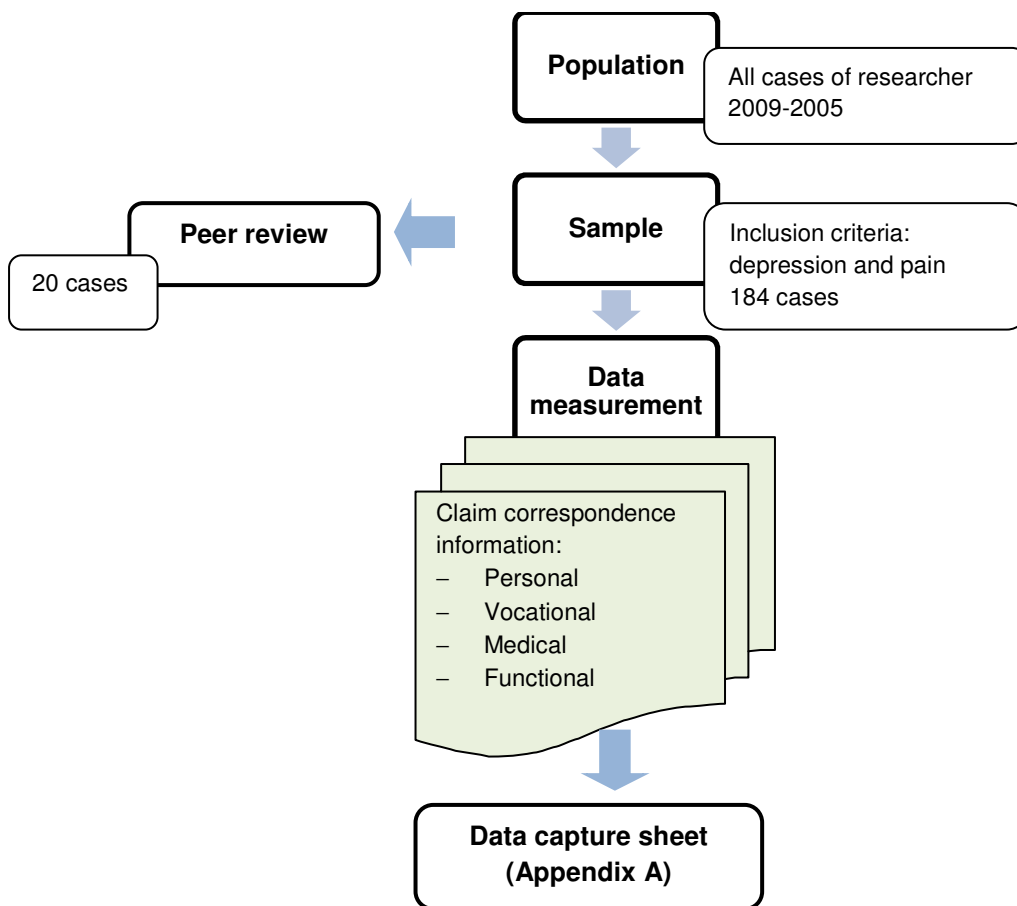
3.6 Method

Data was obtained from the claim correspondence and researcher's FCE reports from 2009 to 2005. Typically the data included:

- Personal demographical information, such as age, personal and family history.
- Vocational information, such as education, training, work history, work demands and productivity reports.
- Medical information, which at least included reports from the attending general practitioner and specialist, but could also have included independent specialist opinions and rehabilitation progress reports. Available test results were included, such as radiological evidence, or psychometric test results (if available).

- Functional information regarding symptoms or Activities of Daily Living, often available on completed claim forms or in FCE reports.
- Results of FCE which were predominantly performed by the researcher.
- Collateral information from family members, claim consultants or forensic investigators.

As the original scoring criteria of the CAT and Slick were used, there was no need to develop further instruments or questionnaires or to perform a pilot study. An overview of the method is depicted in figure 3.2.



Demographical information				Slick criteria (Appendix B)				CAT criteria (Appendix B)				
Claimant number	Gender	Age	Diagnosis	Criterion A (Incentive)	Criterion B (Testing)	Criterion C (Self Reports)	Criterion D (Alternative factors)	Criterion A (ADL)	Criterion B (Medication)	Criterion C (Treatment)	Criterion D (Collateral - med)	Criterion E (Allegations, statements and evidence)

Figure 3.2 Method Flowchart

3.7 Measurement Instrument

The data was captured on a data capture sheet in Microsoft Excel™ (as indicated in appendix A) after reading through the available claims correspondence. The variables included:

- Claimant number (to ensure anonymity).
- Gender
- Age categories
- Diagnosis of pain, depression or fibromyalgia
- Scores for each of the adapted Slick criteria
- Slick classification in terms of probable, possible and not malingering (refer to table 2.6)
- Scores for each of the CAT criteria
- CAT classification in terms of fully, partial and not credible (refer to table.2.9).

Emerging data and/or trends were captured in separate columns and reviewed for relevance. Two additional factors were eventually included: injury and financial incentive.

3.8 Interpretation of Measurement Criteria

Given that no standardised instructions were provided in the CAT by Holmes ⁽³²⁾, the manner in which the criteria were interpreted will be indicated hereafter for replication purposes. The same interpretation (as indicated in appendix B) was also provided to the peer during the peer review. The interpretations were based on available information in the literature review, as well as accepted practice in the insurance industry.

3.8.1 Slick criteria

Even though the adapted Slick criteria include definite, probable or possible subcategories as can be seen in table 2.6, the requirement for definite malingering includes psychometric or neuropsychological testing to establish definite negative response bias (B1 criterion). Given that this could not be consistently applied with all retrospective data as all claimants were not necessarily referred for such testing, the category of definite malingering was omitted. The Slick classification categories were therefore Not, Possible and Probable Malingering.

This was not considered problematic given that the CAT is a screening tool and hence Definite Malingering is not necessarily a realistic outcome. In addition, it was also found that other studies ⁽¹⁰⁹⁾ ⁽¹¹⁰⁾ have not always utilised all the criteria, dependent on the research objective and methodology. These studies concluded that definite and probable malingering are essentially indistinguishable, especially when measured against standard of “more probably than not” or “to a reasonable degree of scientific certainty” on the preponderance of evidence from a medico-legal standpoint ⁽¹⁰⁹⁾.

3.8.1.1 Slick criterion A (Incentive)

The Slick criterion A, which is defined as “the presence of substantial external incentive” ⁽⁵⁸⁾ ⁽¹⁸⁾, and applicable to all the cases given the nature of the referrals, was extended to include information regarding financial and litigious incentive which was one of the emerging factors during the data analysis. Additional literature review confirmed that financial incentive motivates intentional symptom exaggeration. In countries where there are no compensation, patients with whiplash syndrome had a lower incidence of chronic neck pain and headache which did not differ significantly from a control group of uninjured healthy subjects ⁽¹¹¹⁾. Following a change in rules for the compensation of pain and suffering in Canada, the claims incidence declined by 43% for men and 15% for women ⁽¹¹²⁾. The same trend was noticed in Australia ⁽¹¹³⁾.

In the USA, claims against Federal workers compensation laws, with high financial incentive, showed higher rates of diagnosable malingering than claims under State law, with its limited incentive. Furthermore, when the ratio between a worker’s compensation income and salary

increases, the frequency and duration of the claims increases as well ⁽¹¹⁴⁾. A meta-analysis of 32 studies ⁽¹¹⁵⁾ showed that claimants receiving compensation consistently reported more pain than the control claimants. The finding that financial compensation motivates malingering was also found in brain injuries ⁽¹⁰¹⁾.

In South Africa, disability benefits are usually paid in terms of monthly income replacement as well as lump sum awards. Whereas the technical detail of monthly income replacements are not always understood by claimants, there is a perception that lump sum benefits are more beneficial. The use of lump sum disability benefits has been found to act as an extra incentive to exaggerate symptoms ⁽¹⁴⁾.

Additionally, there is also more incentive with disputed or litigious cases where attorneys or the Ombudsman for Long-term Insurance is involved ^{(14) (116)}.

Given the above factors, the degree of financial incentive in terms of monthly benefit, lump sum benefit and litigious cases were also measured. This would not have impacted on the Slick scoring classification, but allowed for additional comparison on the overall outcome.

3.8.1.2 Slick criterion B (Testing)

Criterion B was measured according to each available criteria and the corresponding abbreviation inputted on the data sheet, as:

- Probable response bias (RB): In this instance, results were indicated if performance on one or more indices designed to measure exaggeration or fabrication were consistent with feigning such as during grip strength testing.
- Discrepancy between test data and known patterns of brain functioning (TD≠KP): If test data was markedly discrepant from accepted models of normal, such as improved concentration and accuracy over time with a depressive subject.
- Discrepancy between test data and observed behaviour (TD≠OBS): This is further qualified if performances on two or more tests within the same domain are discrepant with observed level

of function, such as range of movement and functional testing of squatting is discrepant with observed transfers.

- Discrepancy between test data and reliable collateral reports (TD≠COLL): It is required that performance on two or more tests within the same domain are discrepant with day-to-day level of function as described with at least one collateral informant. In this regard, the correlation of video surveillance of gait pattern with the available gait analysis by the physiotherapist and muscle strength testing or electromyography studies.
- Discrepancy between test data and documented background history (TD≠HIST): Poor performance on two or more standardised tests of function or symptoms within specific domain that is inconsistent with medical history. Standardised questionnaires, such as the Beck Depression Inventory and the Hospital Anxiety and Depression Scale ⁽¹¹⁷⁾, should therefore be consistent with the mental status examination according to the medical history as documented by the psychiatrist.
- Not applicable (NA): This would be indicated if none of the above factors are relevant.

3.8.1.3 Slick criterion C (Self Reports)

Criterion C was measured according to each available criteria and the corresponding abbreviation inputted on the data sheet, as:

- Self-reported history is discrepant with documented history (SR≠HIST): A marked difference between the documented and self-reported history is indicated, such as omitting to report a pre-existing medical problem which coincides with the onset of disability cover.
- Self-reported symptoms are discrepant with known patterns of functioning (SR≠KP): In this instance, focus is on the number, pattern and severity of symptoms, such as for example, the likelihood that self-reported memory impairment due to a major depressive episode would prevent one from recalling one's name or date of birth.
- Self-reported symptoms are discrepant with behavioural observations (SR≠OBS): Observed behaviour, rather than test behaviour, is considered, such as the ability to bend to retrieve an object despite self-reported complaints that pain prevents bending.

- Self-reported symptoms are discrepant with information obtained from collateral informants (SR≠COLL): Collateral information on file from the treating team is contrary to the reported symptoms.
- Evidence from exaggerated or fabricated psychological dysfunction (EXAG): Exaggerated response of psychological adjustment or distress, such as indicated in scores of catastrophising which exceeds normative values for chronic pain patients.
- Not applicable (NA): This would be indicated if none of the above factors are relevant.

3.8.1.4 Slick criterion D (Alternative Factors)

Indications were made on the data capture sheet when behaviour meeting the criteria for B (testing) and C (self-reports) were not fully accounted for by psychiatric, neurologic, or developmental factors.

3.8.2 CAT criteria

The test administration of the CAT did not include any specific instructions, whereas the scoring did not indicate specific criteria. Even though the scoring method and three case examples provide by Colledge AL, Holmes EB, Randolph Soo Hoo E, Johns RE, Kuhnlein J, DeBerard S ⁽³²⁾ were perused, this was not always mutually exclusive and/or exhaustive. Additional criteria included:

3.8.2.1 Effects of symptoms or impairment on performance of ADL

As no standardised checklist or rating scale was recommended, information from available self-reported and/or self-administered questionnaires were used. With regards to guidelines used by disability examiners, also in South Africa, it is noted that the performance of Activities of Daily Living (ADL) implies sustained and regular performance ⁽¹⁴⁾.

3.8.2.2 Type, dosage, effectiveness and side effects of medication

Medication use was evaluated with regards to the need for large doses or multiple medication, addictive behaviour, pattern of increasing use, alternative medicine, side-effects, cessation of any medication and self-medicating behaviour. In this regard, the suggested Life Offices Association (LOA) guidelines for depression and back pain were used given the clear criteria for optimal treatment ⁽¹²⁾ ⁽¹⁴⁾.

3.8.2.3 Treatment sought and received

Reasonable treatment in terms of professional help, compliance, attempts to treat the condition or find relief by attempting multiple treatments, frequency of visits, was evaluate. Treatment was interpreted to include rehabilitation programmes. The LOA guidelines for reasonable medical treatment was used given the cognisance of cost, risk, success rate, and what the average reasonable patient would be prepared to undergo (“the test of the reasonable man”) ⁽¹⁴⁾.

3.8.2.4 Opinions that have been recorded by professionals who have treated and/or examined the patient

The recommendations by Colledge et al. ⁽³²⁾ were followed in terms of the weight of the opinion given. Referrals from treating and/or independent professionals are usually included with the referral. In this regard, the authors recommend that:

- The opinions of practitioners who have examined the patient are given greater weight than the opinions of those who have not (e.g. insurance company file reviews).
- Treating sources are given greater weight rather than providers of one-time examinations.
- A source that provides supporting evidence to substantiate the opinion about functional ability should be given more weight than should a source that does not have supporting evidence.
- Opinions most consistent with the preponderance of evidence are given greater weight.

- The opinion of a specialist in the field may be given greater weight than would that of a generalist, even if the length of treatment by the specialist was much less. Furthermore, the opinion of a physician who is more familiar with the demands and tasks in the workplace is likely to be given greater weight than would the opinion of a physician who is unaware of such demands.

3.8.2.5 Conflicts in the allegations, statements or medical evidence in the file

The criterion reads that the “inconsistencies or conflicts in the allegations...” should be reviewed in terms of its’ consistencies. This appears to be an error, as the intention was likely that the “...allegations, statements or medical evidence...” should be evaluated in terms of consistencies. The presumed intent of the criterion was used by deleting the words “inconsistencies or conflicts in the”. This allowed the “conflicts in the allegations, statements or medical evidence” to be reviewed in terms of the level of consistency.

Furthermore, the authors indicate the review of all aspects for inconsistencies or conflicts including:

- The degree to which the allegations are consistent with the objective evidence.
- The history given at different examinations.
- The consistency of the history of the injury/illness, the onset and duration of symptoms, and the functional effects on ADL as reported to various medical professionals.

3.9 Data Analysis

Data was analysed by comparing the different variables of the Slick and CAT criteria, as well as emergent factors, to determine associations between concepts. This included:

- The association between self-report on the CAT classification
- The association between testing on CAT classification, with specific reference to discrepancy methods such as probable response bias, known patterns of functioning, observed behaviour, collateral reports and background history
- The interplay between ADL on Slick classification to determine the importance of delineation thereof with the Slick classification

- The association between compliance with treatment and medication and the Slick classification
- The association between medical collateral information on the Slick classification.

The results of the analyses were used to determine the construct validity, whereas testing concurrent validity was focused on:

- The overall comparative classification between the Slick and CAT total scores
- The analysis of the scoring classification between the CAT and Slick.

3.10 Ethics

Ethics in malingering research is of utmost importance ⁽³⁶⁾ ⁽⁴⁰⁾. Specific to this research study, the legal ownership of the claims correspondence, which refers to medical and financial information about claimants, was carefully considered.

Within the disability claims process, it should be noted that claimants authorise insurers to investigate any information related to their medical condition when submitting a claim. It is argued that consent to the disclosure of any information reduces the incidence of fraudulent claims, and hence benefit clients as premiums can be limited. This may change in the future with the promulgation of the Protection of Information Bill ⁽¹¹⁸⁾, however is probably excluded as personal information has been de-identified to the extent that it cannot be re-identified again.

Disability determination reports performed by independent medical service providers are legally owned by the insurer. Therefore informed consent (as indicated in appendix C) was obtained from the insurers on the grounds that:

- The study consisted of a paper-based analysis of the researcher's own reports and accompanying referral information from 2009 to 2005. No contact was established with the claimant, his employer or treating medical or rehabilitation practitioners.

- The research study would have no impact on the outcome of the payment of the claim as retrospective data was used and the research findings was not documented in the claim file.
- No personal identifiers were recorded on the data capture sheet used for analysis.
- Analysis was performed collectively, thereby insuring anonymity of all cases and insurers.

Ethical approval for the study was obtained from the Health Research Ethics Committee at Stellenbosch University with the reference number of N08/02/045, as indicated in appendix D.

3.11 Summary

The methodology was described in which the construct validity of the CAT will be reviewed. In this regard, it should be noted that the 1999 APA standards were used in that construct validity is considered a unitary concept. Nevertheless, the method in which face and content validity was reviewed where described in terms of focus on scale construction and the operationalised malingering construct respectively.

The methodology was also described in which the concurrent and construct validity of the CAT would be reviewed by comparison with the criterion standard of the adapted Slick criteria.

The research design was a descriptive analytical design of known-groups, which was performed retrospectively with a report review from 2009 to 2005 from insurance referrals to the researcher. Informed consent was obtained from the insurance companies given the legal ownership of the claim correspondence.

A saturated sample of convenience of 184 cases with depression and pain as predominant symptoms were analysed. Bias was minimised through omission of personal identifiers and the use of a peer check of 20 random cases. Results indicated consistency between ratings between the researcher and peer (ICC=0.654), however poor correlation between numerical ratings

(ICC=0.516) were obtained. Responses were suggestive of poor inter-rater reliability of the CAT, rather than recall bias.

Measurement consisted of capturing related data on a data capture sheet by following the guidelines from the respective authors ⁽¹⁸⁾ ⁽³²⁾. The guidelines were further defined to ensure that the study can be replicated by information obtained from the literature review and usual insurance practice. The data method was depicted in figure 3.2.

The analysis of the results is reported in the next chapter.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4. DATA ANALYSIS AND RESULTS

4.1 Introduction

In this chapter, the results of the face, content and concurrent validity of the CAT will be measured by the comparison with concepts from the literature review and the adapted Slick criteria ⁽¹⁸⁾. The statistical methods used will be described, followed by the analysis of interrelationships between the various CAT and Slick criteria. References to the various criteria can be found on appendix B. The collective results were then used to determine an overall view of construct validity.

4.2 Results of Face Validity of CAT

Face validity was reviewed in terms of purpose, instruction, item selection, scaling and weighting as well as level of measurement.

4.2.1 Purpose

The purpose of the CAT is to assess the credibility of the functional limitations and their effect on residual functional capacity, following which a determination is made in terms of credible, partially credible and not credible. As such, it is considered a predictive instrument in that it uses certain criteria to classify individuals in order to predict if they have a certain trait in comparison to set criteria.

The use of the term “credibility” in the title and purpose of the CAT requires further mention. The term “credibility” was retained in reference to the original Social Security Administration (SSA) guidelines. Reference is made to the term “consistency” in the SSA guidelines as an indication of the credibility of an individual’s statements ⁽¹¹⁹⁾. The developers of the CAT also concluded that the credibility determination is better labelled a “consistency assessment” given that inconsistencies

and conflicting statements contributes to the overall credibility assessment ⁽³²⁾. This therefore underscores the conceptual underpinning of the malingering construct.

4.2.2 Instruction

There are no standardised instructions for the CAT, although an article summarises the credibility determination process in terms of the Social Security Administration (SSA) rules. Guidelines and examples are given in terms of the Performance APGAR and indicate that it can be used “for many different types of impairments” and scored “at each visit or over a series of visits”. Scoring interpretation is indicated in terms of that “each of the five areas should be scored 0, 1 or 2 points. The points are then totalled for an overall credibility score...” The resultant three credibility determinations of credible, partially and not credible is then summarised in terms of overall credibility and consistency with diagnosis and objective evidence.

4.2.3 Item selection

The five criteria which are considered in terms of the CAT include:

- Effects of symptoms or impairment on performance of ADL
- Type, dosage, effectiveness and side effects of medications
- Treatment sought and received
- Opinions that have been recorded by professionals who have treated and/or examined the patient
- Inconsistencies or conflicts in the allegations, statements, or medical evidence in the file.

The original factors described in the SSA guidelines included:

- The individual's daily activities;
- The location, duration, frequency, and intensity of the individual's symptoms;
- Factors that precipitate and aggravate the symptoms;

- The type, dosage, effectiveness, and side effects of any medication the individual takes or has taken to alleviate symptoms;
- Treatment, other than medication, the individual receives or has received for relief of symptoms;
- Any measures other than treatment the individual uses or has used to relieve pain or other symptoms; and
- Any other factors concerning the individual's functional limitations and restrictions due to symptoms.
- The medical signs and laboratory findings;
- Diagnosis, prognosis, and other medical opinions provided by treating or examining physicians or psychologists and other medical sources; and
- Statements and reports from the individual and from treating or examining physicians or psychologists and other persons about the individual's medical history, treatment and response, prior work record and efforts to work, daily activities, and other information concerning the individual's symptoms and how the symptoms affect the individual's ability to work.

Even though the CAT criteria summarises the original factors of the SSA guidelines, it does not provide detailed information in terms of the criteria, such as for example the impact of the effect of symptoms on performance of ADL (criterion A) which should consider the “intensity, persistence or functionally limiting effects of symptoms”.

It also does not provide detailed information regarding interpretation of behaviour, such as for example reasons for not seeking treatment in a consistent manner (criterion C), which was given in the SSA guidelines as:

- The individual's daily activities may be structured so as to minimise symptoms to a tolerable level or eliminate them entirely, avoiding physical or mental stressors that would exacerbate the symptoms. The individual may be living with the symptoms, seeing a medical source only as needed for periodic evaluation and renewal of medications.
- The individual's symptoms may not be severe enough to prompt the individual to seek ongoing medical attention or may be relieved with over-the-counter medications.
- The individual may not take prescription medication because the side effects are less tolerable than the symptoms.

- The individual may be unable to afford treatment and may not have access to free or low-cost medical services.
- The individual may have been advised by a medical source that there is no further, effective treatment that can be prescribed and undertaken that would benefit the individual.
- Medical treatment may be contrary to the teaching and tenets of the individual's religion.

The scale construction in terms of item selection is considered adequate as it includes most relevant criteria of credibility.

4.2.4 Scaling and weighting

A judgemental method of scaling is used in that the five criteria are subjectively rated based on available information, rather than statistical patterns obtained from the information. The weighting is therefore implicit as five criteria are scored, and the total score is obtained by adding the scores for each criterion together. The scoring quantifiers (i.e. not, partially and fully consistent) attempt to provide weighting based on consistency criteria which is a key underlying concept of malingering. It also considers the comparability of the injury or illness, as well as the severity thereof, when considering “objective evidence” and “expected severity”. The scoring of 0 for not consistent, 1 for partially consistent and 2 for fully consistent follows the fundamental assumption in malingering research that the more inconsistencies a patient presents across multiple or relatively independent domains, the more likely it is that their performance reflects deliberate efforts to misrepresent their true capabilities ⁽⁴¹⁾.

Nevertheless, it does not provide clear indication whether it should be assessed separate or collectively, or to which degree. Therefore, is a single incident of self-reported ADL impairment which does not coincide with the objective evidence and expected severity equivalent in rating to reported ADL impairment and treatment non-compliance which does not coincide with the expected severity. A clear cut-off between Partially Consistent and Not Consistent is therefore not clear.

In addition, there appears to be an error in the description of criterion D which reads that the “inconsistencies or conflicts in the allegations...” should be reviewed in terms of its consistencies.

This should likely have read that the “...allegations, statements or medical evidence...” should be evaluated in terms of consistencies throughout the assessment.

4.2.5 Level of measurement

The level of measurement is considered ordinal in that it classifies in ranked fashion between Not Credible, Partially Credible and Fully Credible.

4.3 Results of Content Validity Measurement of the CAT

The content validity was measured against the malingering construct as defined in the extensive literature review in section 2.4 and compared with the Slick in terms of its utility for diagnostic ability. The results are indicated in table 4.1 and will be discussed in terms of the underlying concepts.

Table 4.1 Comparison between Content of Slick and CAT

	Adapted Slick et al. ⁽¹⁸⁾	CAT ⁽³²⁾
Credibility		
– Self reports	+	+
– Significant others	+	+
– Medical collateral	+	+
– ADL	/	+
Discrepancy methods		
– Internal consistency	+	/
– Disease deficit comparability	+	+
– Inconsistent with severity of injury	+	+
– Inconsistency between test scores and observed behaviours	+	/
– Violations of performance curves	+	/
Effort, motivation, compliance and incentive	+(incentive and effort)	+(compliance)
Exclude other causes	+	/

Key: + present – absent / implied

4.3.1 Concept of Credibility

The concept of credibility as it refers to obtaining various reports, including self-reports, significant others, collateral reports and documentation from the treatment team is comprehensively considered in the CAT. In addition, it also specifically refers to ADL performance, which is only implied indirectly in the Slick.

4.3.2 Discrepancy Methods

The categories of “disease deficit comparability” and “consistency with severity of injury” are clearly indicated in the CAT in terms of the scoring quantifiers of “objective evidence and/or expected outcome/severity”. These discrepancy methods therefore remain the main focus of content validity.

Of note, is that the impact of test performance on the overall rating is not clearly delineated, apart from the scoring quantifier which includes “expected outcome”. It is also assumed that these factors would be indicated in criterion D (“opinions about function given by other treating and examining sources”) and criterion E (“conflicts in the medical evidence in file”). Discrepancy methods which therefore refer to test performance, such as internal consistency, inconsistency between test scores and observed behaviours and violations of performance curves, are therefore rather implied in criteria D and E. In this regard, the differentiation between the two is also not clearly indicated. As an example, should inconsistent test behaviour on more than one report be considered in both criteria, and to which degree in terms of the scoring quantifiers.

The CAT therefore included discrepancy methods however did not necessarily allow for all specific types. This should take into account that it is a screening tool, whereas “violations of performance curves” are often more specialised techniques of malingering detection. More relevant, is that it encompasses theoretical concepts in keeping with malingering test methodology. In this regard, it allows for multitrait-multimethod strategies inclusive of self-reported measures, collateral sources, observations and record review. Multiple sources are also used when reviewing the consistency of allegations, statements and medical evidence (as indicated in criterion E).

4.3.3 Effort, Motivation, Compliance and Incentive

There is a difference between the approach of the Slick and the CAT in terms of the above. The Slick separately indicates “substantial external incentive” and refers to effort in terms of test performance. The CAT focuses predominantly on motivation and compliance with treatment and medication (criterion B and C), whereas the rest are implied. In this regard, the literature review indicated strong interrelationships between these factors.

4.3.4 Expert Review of Content Validity

The Performance APGAR was developed after various experts in the field of disability evaluation performed an extensive literature review. Unfortunately the method which was followed has not been published, apart from stating that preliminary studies validated the construct validity. The origin of the CAT is based on the SSA legislation which has been under litigious scrutiny and amended accordingly ⁽¹²⁰⁾. Therefore, the original construct as related to credibility considered expert review of content. However, further expert review is required in terms of the application to malingering.

4.4 Statistical Analysis of Construct and Concurrent Validity

The statistical analysis was performed on Statistica™ and interpreted with the assistance of the Centre for Statistical Consultation at the Stellenbosch University.

Descriptive statistics were used for descriptive data, such as the demographical information.

The Slick classification included Not Malingering, Possible Malingering and Probable Malingering and was therefore considered categorical variables. The CAT classification included Not Credible, Partially Credible and Fully Credible, but was linked to a score between 1 and 10, and was therefore categorical or ordinal data. Given this, nonparametric statistical procedures such as Chi-

square tests were used to establish statistical significance as the data could not be ordered numerically on a logical basis. It should be noted that chi-square does not refer to the strength of a relationship, only its significance. The sample size was considered sufficiently large to apply Chi-square significance tests.

This unfortunately limited statistical correlation with coefficient variables or factor analysis, which prevents comprehensive review of construct validity. Even if comparisons were made *post hoc*, by assigning groups in Probable Malingering versus Not Malingering, or attributing ordinal states to the category in terms of importance, it was considered statistically dubious.

Nevertheless systematic relations between variables were obtained through correspondence analysis, which is a descriptive and exploratory data analytic technique. Correspondence analysis shows how the variables are related, not just that a relationship exists. It was therefore able to show the relationships between the various criteria of the Slick and CAT.

The overall comparative classification between the Slick and CAT total scores were correlated with Spearman rank correlation.

The reader is reminded that reversed notation is used, i.e.: Not Credible (CAT) versus Not Malingering (Slick).

4.5 Demographics

Demographic information was obtained with regards to gender, age and diagnosis.

4.5.1 Gender

The distribution of male to female was 58:42 as indicated in figure 4.1. There is no known impact of gender on malingering. This is therefore purely descriptive statistics. Even though the study did not attempt to control for factors such as gender, this provides for a reasonable equal distribution in terms of gender.

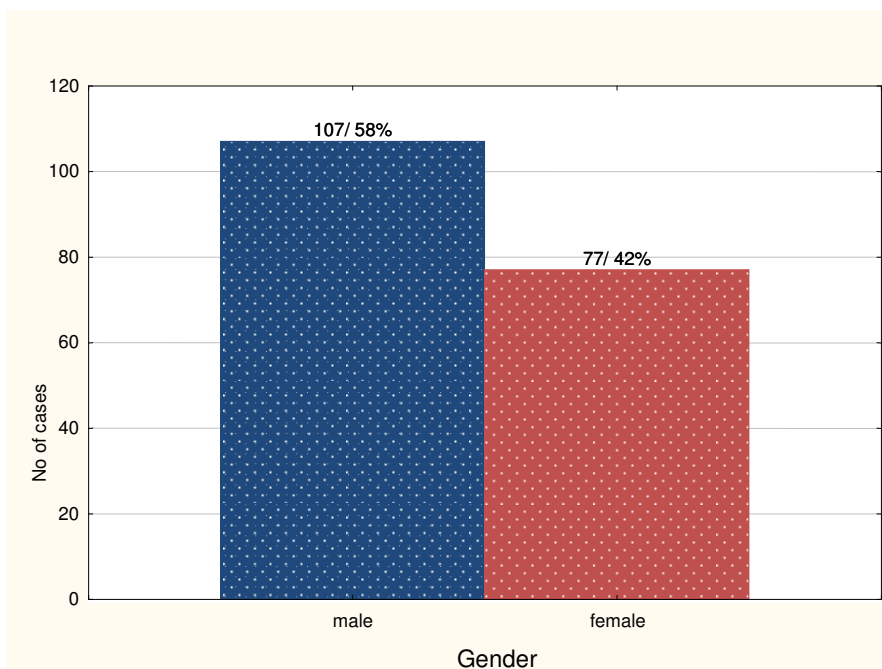


Figure 4.1 Gender distribution

4.5.2 Age

The age distribution is indicated in figure 4.2. The largest proportion of disability applicants (75%) were aged 41-60 years, which appeared feasible in terms of disability and early retirement applications. Comparative information from the insurance industry was however not available, given that data only exists per insurer or reinsurer group, but not collectively for the insurance industry as a whole.

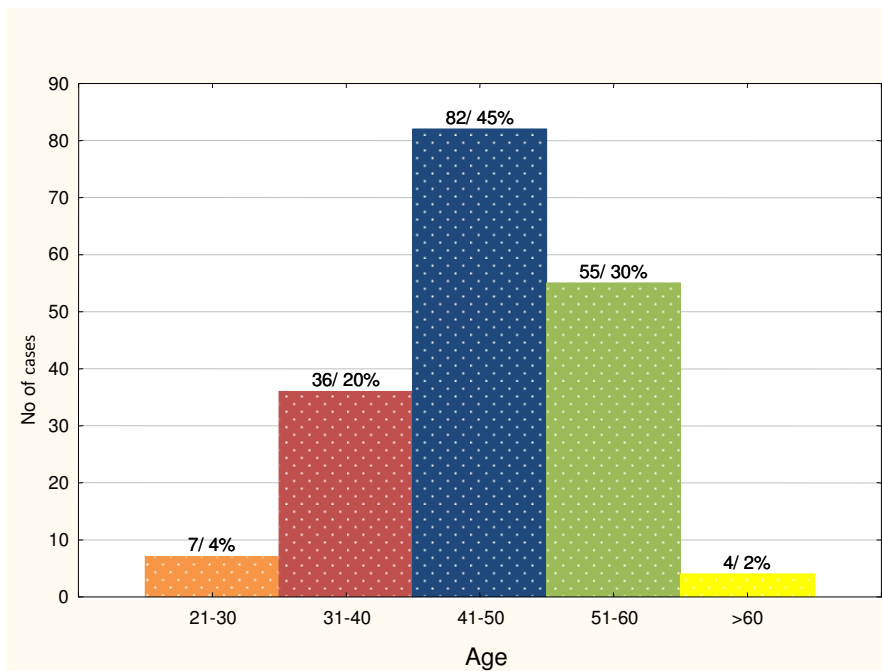


Figure 4.2 Age distribution

4.5.3 Diagnosis

The distribution of diagnoses is indicated in figure 4.3, with 57% of the sample suffering from pain, 30% depression and 14% fibromyalgia. This is considered representative of the population of disability claims in South Africa, where there is a high prevalence of musculoskeletal as well as psychiatric claims. Unfortunately comparative data only exists per insurance or reinsurance company, and not collectively for all insurance companies.

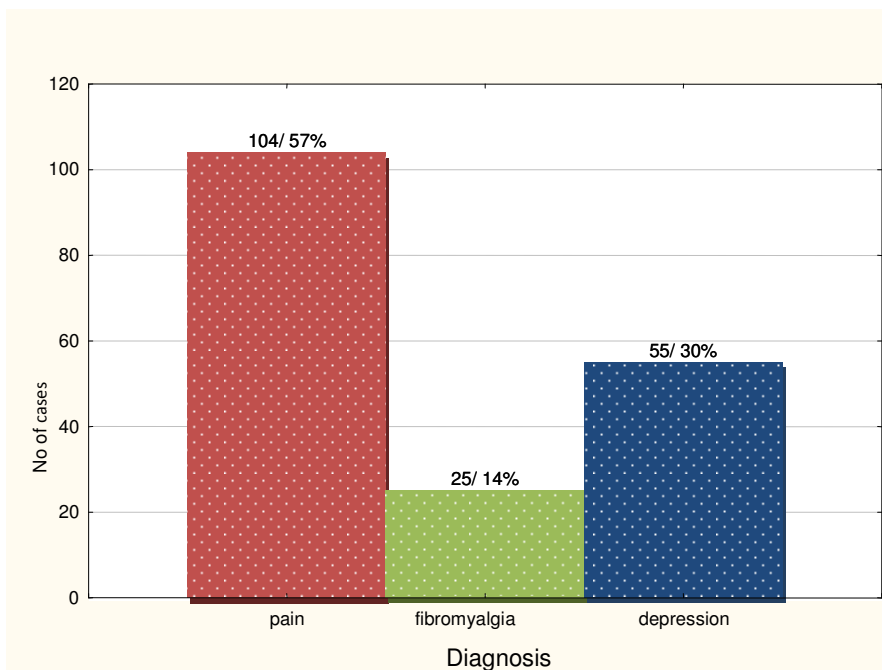


Figure 4.3 Overall diagnosis distribution

4.6 Prevalence of Malingering

The prevalence of malingering according to the Slick classification measured 20% as indicated in figure 4.4. Unfortunately limited information is available in South Africa ⁽¹²¹⁾, however the results were similar to the only other study of the prevalence of malingering in a medico-legal setting in South Africa where it rated 25% ⁽¹²²⁾. Comparatively, the prevalence of malingering is estimated at 30% in the USA ⁽⁶⁾ and 13% in Australia ⁽¹²³⁾ respectively. This study is therefore considered representative of the malingering phenomenon.

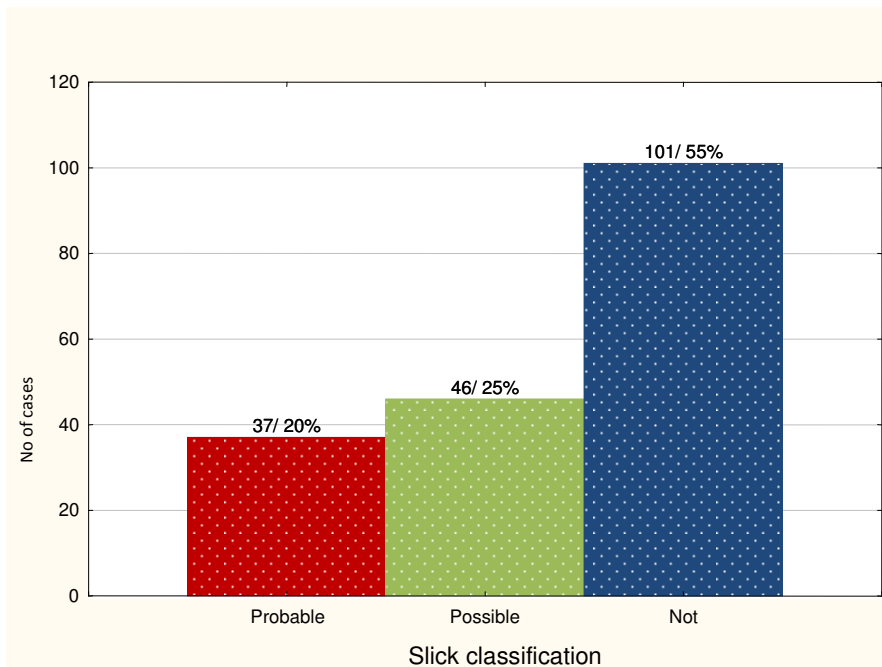


Figure 4.4 Prevalence of Malingering

The base rates of malingering, i.e. the proportion of a population that falls within a diagnostic category, are 18% for pain, 32% for fibromyalgia and 18% for depression (as indicated in figure 4.5). Analysis of the results of the pain and depression sub-groups displays an equivalent distribution pattern, however this was not replicated in the fibromyalgia sub-group. This is however not of statistical significance ($p > 0.05$), which is likely linked to the relative small amount of fibromyalgia cases ($n = 25$) considered.

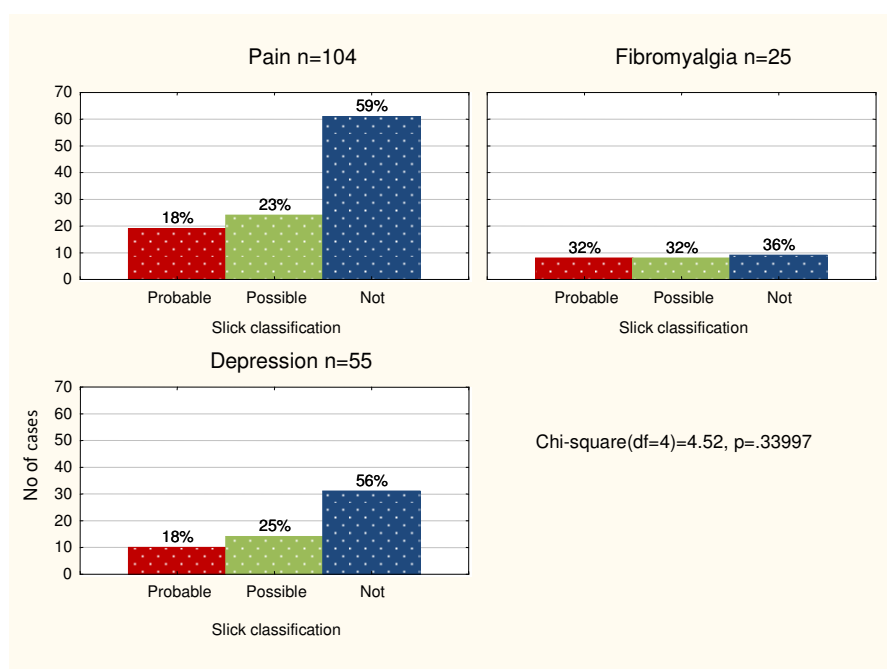


Figure 4.5 Base Rates of Malingering

4.7 Comparison between CAT and Slick Classification of Malingering

In general, there is a positive association ($p < 0.05$) between the Slick and CAT overall classification of malingering according to the total scores. As indicated in figure 4.6, there is an 86% association between the Not Malingering category (Slick) and Fully Credible category (CAT). Reversely, the Probable Malingering category (Slick) measured a 35% association with the Not Credible category (CAT) although a clear direct association is not indicated seeing that Partial Credible (CAT) rated higher at 59%.

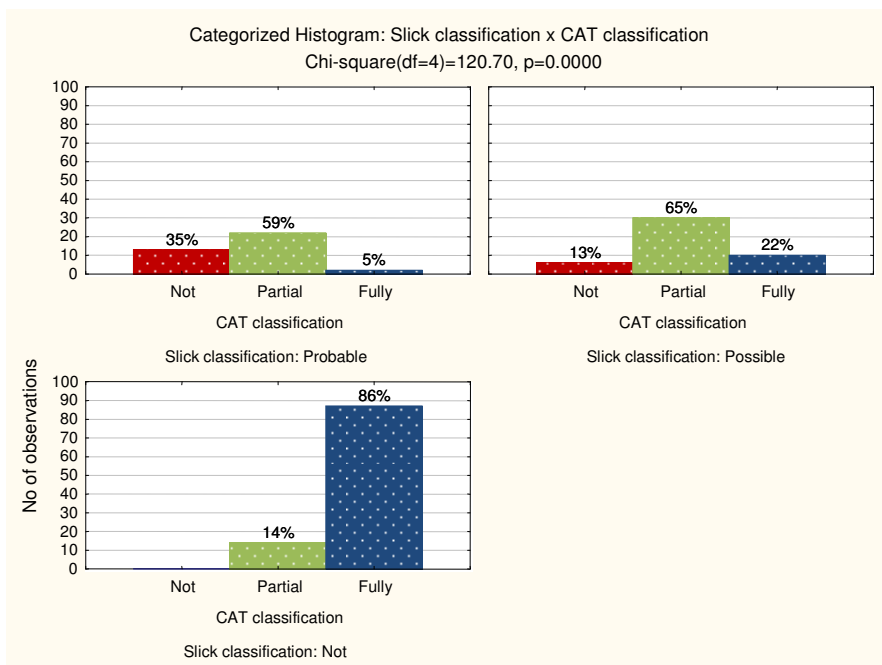


Figure 4.6 Comparison between CAT and Slick Classification of Malingering

4.7.1 Correlation per diagnosis

Using Spearman analysis, there is a positive correlation between the overall CAT and Slick classification of malingering (where $p < 0.01$) in all diagnostic subgroups, more so with pain and depression given the scores and group size (table 4.2). This therefore provides evidence of concurrent validity between the CAT and Slick.

Table 4.2 Correlation of CAT and Slick classification per diagnosis

Diagnosis	N	Spearman correlation
Total	184	0.74
Pain	104	0.75
Fibromyalgia	25	0.52
Depression	55	0.80

4.7.2 Analysis of CAT scoring

According to figure 4.6, the mean score was 7.17 (sd=2.3676). The median score was 8 with 25-75% of cases between 6 and 9. The median score of 8 is considered Fully Credible according to the scoring criteria. Given the results which indicated a high frequency of scores classified between Partially and Fully Credible¹, the ability to discriminate between these categories were further analysed.

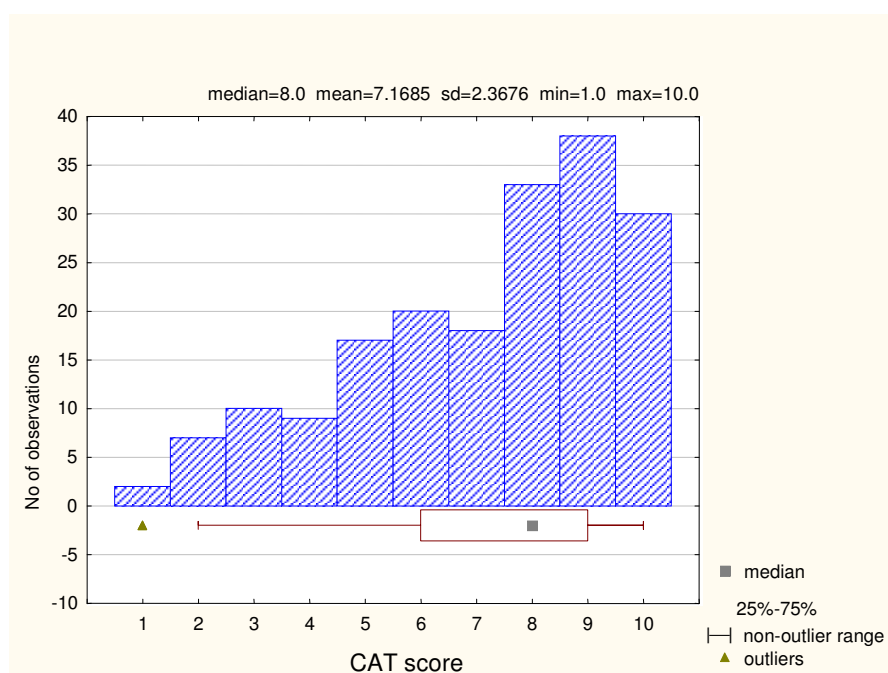


Figure 4.6 CAT overall score distribution

¹ Not credible = 0-3 / Partially credible = 4-7 / Fully credible = 8-10

Using Kruskal-Wallis (with $p < 0.01$) to compare the CAT score with the Slick classification as indicated in figure 4.7, it was found that there was an even distribution within usual CAT scoring parameters.

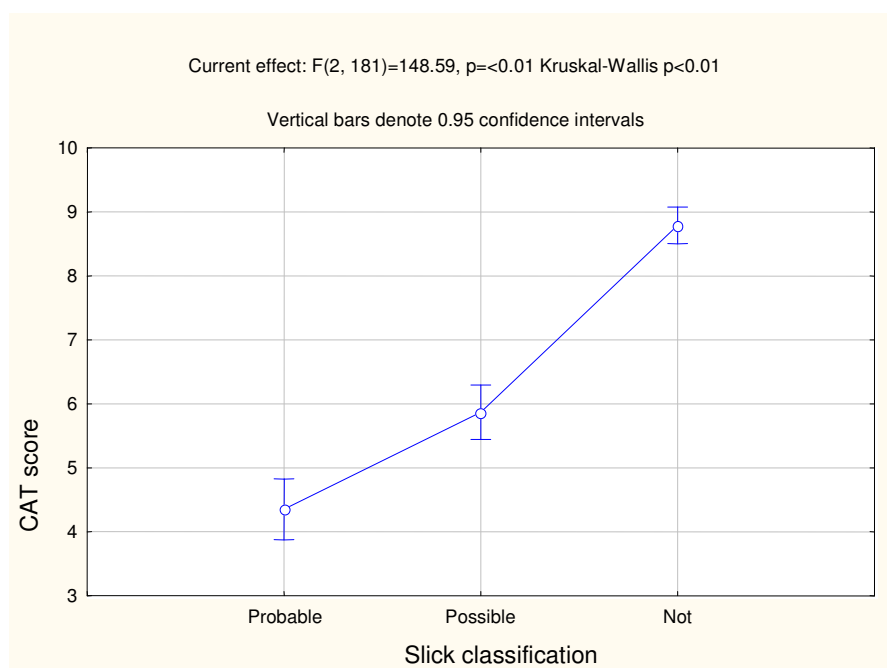


Figure 4.7 Comparison between CAT score and Slick classification

Of interest, is that the classification of Aronoff, Mandel, Genovese, Maitz, Dorto, Klimek and Staats⁽¹⁸⁾ includes the description of probable malingering as “noncredible” and possible malingering as “questionable”. The similarity of the terminology between the adapted Slick criteria and the CAT is thus highlighted, and therefore underscores the validity of the tool.

In addition, there was a clear difference between Not Malingering (CAT score: $m = 8.80$ SD 1.08) and Possible Malingering (CAT score: $m = 5.87$ SD 1.77) subcategories, which falls within the CAT scoring criteria of Fully Credible (8-10) and Partially Credible (4-7). However, there was not a clear cut-off score for Probable Malingering (CAT score: $m = 4.35$ SD 2.37), which showed an overlap with the Not Credible scoring criteria of 0-3. (See Table 4.3)

Table 4.3 Descriptive statistics of CAT score and Slick classification

Effect	Level of Factor	N	CAT score Mean	CAT score Std. Dev.	CAT score Std. Err.	CAT score -95%	CAT score +95%
Total		184	7.17	2.37	0.17	6.82	7.51
Slick classification	Probable	37	4.35	1.90	0.31	3.71	4.99
Slick classification	Possible	46	5.87	1.77	0.26	5.34	6.40
Slick classification	Not	101	8.79	1.08	0.10	8.58	9.01

Even though it could be argued that one could expect an additional category for a lower score when using Definite Malingering criteria (which was not used during this research), the researcher opined that this would negate the research objective of establishing a screening tool.

The discrimination Not Credible and Partially Credible would require further research to determine cut-off scores.

4.8 Analysis of Test Criteria on Overall Outcome

The relationship of the individual test criteria of the Slick and CAT were compared. The review of the Slick criteria analysed the association of quantitative and qualitative factors with the overall outcome. The review of the CAT focused on the association between test criteria and the overall score as compared with the Slick.

4.8.1 Association between Testing (Slick criterion B) and CAT

According to the Slick criterion B, evidence from testing is defined by:

- Probable response bias (RB)
- Discrepancy between test data and known patterns of functioning (TD ≠ KP)
- Discrepancy between test data and observed behaviour (TD ≠ OBS)
- Discrepancy between test data and reliable collateral reports (TD ≠ COLL)
- Discrepancy between test data and documented background history (TD ≠ HIST)

Based on the descriptors of the CAT categories, criterion E (Conflicts in the allegations, statement or medical evidence in the file) is the only category where a review of the combined inconsistencies or discrepancies is made. This is therefore the category where matters pertaining to test data must be interpreted.

Therefore, when comparing the association between Slick criteria B with CAT criterion E, the following results were found:

4.8.1.1 Probable Response Bias (RB)

Even though there was a positive association between the use of probable response bias as indicated in figure 4.8 where $p < 0.05$, the largest amount of observations were made of cases not using probable response bias. This is in keeping with the nature of the CAT as a screening tool, which can be used without and/or prior to referral for specialised malingering testing.

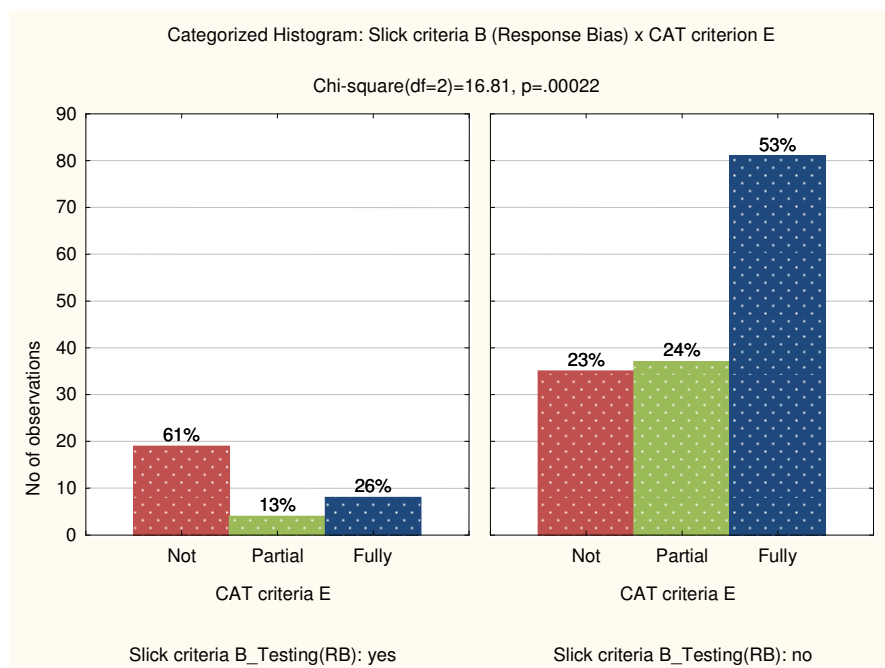


Figure 4.8 Association between Response Bias and CAT criterion E

4.8.1.2 Test Data and Known Patterns of Functioning (TD ≠ KP)

A positive association was between test data and known patterns as shown by the reversed categorised histogram (with $p < 0.05$) in figure 4.9. As indicated, 52% of discrepancies between test data and known patterns led to a Fully Credible CAT score, whereas 0% of consistency between test data and known patterns led to a Fully Credible score. This is in keeping with the design of the CAT where data is interpreted according to “expected outcome/severity”.

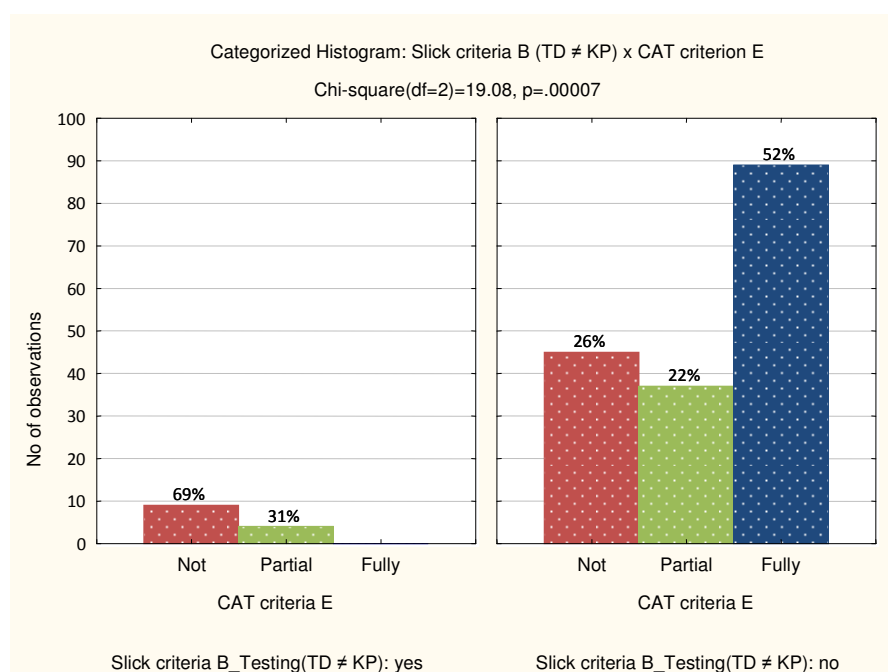


Figure 4.9 Association between Test Data and Known Patterns with CAT criterion E

4.8.1.3 Test Data and Observed Behaviour (TD ≠ OBS)

A clear positive association was shown between test data and observed behaviour as indicated in figure 4.10, where 89% of discrepancies between test data and observations led to a Not Credible CAT interpretation, whereas 56% of consistency between test data and observation led to a Fully Credible CAT interpretation.

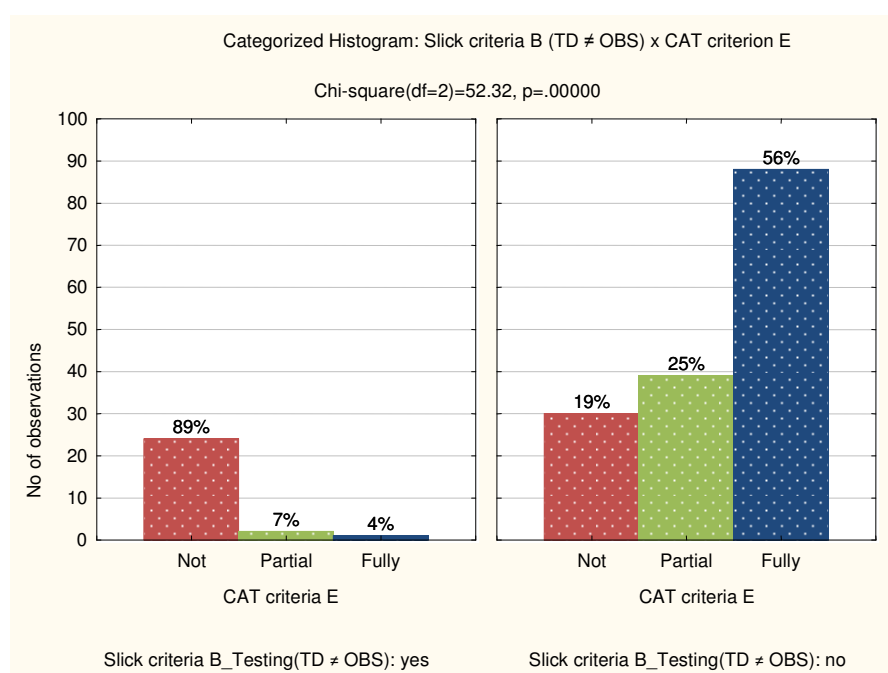


Figure 4.10 Association between Test Data and Observed Behaviour with CAT criterion E

4.8.1.4 Test Data and Reliable Collateral Reports (TD ≠ COLL)

There was insufficient information in these categories to obtain results of statistical significance as it amounted to 2% of the total scores on Slick criterion B (testing).

4.8.1.5 Test Data and Documented Background History (TD ≠ HIST)

There was insufficient information in these categories to obtain results of statistical significance as it amounted to 2% of the total scores on Slick criterion B (testing).

4.8.1.6 No Significant Test Data

There was also a positive association between data where test results were found to be non-contributory as indicated in figure 4.11. In this regard, it therefore showed that where formal test methodology did not feature as strongly, the use of other qualitative factors shows high association with the CAT.

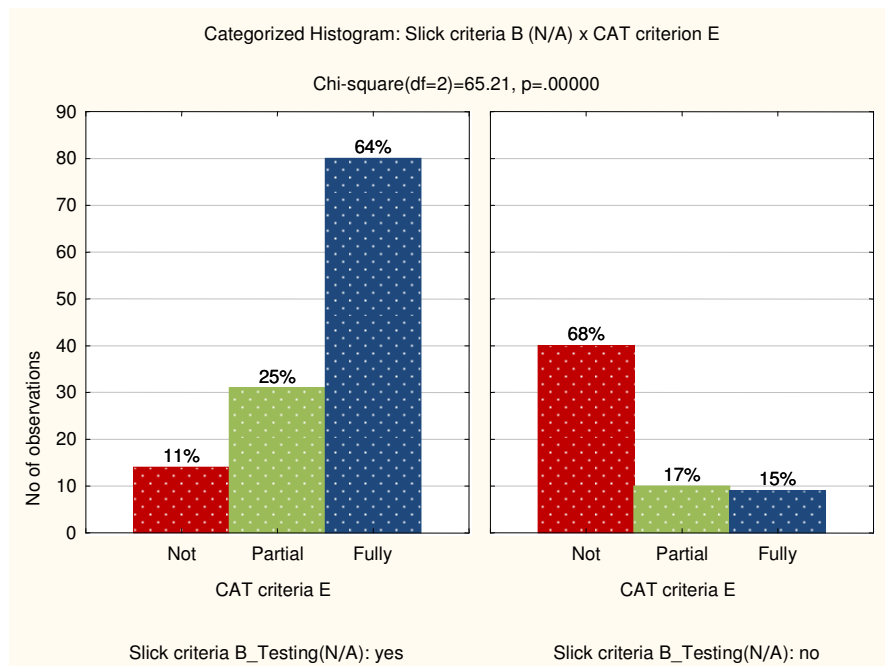


Figure 4.11 Association between No test data and CAT criterion E

4.8.1.7 Overall Association between Test Data and CAT criterion E (Conflicts in the allegations, statements or medical evidence in the file)

Given that a number of categorical variables were reviewed in the above section, a correspondence analysis (figure 4.12) was performed by cross-tabulating test data criteria on CAT criterion E. Even though this statistical technique reduces the accuracy of data as it standardises the row and column profiles, it is beneficial in providing an overview of the interrelationships when looking at the proximity of the variables. The joint display of row and column coordinates shows the relation between a point from one set and all points of another set, and not between individual points between each set.

The results show that most category values are close to Partially Credible CAT classification, specifically with regards to the association between the CAT score with test data and observations (TD≠OBS), history (TD≠HIST) and response bias (TD≠RB). No significant test results (NA) is nearer to Fully Credible, which highlights the importance of qualitative factors in the CAT. Discrepancy between test data and collateral (TD≠COLL) is closely linked with Not Credible. There was therefore a positive association of the test variables on the ability to identify Not Credible and Partially Credible, and qualitative factors (i.e. not formal test data) on Fully Credible. This therefore supports the concept of discrepancy methods as it relates to the malingering construct, which highlights a degree of construct validity of the CAT.

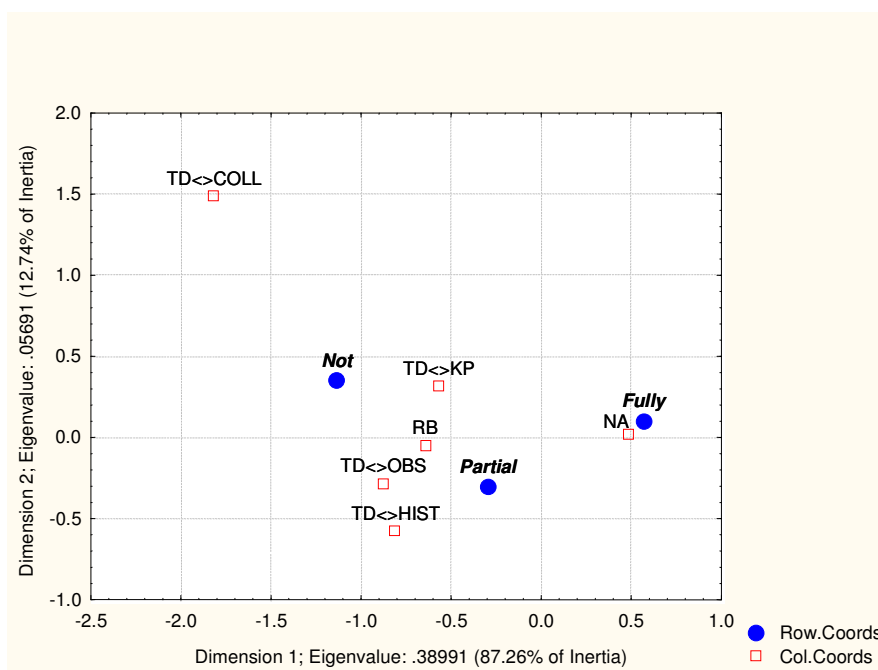


Figure 4.12 Correspondence Analysis of Test Data on CAT criterion E

4.8.2 Association between Self-report (Slick criterion C) and CAT

It has been argued that the Slick criterion C (Self-report) is not easily quantifiable and at best can achieve the result of Probable Malingering provided that at least two indicators are used. It was therefore suggested that two of these criteria should be positive to increase the likelihood of malingering⁽¹⁰¹⁾. As the classification of Probable Malingering was sufficient for the purpose of this study, the relationship of the self-reported qualitative indicators with the total score was investigated.

According to the Slick criterion C, evidence from self-report is defined by:

- Self-reported history is discrepant with documented history (SR ≠ HIST)
- Self-reported symptoms are discrepant with known patterns of functioning (SR ≠ KP)
- Self-reported symptoms are discrepant with behavioural observations (SR ≠ OBS)
- Self-reported symptoms are discrepant with information obtained from collateral informants (SR ≠ COLL)
- Evidence from exaggerated or fabricated psychological dysfunction (EXAG)

The overall CAT classification was used in this instance for comparison, as self-reported symptoms could relate to other CAT criteria (such as criterion A).

When comparing the association between Slick criteria C with the overall CAT classification, the following results were found:

4.8.2.1 Self-report and Exaggerated or Fabricated Psychological Dysfunction (EXAG)

There was not a clear association between self-reported exaggerated psychological dysfunction and the CAT classification as indicated in figure 4.13a. When there was no exaggeration, 55% was considered fully credible (with $p=.39358$).

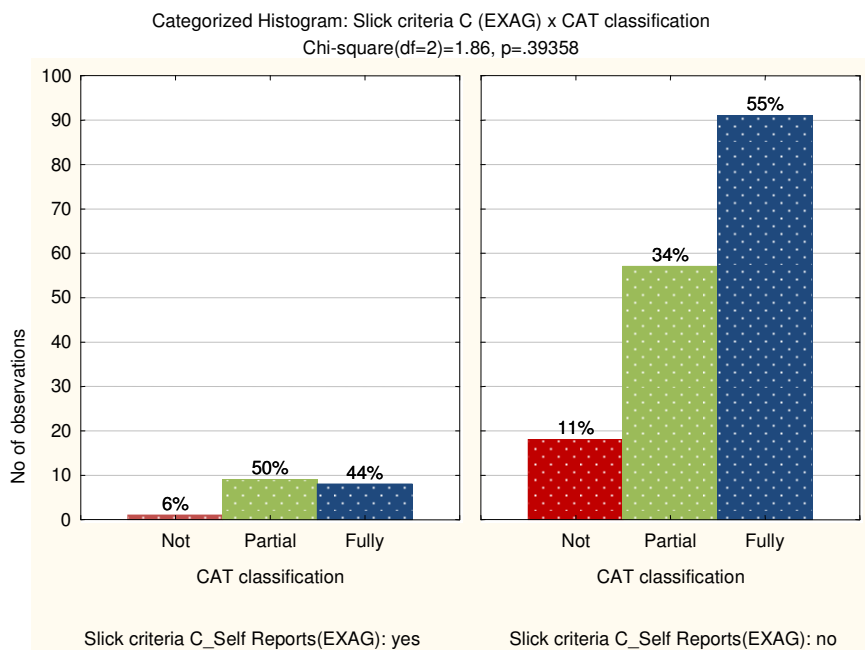


Figure 4.13a Association between Exaggeration and overall CAT classification

The same trend was also found when comparing self-reported exaggerated psychological dysfunction with the Slick classification (with $p=.61519$) as can be seen in figure 4.13b. It is postulated that exaggerated self-reports may be linked to illness behaviour which require further research and is likely not used in isolation.

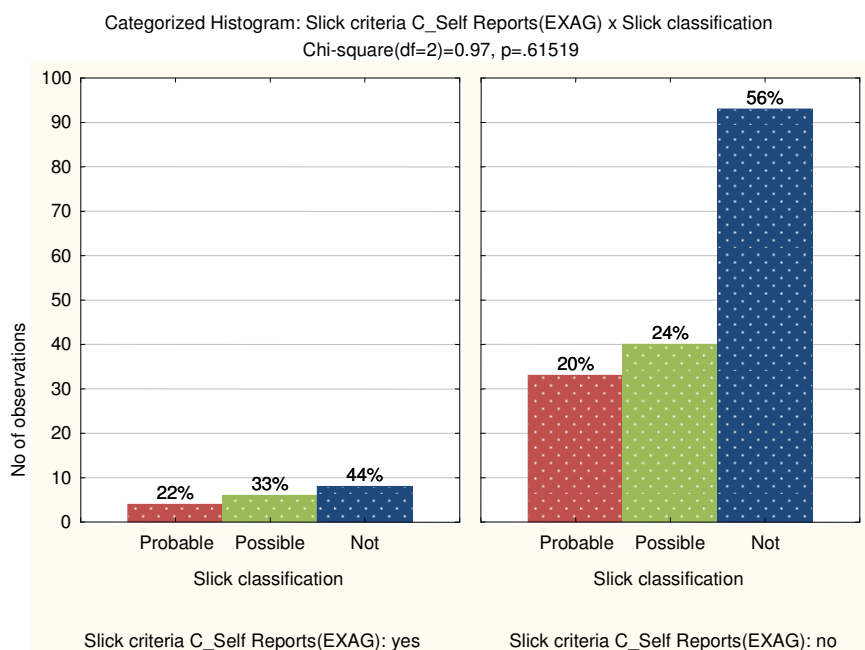


Figure 4.13b Association between Exaggeration and overall Slick classification

4.8.2.2 Self-report and Known Patterns of Functioning (SR ≠ KP)

A positive association was shown by absence of discrepancy between self-reports and known patterns, such as indicated by frequency of 71% Fully Credible cases (figure 4.14). A clear linear relationship was however not shown with the discrepancy with Partially Credible cases where the highest frequency was 65%. This could confirm the Slick criteria that self-report should not be used in isolation.

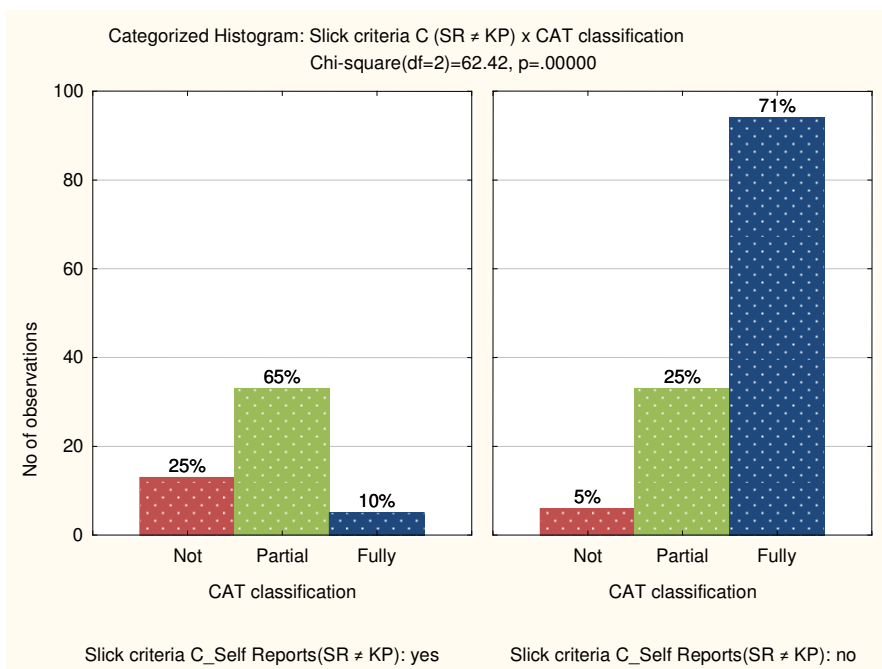


Figure 4.14 Association between Self Report and Known Patterns with CAT classification

4.8.2.3 Self-report and Observed Behaviour (SR ≠ OBS)

The same pattern was obtained with observed behaviour than with self-report and known patterns as can be seen in figure 4.15.

A clear linear relationship was again not shown in this instance, however the absence of discrepancy between self-reports and observation shows a high frequency of 78% of cases classified Fully Credible. Discrepancy between self-reports and observation shows the highest frequency of 58% of Partially Credible cases. The importance of multi-modal assessments are therefore again highlighted.

The above finding is considered significant given that the CAT criteria do not directly refer to observation, however this is often used in FCE and other assessments.

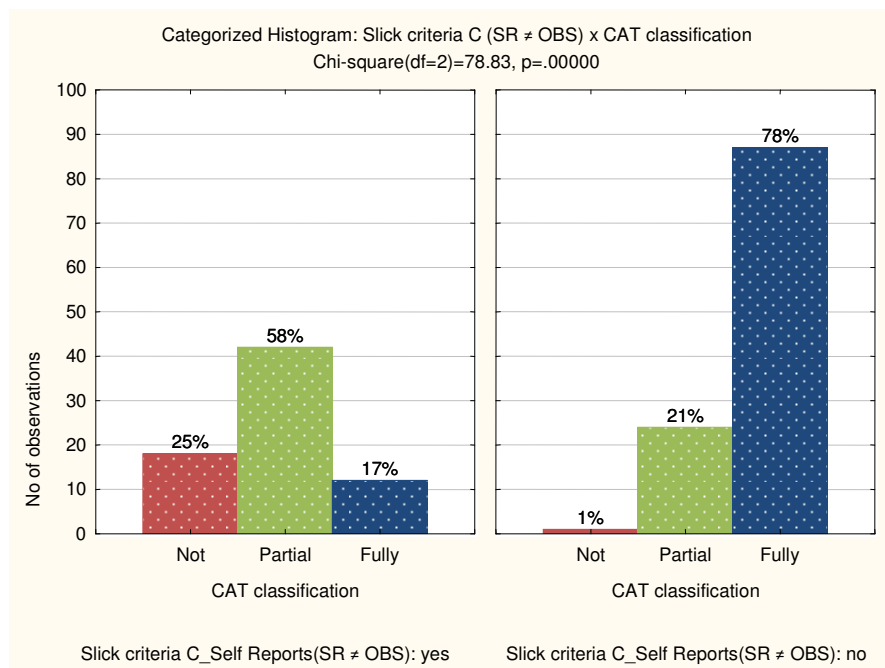


Figure 4.15 Association between Self Report and Observed Behaviour with CAT classification

4.8.2.4 Self-report and Collateral Information (SR ≠ COLL)

The same pattern was obtained than with self-report, known patterns and observed behaviour as can be seen in figure 4.16.

The absence of discrepancy between self-reports and observation shows a high frequency of 64% of cases classified Fully Credible. Discrepancy between self-reports and collateral shows the highest frequency of 55% of Partially Credible cases. The investigation of multiple factors is therefore again recommended.

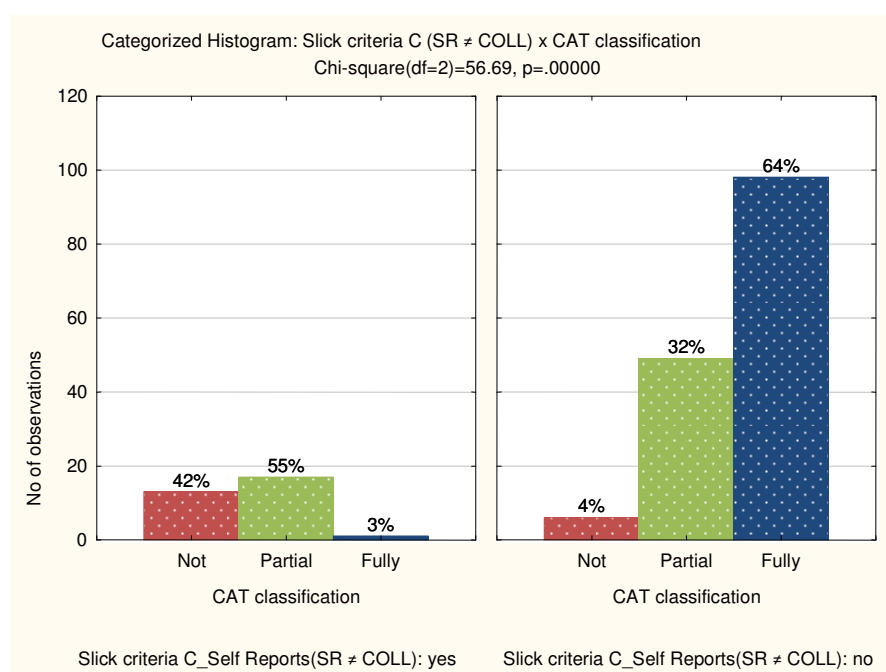


Figure 4.16 Association between Self Report and Collateral Information with CAT classification

4.8.2.5 Self-report and Documented History (SR ≠ HIST)

The same pattern was obtained than with self-report, known patterns, observations and collateral information as indicated in figure 4.17. Even though there was an association between the absence of discrepancy between self-reported and documented history with cases classified Fully Credible of 62%, the reverse is not true. The highest frequency again referred to partial credibility.

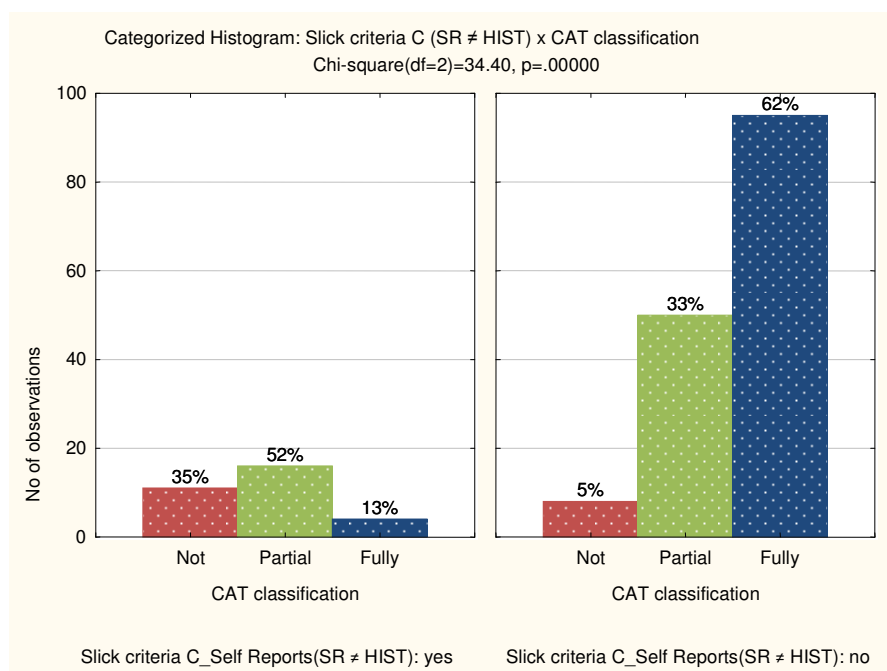


Figure 4.17 Association between Self Report and Documented History with CAT classification

4.8.2.6 Overall Association of Self Report and CAT classification

Even though there was a positive association between the absence of discrepancies between self report and the different variables, the relationship was not as clear as found with test data. Specifically the association with the high frequency of Partially Credible supports the research ⁽¹¹⁰⁾ that a discrepancy in criterion C (self report) alone is not sufficient for a diagnosis of malingering in the absence of a B criterion (test data).

Another correspondence analysis was performed which showed closer association between collateral and documented history with Not Credible, and known patterns and observations with Partially Credible is indicated in figure 4.18. This indicates a degree of construct validity when compared to credibility as an underlying concept of the malingering construct, specifically with regards to reports by collateral and documented sources.

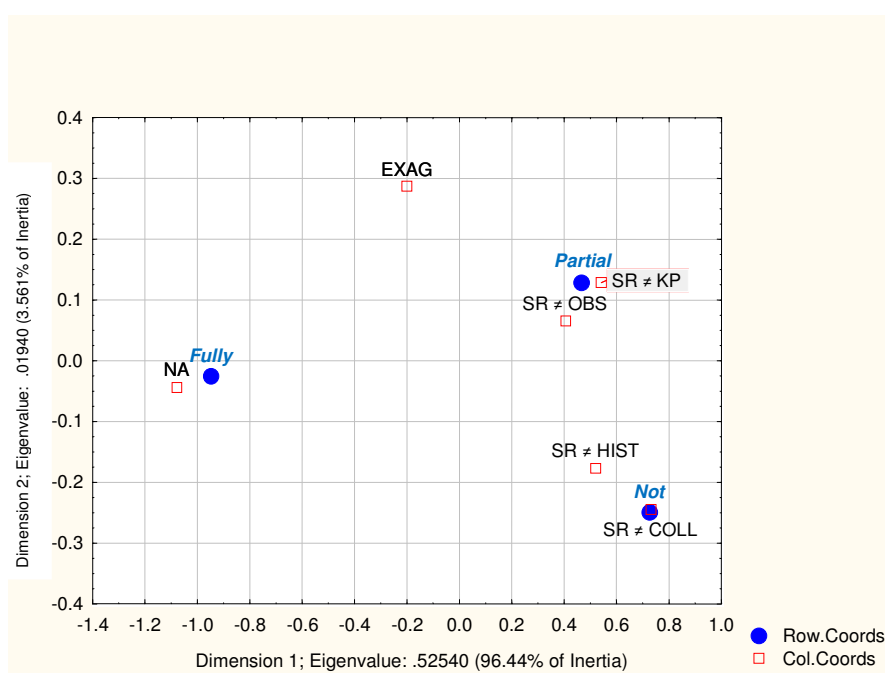


Figure 4.18 Correspondence Analysis of Self Report

4.8.3 Association between CAT criteria and Slick classification

Given that the association between criterion E and the Slick classification was already explored in section 4.8.1.7, the focus of this section is related to the other CAT criteria which were:

- Criterion A (ADL) – Impact of symptoms or condition on ADL
- Criterion B (Medication) – Type, dosage, effectiveness, and side effects of medication
- Criterion C (Treatment) – Treatments sought and received
- Criterion D (Collateral – med) – Opinions about function given by other treating and examining sources

4.8.3.1 Association between ADL and Slick classification

A positive association was shown by the reversed pattern on the categorised histogram where 78% of the Not Malingering group was considered Fully Credible in terms of ADL participation. The reverse was found in that only 5% of the Not Malingering group was considered Not Credible as illustrated in figure 4.19.

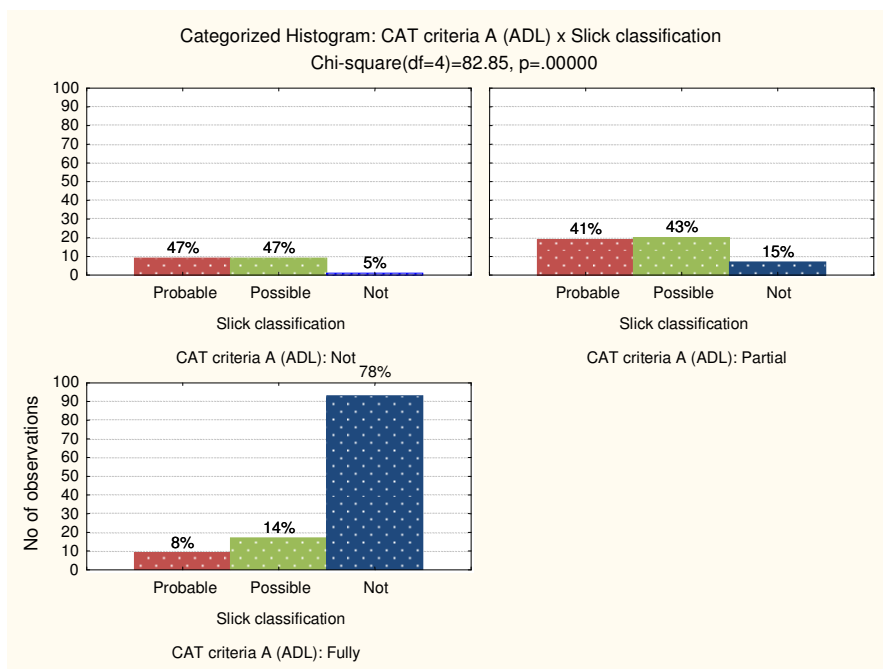


Figure 4.19 Association between ADL and Slick classification

4.8.3.2 Association between Medication and Slick classification

Even though there was an indication that 65% of those fully compliant with medication were not malingering (with $p < 0.05$), there was not a reversed or linear relationship with those not compliant with medication as indicated in figure 4.20. The overall distribution pattern in the partial compliance was not similar to the other CAT Partial Credible histograms in this section. The interpretation that this may have been influenced by factors such as availability of medication, financial concerns and personal preferences are not founded as scoring instruction indicated that these factors should not have influenced scoring outcome. Therefore an association between the use of medication and malingering classification of Possible and Probable Malingering did not exist.

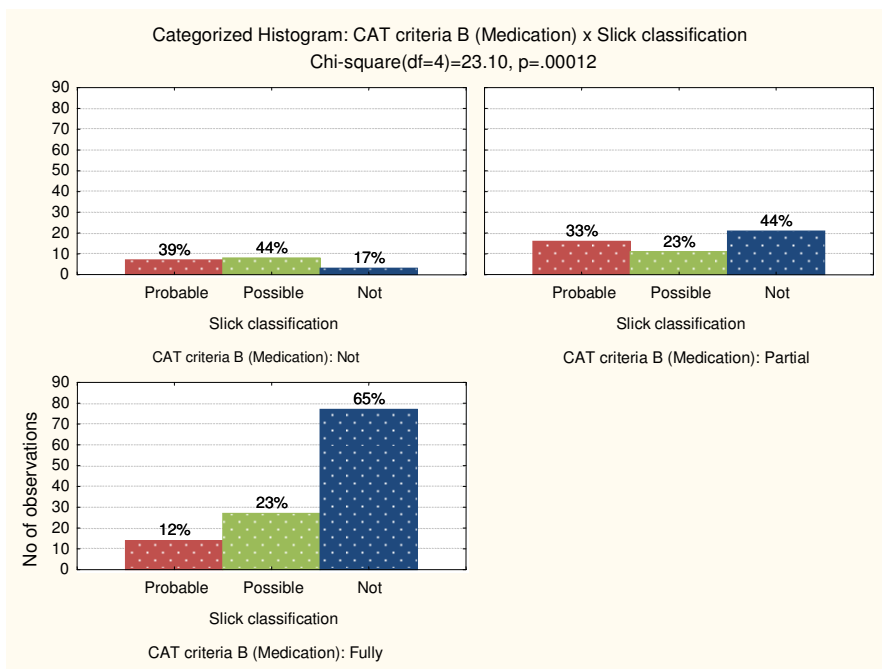


Figure 4.20 Association between Medication and Slick classification

4.8.3.3 Association between Treatment and Slick classification

A positive association (with $p < 0.05$) was shown between those fully compliant with treatment and rehabilitation (68%) and considered Not Malingering according to Slick criteria, whereas those not compliant (38%) were shown as Probable or Possible malingerers. (Figure 4.21)

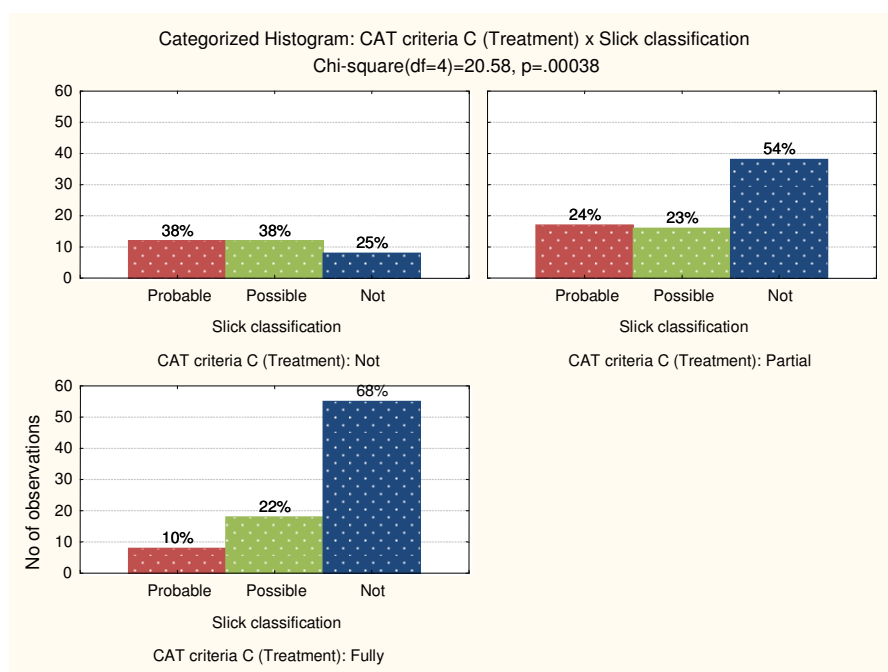


Figure 4.21 Association between Treatment and Slick classification

4.8.3.4 Association between Medical Collateral and Slick classification

A linear relationship was found with those considered not consistent with collateral medical opinion and Probable Malingering (67%), and Not Malingering (0%). The reversed is found with those Fully Consistent with collateral medical opinion and Probable Malingering (8%) and Not Malingering (71%). (See figure 4.22) Medical collateral information therefore seems to weigh heavier than collateral information from significant others (as indicated in section 4.8.2.4).

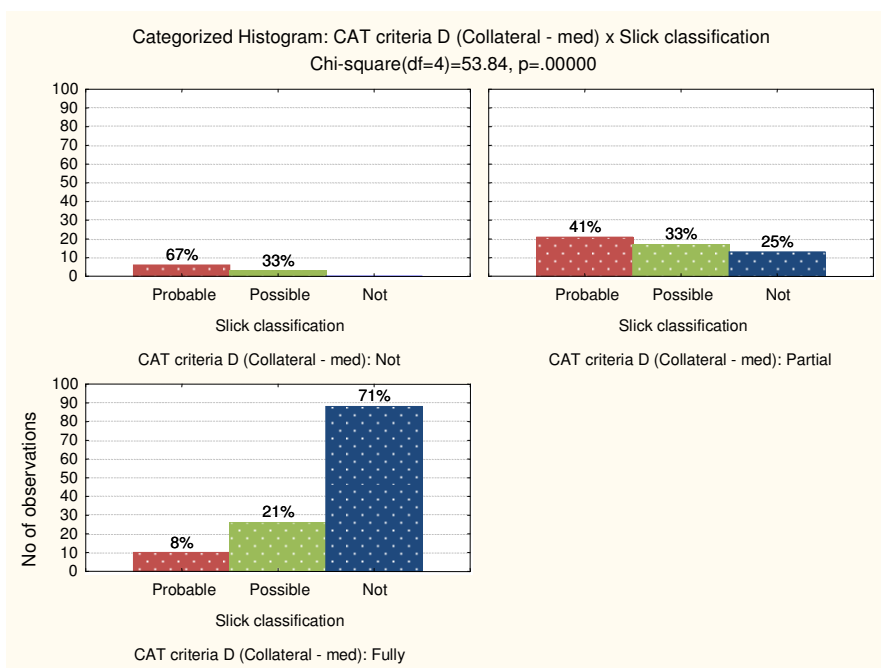


Figure 4.22 Association between Medical Collateral and Slick classification

4.9 Alternative Factors

The following factors, which were emergent from the literature review and data analysis, were also analysed in terms of the association thereof on malingering.

4.9.1 Financial Incentive

As indicated in section 3.8.1.1, the degree of financial incentive in terms of monthly benefit, lump sum benefit and litigious cases were also analysed. This would not have impacted on the Slick scoring classification, but allowed for additional comparison. In both the CAT and Slick, the strongest relationship with financial incentive was linked to litigation, followed by lump sum payments as indicated in table 4.4.

Table 4.4 Association between Financial Incentive with Slick and CAT classification

Marked cells have counts > 10. Chi-square(df=4)=31.55, p=.00000				
Slick criteria A (Incentive)	Slick classification Probable	Slick classification Possible	Slick classification Not	Row Totals
Monthly benefit	5	22	56	83
Row %	6.02%	26.51%	67.47%	
Lump sum	16	12	36	64
Row %	25.00%	18.75%	56.25%	
Litigation	16	12	9	37
Row %	43.24%	32.43%	24.32%	
Totals	37	46	101	184

Marked cells have counts > 10. Chi-square(df=4)=33.46, p=.00000				
Slick criteria A (Incentive)	CAT classification Not	CAT classification Partial	CAT classification Fully	Row Totals
Monthly benefit	5	23	55	83
Row %	6.02%	27.71%	66.27%	
Lump sum	6	19	39	64
Row %	9.38%	29.69%	60.94%	
Litigation	8	24	5	37
Row %	21.62%	64.86%	13.51%	
Totals	19	66	99	184

4.9.2 Injury

The impact of physical injury and psychological trauma on the claimant's sense of entitlement of compensation was considered as a possible alternative factor. However the results indicated no association (with $p > 0.05$) between injury and both the Slick classification (figure 4.23a) and CAT classification (figure 4.23b). A similar distribution of Not Malingering was found with little difference between injury or not in the Slick ($p = .64705$) and the CAT ($p = .90360$).

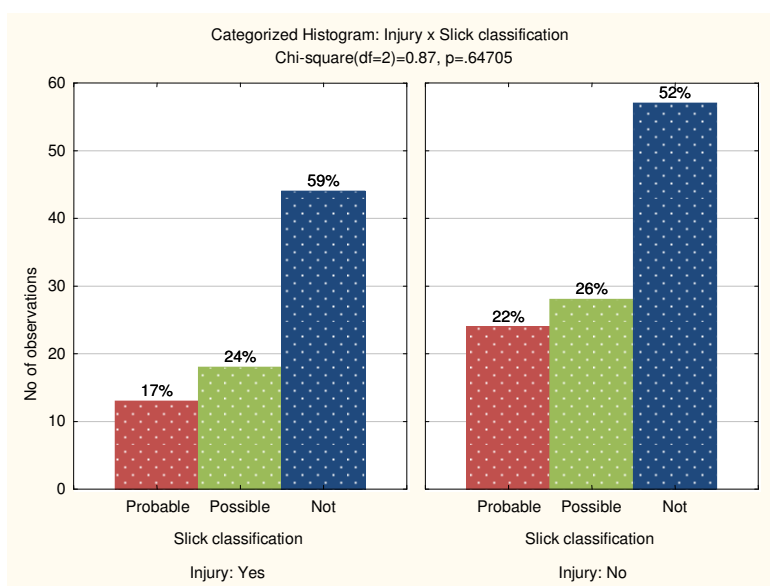


Figure 4.23a Association between Injury and Slick classification

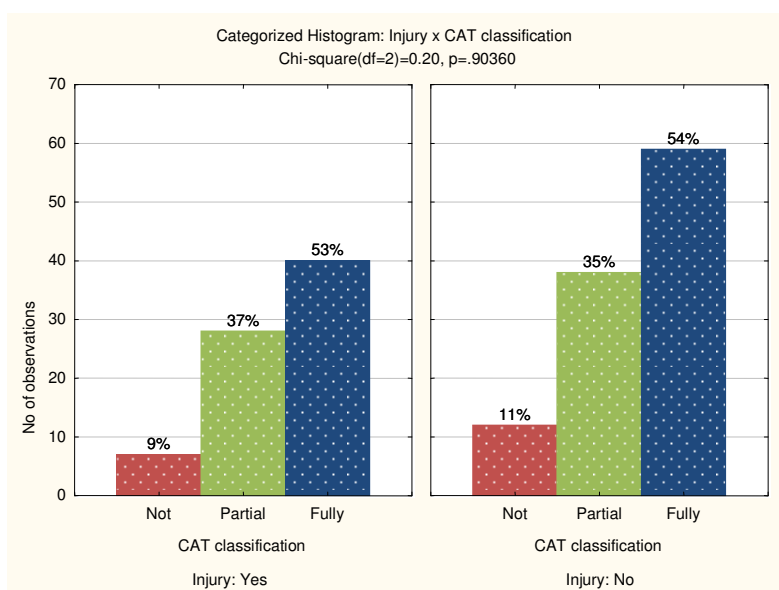


Figure 4.23b Association between Injury and CAT classification

4.10 Summary

Face validity was displayed by the purpose of the tool in that the term “credibility” is directly linked to the malingering construct. The item selection and level of measurement was adequate, however the scale and weighting was problematic. Content validity was supported by the malingering construct as operationalised and discussed according to table 4.1.

Demographical information was representative of the population of disability claims in South Africa and available information from literature.

The construct validity of the CAT was investigated with the analysis of interrelationships between the different variables of the CAT and Slick. In general, the results indicated positive association with most aspects of the CAT and Slick, thereby indicating a degree of construct validity. This was also supported with the associations obtained from correspondence analysis between CAT criterion D and Testing and Self-reports respectively, which is in keeping with the underlying construct of malingering in terms of testing methodology. The association between CAT criteria and Slick classification showed a positive relationship with ADL and collateral medical opinion. Even though there was a positive association between Not Malingering and compliance with medical and treatment, a reversed linear relationship with non-compliance was not found.

Concurrent validity was supported by positive correlation between the Slick and CAT, however requires further research. Scoring was comparable, but a clear cut-off score was not replicated with the Not Credible category which requires further research. Comparison of test data with the CAT score showed better association than with self-reports, which is consistent with the literature review findings.

There was also a positive association with financial incentive on malingering outcome, especially with regards to litigation, whereas the same was not found with the impact of injury.

The interpretation of the above results will be discussed in the following chapter, followed by recommendations for further research and practical application.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The aim of the research project was to perform a critical review of the validity of the CAT and its application to the screening of suspected malingering. Accepting the APA standards, this would initiate the review of construct validity especially as validity is a unitary concept that considers the appropriateness, meaningfulness and usefulness of the specific inferences made from tools. The objectives therefore included determining face and content validity through a literature review and concept analysis of the malingering construct. It also included a comparison between the CAT and adapted Slick criteria, which was identified as the most suitable criterion standard for diagnostic purposes. Data was analysed by subjecting the different variables of the Slick and CAT criteria, as well as emergent factors, to chi-square and correspondence analysis to determine construct validity. Concurrent validity was reviewed with coefficient correlations. The interpretation of the results will be discussed with recommendations for further research.

5.2 Face Validity of the CAT

The use of the term “credibility” within the title of the CAT is reminiscent of criminological theory as applied to the credibility of a witness, which is generally cautioned against in malingering literature especially when used by clinicians ⁽⁵⁹⁾. Even though it was argued by the authors of the CAT ⁽³²⁾ that it does not refer to a moral judgement and is based on original terminology referred to in the Social Security Administration guidelines, the researcher agrees that it would be better to consider it a Consistency Assessment Tool. This would allow for ease of use as the same acronym can be used. More noteworthy is that it refers to the conceptual importance of consistency and discrepancy methods it relates to malingering.

The CAT displayed adequate face validity in that:

- The purpose was clearly defined as a predictive instrument and screening tool for the overall credibility of persons claiming for financial compensation following injury or illness.
- The association between credibility and consistency criteria of malingering are shown in literature and indicated as such by the authors in terms of the reference to a “consistency assessment”.
- The item selection includes most relevant criteria of credibility.
- Scoring is judgemental and implicit. It underpins malingering concepts that the more inconsistencies presented over multiple or relatively independent domains, the more likely it is that performance reflects deliberate efforts of misrepresentation ⁽⁴¹⁾.

The following was of concern and requires further review:

- It did not provide adequate standardised instruction, despite the available explanation in the SSA guidelines. It is recommended that the criteria should include a detailed explanation of the key requirements, what is included or excluded and considered grounds for partially consistent ratings.
- The scaling requires further research in terms of weighting of consistency with “objective evidence” and “expected severity”, or alternatively should consider further instruction in terms of the difference between scoring quantifiers. Even though the scoring quantifiers (“with the objective evidence and/or expected outcome/severity”) and scoring system (0, 1, 2) follow fundamental assumptions in malingering research, it does not provide any specific weight in terms of the criteria. For example, it does not indicate whether a discrepancy between test results and diagnosis carried more or less weight than a discrepancy between a diagnosis and expected severity of impairment. It also does not indicate whether scoring should be performed numerically, for example, calculating the amount of discrepancies, or weighed qualitatively in terms of determining whether it is partially or fully consistent. Albeit that this remains a screening tool, the scoring system could possibly be adapted to include one point each for inconsistencies within available test results, and clinical substantiation, allowing the same total of ten points.
- The incorrect wording of criterion D.

5.3 Content Validity of the CAT

Content validity examines the extent to which the method of measurements includes all the major elements relevant to the construct being measured, and is usually obtained from the literature, representatives of the relevant populations, and content experts. In this regard, it was determined by literature review where a content analysis revealed adequate association between the concepts of credibility, discrepancy methods and effort/compliance.

Furthermore content validity was supported by the following underlying concepts in malingering detection:

- The scoring quantifiers refer to various levels of consistency “with the objective evidence and/or expected outcome/severity”. This therefore refers to discrepancy with objective test results (internal consistency), clinical presentation, severity and expected outcome of illness and injury (concurrent validity) and observed functioning and behaviour (ecological validity) which has been highlighted in most malingering models ^{(24) (91) (58) (39)}.
- Criterion D (“opinions about function given by other treating and examining sources in the file”) and criterion E (“inconsistencies or conflicts in the allegations, statements, or medical evidence in the file”) complies with the principle of using multiple sources of information ^{(15) (49) (95)}. Even though test results are implied in the term “medical evidence”, it does not provide specifically for specialised malingering testing or techniques such as test-retest over various domains, or comparison of self-report with collateral sources. Accepting that this is a screening tool, where specialised test results may or may not be available, this is however not considered vital.
- There is a strong focus on optimal treatment in criterion B (“type, dosage, effectiveness, and side effects of medications”) and criterion C (“treatment sought and received”). This is directly linked to insurance requirements that optimal treatment should be sought to ensure maximal medical improvement in order for permanence of a condition to be established. In addition, non-compliance is an indicator of poor motivation. Motivation, including external motivation through financial incentive, was proposed as one of the domains of malingering by Samuel and Mittenberg ⁽⁹⁹⁾. Motivation and effort is often also used simultaneously, as the level of effort produced is often affected by motivation ⁽²⁶⁾. The link between effort and compliance has been

modelled by Frederick ⁽⁶⁰⁾ when he included intent, which is often the key to the definition of malingering ⁽²⁾, in his model. The interplay between compliance, motivation, effort and intent as it relates to malingering is therefore clearly shown by literature. The delineation of motivation from compliance may require further exploration in terms of the impact of scoring on the CAT.

- In addition, it is noted that the CAT separates ADL as a category, whereas reported function in ADL is often implied in other malingering models. Criterion A (“impact of symptoms or condition on ADL”) specifically refers to the correlation of symptoms with Activities of Daily Living (ADL). Even though ADL is occasionally alluded to in other models of malingering, it is not usually a separate or distinct category. Interview information usually includes pre-injury ⁽⁶³⁾, clinical, socio-economic ⁽³⁷⁾, occupational and recreational history ^{(39) (99)}. It would appear that ADL has been covered under various self-reported measures, however this often refers to symptoms rather than function ⁽⁶⁵⁾. The combination of the above factors of symptoms and personal history provide similarities with ADL. This category is important to the South African disability insurance market for two reasons: Firstly, it is in keeping with policy guidelines that functional impairment relates to sustained and regular performance of ADL ⁽¹⁴⁾, which is especially relevant to insurance products designed on the grounds of functional impairment in ADL. Secondly, it also provides for opportunity to ensure valid information if language barriers exist as abstract symptoms can be related to difficulties in daily living tasks. Further exploration is required in terms of the possible use of ADL scales in terms of malingering and would require specific review in terms of convergent and divergent validity.

Content validity was therefore rated as adequate to excellent ⁽¹⁰⁴⁾, as the CAT is comprehensive and includes criteria suited to malingering, however the method was judgement based on literature rather than consensus method of content experts or statistical method.

Unfortunately the reported validity studies of the CAT ⁽³²⁾ was not published or available, which prevented empirical literature relevant to the construct, such as studies on construct validity of potential items or criteria, or related to other assessment instruments that have demonstrated validity. This therefore meant that content validity was limited to a literature review.

According to Haynes guidelines ⁽¹⁰⁵⁾, it is necessary that the concepts and constructs be delineated before judged by experts. In this instance, it was considered particularly relevant especially given that the tool needs to be generalised for multidisciplinary use. It is therefore recommended that the CAT is further reviewed by an expert panel that consists of multidisciplinary disability evaluation

professionals to judge the content, the extent to which the criteria fit the definition of the construct, the clarity and importance of the CAT. This should include numerical values reflecting the level of content-related validity to ensure statistical interpretation such as factor-analyses. This would then allow for more detailed content validity research as both logical analyses and experts' reviews would have been done.

5.4 Construct and Concurrent Validity of the CAT

The Slick criteria for MND ⁽⁵⁸⁾ have been used as a criterion standard for determining diagnostic categories of malingering. As such, it was considered a suitable measure for comparison with the CAT. It also fulfilled the concepts of the malingering construct as operationalised for this study. Given the need for multidisciplinary use, the adapted Slick criteria as proposed by Aronoff, Mandel, Genovese, Maitz, Dorto, Klimek and Staats ⁽¹⁸⁾ were therefore used to evaluate the construct and concurrent validity of the CAT.

The findings confirmed that:

- The prevalence measured 20% (as indicated in figure 4.4) which is consistent with other available malingering research performed in South Africa. This therefore highlights its applicability to the South African insurance industry and its relevance for further scientific research and test development.
- There was a positive association between the outcome of the CAT score (CAT classification as not, partially or fully credible) and Slick classification (as not, possible and probable malingerers).
- There was a positive correlation between the overall CAT classification of malingering when compared to the Slick classification in all diagnostic subgroups. In this regard, it is noted that the size of fibromyalgia group was relatively small (n=25) and would therefore require further research. The CAT therefore displayed concurrent validity when compared with the Slick criteria.
- The scoring of the CAT and Slick was comparable in terms of the levels between Not Credible and Probable Malingering, Partially Credible and Possible Malingering, and Fully Credible and

Not Malingering (figure 4.6). However there was not a clear cut-off score for the Not Credible and Probable Malingering category (figure 4.7). Further research would be necessary with regards to the sensitivity, specificity and predictive validity of the tool, especially with regards to cut-off scores.

- Comparison of test data with the CAT score (as indicated in figure 4.12) revealed positive association with specialised testing (TD≠RB), clinical substantiation (TD≠KP) and behavioural observations (TD≠OBS). Further review of collateral reports (TD≠COLL) and background history (TD≠HIST) is recommended as this study did not present with sufficient information for statistical analysis.
- Comparison of self-report with the CAT score (as indicated in figure 4.18) revealed poorer association with exaggerated psychological dysfunction, clinical substantiation, behavioural observations and collateral information. However, the pattern supports research of the Slick criteria ⁽⁵⁸⁾ that self-report should not be used in isolation, but in addition to quantitative test results. The impact of exaggerated self-reports as it relates to illness behaviour also requires further research.
- The use of ADL as one of the CAT criteria is supported by its positive association with the outcome of the Slick classification (figure 4.19).
- There was a positive relationship with compliance with medical and treatment and the outcome of Not Malingering (figure 4.20 and 4.21).
- Collateral medical opinion displayed a strong relationship with classification of malingering (figure 4.22), and provided with stronger statistical evidence than collateral information from significant others.
- There is a positive association with financial incentive on malingering outcome, with the strongest association linked to litigation and followed by lump sum payments (table 4.4).
- Injury did not display a positive association with malingering (figure 4.23).

The results therefore indicated that the CAT presented with adequate construct validity, as demonstrated by the confirmation by theoretical concepts. The strength of association between the overall CAT classification and the Slick classification is considered large considering that $r > 0.5$. However, this is not considered sufficient analysis to provide a conclusive opinion about concurrent

validity. In this regard, it should be noted that no published psychometric results are available to determine the validity of an equivalent screening tool.

Construct validity should be demonstrated from a number of perspectives and an accumulation of evidence. Thus, using the APA definition, considering whether the accumulated evidence supports the interpretation of the test scores of the CAT for the purpose of discriminating between various degrees of credibility, the findings of this research study was that it provided adequate validity in the different aspects of face, content and concurrent validity.

5.5 Limitations of the Study

It is acknowledged that there are a number of challenges in the field of malingering research, as explored in section 2.5 under Methodological Challenges. Practically this translates into ethical and definitional complexities in classifying malingerers, as well as the availability of information. Inevitably, analysis is often performed of cases referred for disability determination and/or malingering detection, which leads to sampling and researcher bias.

- During this research study, a saturated sampling method of convenience was used to ensure an adequate sample size. Even though this was limited to cases of the researcher, it did not necessarily limit generalisability given that it included national cases. In order to obtain sufficient cases, the inclusion criteria comprised of diagnoses associated with malingering. Even though this risks selection bias, it does not detract from the overall application of the tool. Nevertheless, it is recommended that sampling bias be minimised by the application of the tool at insurance companies where a large amount of data can be analysed irrespective of diagnosis.
- Retrospective data was used which limited further analytical research designs, such as performing a case-control study where malingers were compared to non-malingerers on the CAT.
- There may have been recall bias as the data analysis was not performed blind.

- This was further seen by the poor correlation with the peer check. However, it is strongly suggested that there may be poor inter-rater reliability given the limited instructions, training and available validity and reliability studies of the CAT.
- The unavailability of validity studies hampered the review of concurrent validity.
- Exploratory studies typically take place prior to highly focused studies, which limit stronger research designs with statistical interpretation of e.g. parametric tests.

5.6 Recommendations

Recommendations in terms of suggestions for further research, as well as practical applications for future use, include the following:

5.6.1 Recommendations for Further Research

Even though it is accepted that the CAT is a screening tool, there are a number of aspects that would benefit from further research to enable it to be a multi-modal screening tool, which can be used solely or in addition to available malingering detection tools.

1. Further review of content validity of the CAT

It is recommended that the CAT is further reviewed by a multidisciplinary expert panel to judge the content, the extent to which the criteria fit the definition of the construct, the clarity and importance of the CAT.

2. Review of criterion and predictive validity of the CAT

The development of tools used for the detection of suspected malingering requires a rigorous scientific approach given the ethical and legal implications thereof. Greve and Bianchini ⁽⁹⁶⁾ recommended that focus should be on indices of classification accuracy, such as sensitivity and specificity, as well as the predictive power, which is an index of the confidence one has that the classification of an individual is correct. Given the impact of falsely diagnosing someone as a malingerer, it is suggested that focus remains on specificity rather than sensitivity.

It is therefore crucial that a follow-up case-control study is performed where results are compared between a malingerer and non-malingerer group. Malingerer groups can be ascertained by using the Slick criteria. The non-malingerer group should include both healthy and those free of financial incentive.

It is recommended that the sample size at least exceed 20 controls to ensure adequate review of cut-off scores and allow for adequate variance and diversity of conditions to render it useful for multiple diagnoses. In this regard, it is imperative that adequate controls such as an appropriate sample size and standardised tests be used for diagnoses, such as fibromyalgia, or cases where illness behaviour are suspected.

It is recommended that further research into the generalising of the results to other diagnoses be conducted, specifically with those conditions with neurocognitive symptoms where literature also shows a high prevalence of malingering ⁽⁶⁾.

3. Review of the reliability of the CAT

There are a number of factors that questions the inter-rater reliability of the CAT. This includes the lack of standardised instructions as well as information regarding the scoring quantifiers (i.e. what is considered fully, partial and not consistent). Further research would require peer review of the CAT by more than one professional discipline, standardisation and validation of instructions and further development of the scoring quantifiers and criteria.

Should the use of the CAT be considered for the use of claim consultants at insurance companies, the review of the tool in terms of inter-rater reliability, practicality and utility should be performed. This may also provide for opportunity for research as claims consultants are generally multidisciplinary staff who would be able to apply the CAT to available prospective claim data.

Should the use of the CAT be considered for various multidisciplinary professionals involved in disability determination, it is recommended that it can be used in conjunction with other assessment protocols given the short administration time. It is noted that most international assessment protocols do not necessarily include malingering detection tools^{(19) (20) (31)}, as this is often considered specialised testing of long duration. It may also be considered for medical professionals who do not have specialised training or experience in disability determination⁽¹²⁴⁾.

4. Specialised ADL indices for Malingering

Another research question indirectly related to the results of this study, is whether ADL indices could be adapted or developed to include consistency criteria, which can be used for malingering screening. This would then also further substantiate criterion A of the CAT. This may be in the form of self-reported checklists of functional impairments or existing disability questionnaires, where cut-off points are developed and reviewed for specific medical conditions. The assessment of ADL has predominantly been in the domain of occupational therapy in terms of activity analysis and assistive devices^{(125) (126)}.

It would be worthwhile if specialised ADL indices can be developed especially given the nature of insurance disability products, which does not only focus on occupational disability, but also functional impairment as it relates to ADL. Given that the claimant is aware of the description of each ADL sub-item in order to qualify for benefits, the traditional use of interview is no longer adequate with suspected malingerers. Performance on Functional Capacity Evaluations (FCE) do also not necessarily compare with ADL impairments⁽¹²⁷⁾.

5. Future Statistical Modelling of the CAT

It is recommended that further 3D statistical modelling of the CAT is performed to allow for various scoring criteria and grading thereof. This would allow for the two current scoring quantifiers to be expanded to include self-reported and collateral analysis, objective test results inclusive of intra-test validity, and clinical substantiation based on established diagnostic cut-off scores.

5.6.2 Practical Applications for Further Use

The CAT has two advantages in terms of use: Firstly, it provides opportunity for application with many professionals involved with disability determination, including claims consultants, and is not specific to a specific discipline. It could therefore be included in guidelines for disability assessments, and the rating can be added to usual reports. Of note, is that there are no specific guidelines regarding the assessment of malingering for the insurance industry in South Africa. The application of the CAT will therefore be beneficial in this regard. In addition, it can easily be programmed with available claims assessment software to aid decision-making and provide a wealth of data for further research.

Secondly, it provides a standardised guideline for observations that are often considered qualitative in nature. In so doing, it provides a background for a multimodal assessment, independent of whether specialised malingering testing is performed or required. It is also not time or context specific, therefore allowing for application at various intervals.

It is recommended that the following changes (as indicated in table 5.1) are made to the existing CAT based on the results of this study, which supports the delineation of correlation with test results (internal validity) and clinical substantiation (concurrent validity).

Table 5.1 Proposed Consistency Assessment ToolInstructions for Use:

- The Consistency Assessment Tool requires the indication of five criteria, either based on record review or actual assessment.
- Agreement with the detail should be indicated with a mark that denotes one point, to a maximum of two points per criterion. The total score is calculated and compared with the scoring guidelines below.
- A score of between 0-3 as indicated in the red zone warrants immediate referral for additional testing and/or forensic investigation.

Criteria	Detail	Score	Score Total
Activities of Daily Living	Consistency with objective evidence	(___/1)	(___/2)
	Consistency with expected outcome/severity	(___/1)	
Compliance with medication, treatment and rehabilitation	Consistency with treating or rehabilitation team reports	(___/1)	(___/2)
	Consistency with expected outcome/severity	(___/1)	
Motivation	Increased financial incentive (e.g. potential termination of benefit, lump sum, accelerated benefit based on medical deterioration)	(___/1)	(___/2)
	Involvement with litigation	(___/1)	
Allegations, statements and medical opinion of function by treating and examining sources	Consistency with collateral information from significant others	(___/1)	(___/2)
	Consistency with collateral medical reports	(___/1)	
Conflicts in the medical evidence or test results	Consistency with objective evidence	(___/1)	(___/2)
	Consistency with expected outcome/severity	(___/1)	
	Total consistency score		(___/10)
	Not consistent	0-3	
	Partially consistent	4-7	
	Fully consistent	8-10	

(Adapted from *Credibility Assessment Tool*, by Colledge et al., 2001)

5.7 Conclusion

This research study aimed to perform a critical review of the validity of the CAT, a tool which was initially developed as part of the Performance APGAR for the purpose of disability determination, and its application to malingering. It was undertaken given the lack of standardised evaluation protocols across all specialities involved with disability determination, specifically with regards to malingering which is of relevance to the insurance industry.

Test validation has been described by the APA⁽¹⁰³⁾ as a process of accumulating evidence to support inferences made from instrument scores, which in this instance included the critical review of the validity of the CAT to be used as a screening tool for malingering detection. The methodology therefore included an extensive literature review, concept analysis and comparison with the most suitable criterion standard. However, it is cautioned that the study was exploratory in nature and as such further studies are recommended.

The results of the study found that the CAT presented with adequate face validity in terms of the purpose, item selection and association between consistency criteria of malingering as shown in literature, however required standardised instruction and improved weighting of the scale.

Content validity was rated as adequate to excellent, as the CAT is comprehensive and includes criteria suited to malingering, however the method was judgement based on literature rather than consensus method of content experts or statistical method.

Construct validity was adequate as demonstrated by the confirmation by theoretical concepts, and interrelationships between the CAT and Slick is indicated in section 5.4. The strength of association between the overall CAT classification and the Slick classification is considered large considering that $r > 0.5$. However, this is not considered sufficient analysis to provide a conclusive opinion about concurrent validity. In this regard, it should be noted that no published psychometric results are available to determine the validity of an equivalent screening tool.

Recommendations for further research included the review of the content validity with subject experts, as well as criterion and predictive validity by a follow-up case-control study between a malingering and non-malingering group. Research into the reliability of the CAT was also recommended, especially if it is to be utilised by multidisciplinary professionals involved in disability determination. Other research suggestions included specialised ADL indices for malingering detection, which is especially relevant to occupational therapists given their scope of practice. Future 3D statistical modelling of CAT scoring quantifiers is recommended to allow for scoring between multiple variables. Finally the proposed Consistency Assessment Tool was depicted, with alterations based on the results of this research project.

REFERENCE LIST

REFERENCE LIST

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: APA; 1994.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Revised 4th ed. Washington, DC: APA; 2000.
3. Dempsey, P. Life industry foils fraudulent and dishonest claims worth millions. [Online]. 2009 [cited 2009, July 7]. Available: <http://www.asisa.org.za/downloads/051-2009.pdf>.
4. Van der Werf SW, Prins JB, Jongen PJ, Van der Meer JW, Bleijenberg G. Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment: a matched comparison between chronic fatigue syndrome and multiple sclerosis. *Neuropsychiatry Neuropsychol Behav Neurol* 2000;13:199-203.
5. Gervais RO, Green P, Allen LM, Ferrari R, Pieschl SD. Effort testing in patients with fibromyalgia and disability incentives. *J Rheumatol* 2001;28:1892-1899.
6. Mittenberg W, Patton C, Canyock EM, Condit D. Baserates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol* 2002;24(8):1094-1102.
7. Shevel A. Work disability payouts in SA are all in the head. [Online] 2007 [cited 2007, July 22]. Available: <http://www.timeslive.co.za/sundaytimes/article110488.ece>
8. Buys T, Van Biljon H. Functional capacity evaluation: An essential component of South African occupational therapy work practice services. *Work* 2007;29(1):31-36.

9. Holmes EB. Impairment Rating and Disability Determination. [Online] 2008 [cited 2008, May 14]; Available <http://emedicine.medscape.com/article/314195-overview>.
10. South African Social Security Agency. [Online] 2010 [cited 2010, November 6]. Available: <http://www.sassa.gov.za/ABOUT-SOCIAL-GRANTS/QUALIFYING-REQUIREMENTS-651.aspx>.
11. SA Department of Labour. [Online] 1993 [cited 2010, November 6]. Available: <http://www.labour.gov.za/legislation/acts/compensation-for-occupational-injuries-and-diseases/compensation-for-occupational-injuries-and-diseases-act>.
12. South African Society of Psychiatrists. Guidelines to the Management of Disability Claims on Psychiatric Grounds (Second Edition). [Online] [Cited 2010, November 6]. Available: http://www.sasop.co.za/A_aboutus_Guidelines.asp.
13. Road Accident Fund Act (No. 56 of 1996). [Online] 1996 [cited 2010, November 6]. Available: <http://www.raf.co.za/Legislation/Pages/RAFAct.aspx>.
14. Coetzer P. Disability assessment of lower back ache [Insert]. S Afr Med J 1997:1-15.
15. Occupational Therapists in Life Assurance (OTLA). Guidelines for Occupational Therapy – Evaluations and Reports. [n.p.];[n.d.]
16. Association for Savings and Investment SA. Quantitative disability evaluation of syndromes presenting with chronic fatigue. [Online] [Cited: 2010, November 6]. Available: <http://www.asisa.co.za/index.php/medical-info.html>.
17. American Academy of Disability Evaluation Physicians (AADEP) [Online] [access 2010, July 25]; Available: http://www.aadep.org/en/about_aadep/.

18. Aronoff GM, Mandel S, Genovese E, Maitz EA, Dorto AJ, Klimek EH, Staats TE. Evaluating malingering in contested injury or illness. *Pain Practice* 2000;7(2):178-204.
 19. Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, Van der Klink JJL. Reliability and validity of the Disability Assessment Structured Interview (DASI): A tool for assessing functional limitations in claimants. *J Occup Rehabil* 2010; 20:33-40.
 20. De Boer WEL, Donceel P, Brage S, Rus M, Willems JHBM. Medico-legal reasoning in disability assessment: A focus group and validation study. *BMC Public Health* 2008; 8:335-344.
 21. Nahm FS, Lee PB, Kim TH, Kim YC, Lee CJ. Comparative analysis of the independent medical examination reports and legal decisions in Pain Medicine. *Korean J Pain* 2010; 23(1):34-38.
 22. Byrne, LJ. The current and future role of occupational therapists in the South African group life insurance industry. [unpublished dissertation]. Pretoria: University of Pretoria; 2001.
 23. King PM, Tuckwell TE, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998;78(8):852-865.
 24. Genovese E, Galper JS, editors. *Guide to the Evaluation of Functional Ability: How to Request, Interpret and Apply Functional Capacity Evaluations*. USA: AMA Press; 2009.
 25. Velozo, CA. Work evaluations: critique of the state of the art of functional assessment of work. *Am J Occup Ther* 1993;47(3):203-209.
-

26. Strong S, Baptiste S, Cole D, Clarke J, Costa M, Shannon H, et al. Functional assessment of injured workers: A profile of assessor practices. *Can J Occup Ther* 2004;71(1):13-23
 27. Innes E, Straker L. A clinician's guide to work-related assessments: 1 - purposes and problems. *Work* 1998;11:183-189.
 28. Innes E, Straker L. A clinician's guide to work-related assessments: 2 - design problems. *Work* 1998;11:191-206.
 29. Allen S, Rainwater A, Newbold A, Deacon N, Slatter K. Functional capacity evaluation reports for clients with personal injury claims: A content analysis. *Occup Ther Int* 2004;11(2):82-95
 30. Strong S, Baptiste S, Clarke J, Cole D, Costa M. Use of functional capacity evaluations in workplaces and the compensation system: A report on workers' and report users' perceptions. *Work* 2004;23:67-77.
 31. Cocchiarella L, Andersson GBJ, editors. *Guides to the Evaluation of Permanent Impairment*. 5th edition. USA: AMA Press; 2001.
 32. Colledge AL, Holmes EB, Randolph Soo Hoo E, Johns RE, Kuhnlein J, DeBerard S. Motivation determination (sincerity of effort): the performance APGAR model. *Disability Medicine* 2001;1(2):5-18.
 33. Bordini EJ, Chaknis MM, Ekman-Turner RM, Perna RB. Advances and issues in the diagnostic differential of malingering versus brain injury. *NeuroRehabilitation* 2002;17:93–104.
-

34. Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS. Chronic pain disability exaggeration/malingering and submaximal effort research. *Clin J Pain* 1999; 15(4):244-74.
 35. Mendelson. *Psychiatric aspects of personal injury claims*. Chicago : Charles C Thomas; 1988.
 36. Drob SL, Meehan KB, Waxman SE. Clinical and conceptual problems in the attribution of malingering in forensic settings. *J Am Acad Psychiatry Law* 2009;37:98–106.
 37. Main CJ, Spanswick CC . “Functional overlay” and illness behaviour in chronic pain: distress or malingering: conceptual difficulties in medico-legal assessment of personal injury claims. *J Psychosom Res* 1995;39(6):737-753.
 38. Slick DJ, Tan JE, Strauss EH, Hultsch DF. Detecting malingering: a survey of experts’ practices, *Arch Clin Neuropsychol*, 2004;19:465–473.
 39. Sreenivasan S, Eth S, Kirkish P, Garrick T. A practical method for the evaluation of symptom exaggeration in minor head trauma among civil litigants. *J Am Acad Psychiatry Law*. 2003;31(2):220-231.
 40. Iverson GL. Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Appl Neuropsychol* 2006;13(2):77-90.
 41. Bianchini KJ, Greve KW, Glynn G. On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine J* 2005;5:404-417.
-

42. Larrabee, GJ. Exaggerated pain report in litigants with malingered neurocognitive dysfunction. *Clin Neuropsychol* 2003;17(3):395-401
43. Etherton JL, Bianchini KJ, Greve KW, Ciota MA. TOMM is unaffected by laboratory-induced pain: implications for clinical use. *Arch Clin Neuropsychol* 2005;20:375-384.
44. Iverson GL, King RJ, Scott JG, Adams RL. Cognitive complaints in litigating patients with head injuries or chronic pain. *J Forensic Neuropsychol* 2001;2:19-30.
45. Meyers JE, Diep A. Assessment of malingering in chronic pain patients using neuropsychological tests. *Appl Neuropsychol* 2000;7:133-139.
46. Gervais RO, Green , Allen LM, Iverson GL. Effect of coaching on symptom validity testing in chronic pain patients presenting for disability assessment. *J Forensic Neuropsychol* 2001;2:1-9.
47. Kay NR, Morris-Jones H. Pain clinic management of medico-legal litigants. *Injury* 1998; 29:305-308.
48. Emsley R, Coetzer P. Disability claims on psychiatric grounds. *S Afr Med J* 1996; 191:237-241.
49. Inman TH, Berry DTR. Cross-validation of indicators of malingering, a comparison of nine neuropsychological tests, four tests of malingering, and behavioural observations. *Arch Clin Neuropsychol* 2002;17:1-23.
50. Brown RJ. Introduction to the special issue on medically unexplained symptoms: Background and future directions. *Clin Psychol Review* 2007;27(7):769-780.

51. Goldenberg DL. Fibromyalgia: why such controversy? *Ann Rheum Dis* 1995; 54:3-5.
 52. Williams DA, Clauw DJ. Understanding Fibromyalgia: Lessons from the Broader Pain Community. *J Pain* 2009; 10(8):777-791.
 53. Wolfe F. The fibromyalgia problem. *J Rheumatol* 1997;24(7):1247-9.0315-162.
 54. Bennett RM. Fibromyalgia and the disability dilemma. A new era in understanding a complex, multidimensional pain syndrome. *Arthritis Rheum* 1996;39(10):1627-1634.
 55. Meyer HP. Fibromyalgia syndrome – current concepts. *SA Pharm J* 2007;8:8-17.
 56. Dersh J, Polatin PB, Leeman G, Gatchel RJ. The management of secondary gain and loss in medicolegal settings: Strengths and weaknesses. *J Occup Rehabil* 2004;14(4):267-279
 57. Freud, S. Introductory lectures on psychoanalysis. London: Hogarth; 1959 In: Dersh J, Polatin PB, Leeman G, Gatchel RJ. The management of secondary gain and loss in medicolegal settings: Strengths and weaknesses. *J Occup Rehabil* 2004;14(4):267-279
 58. Slick DJ, Sherman E, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol* 1999;13 (4):545-561.
-

59. Rogers R. Clinical Assessment of Malingering and Deception. 2nd ed. New York: Guilford Press; 1997.
 60. Frederick RI. Validity Indicator Profile. 2nd edition. Minnetonka: NCS Pearson Inc; 2003
 61. Rudy TE, Lieber SJ, Boston JR. Functional capacity assessment: influence of behavioural and environmental factors. *J Back Musculosk Rehabil* 1996;6:277-288.
 62. Matheson L. The functional capacity evaluation. In: Demeter S, Smith G (editors). *Disability Evaluation* (2nd edition). Chicago: Mosy Yearbook, 2003.
 63. Lees-Haley PR, Williams CW, Zasler ND, Marquillies S, English LT, Stevens KB. Response bias in plaintiffs' histories. *Brain Inj*, 1997;11(11):791-799.
 64. Lally SJ. What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 2003;34(5):491-498.
 65. Sbordone RJ, Seyranian GD, Ruff RM The use of significant others to enhance the detection of malingers from traumatically brain-injured patients. *Arch Neuropsych*, 2000;15(6):465-77.
 66. Sullivan K, Richer C. Malingering on subjective complaint task An exploration of the deterrent effects of warning *Arch Clin Neuropsychol* 2002;17:691-708.
-

67. Rohling ML, Green P, Allen LM, Iverson GL. Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Arch Clin Neuropsychol* 2002;17:205-222.
 68. Sbordone RJ, Seyranian GD, Ruff RM. Are the subjective complaints of traumatically brain injured patients reliable? *Brain Inj*, 1998;12(6):505-515.
 69. Lees-Haley PR, Dunn, JT. The ability of naive subjects to reports symptoms of mild brain injury, post-traumatic stress disorder, major depression, and generalised anxiety disorder. *J Clin Psychol* 1994;50(2):252-256.
 70. Wong JL, Regennitter RP, Barrios F. Base rate and simulated symptoms of mild head injury among normals. *Arch Clin Neuropsychol* 1994;9(3):411-425.
 71. Cano A, Johansen AB, Franz A. Multilevel analysis of couple congruence on pain, interference, and disability. *Pain*, 2005;118(3):369-379.
 72. Cripe LI. Limitations of records reviews. *Division of Clin Neuropsychol*. 2002;40(20. In Samuel RZ, Mittenberg W. Determination of malingering in disability evaluations, *Primary Psychiatry* 2005;12:60-68.
 73. The Oxford Pocket Dictionary of Current English. [Online] 2009. [access 2010, April 20]; Available: <http://www.encyclopedia.com/doc/1O999-credibility.html>
-

74. Rockwell P, Hubbard AE. The effect of attorneys' nonverbal communication on perceived credibility. *J of Credibility Ass and Witness Psych* 1999;2(1):1-13.
 75. Mobbs D, Lau HC, Jones OD, Frith CD. Law, responsibility, and the brain. *PLoS Biol* 2007;5(4):396-700.
 76. Illes J, Bird SJ. Neuroethics: A modern context for ethics in neuroscience. *Trends Neurosci* 2006;29(9):511-517.
 77. Yuille JC, editor. *Credibility Assessment*. Dordrecht: Kluwer Academic Publishers; 1989.
 78. DeBofksy MD. Disability benefits adjudication: Attorneys representing disability claimants. In Morgan JE, Sweet JJ, editors. *Neuropsychology of malingering casebook*. London: Psychology Press; 2008. pp.476 – 487.
 79. Blackmun J. *Daubert v. Merrell Dow Pharmaceuticals, Inc.* Washington, DC: United States Supreme Court; 1993.
 80. Le Page JA, Iverson GL, Collins P. The impact of judges' perceptions of credibility in fibromyalgia claims. *Int J Law Psychiatry*. 2008;31:30-40.
 81. Vickery CD, Berry DTR, Inman, TH, Harris MJ, Orey SA. Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures *Arch of Clin Neuropsychol* 2001;16:45-73.
-
-

82. Bond CF, DePaulo BM. Accuracy of deception judgements. *Personality and Social Psychol*, 2006;10:214-234.
83. Nies KJ, Sweet JJ. Neuropsychological assessment and malingering: A critical review of past and present strategies. *Arch Clin Neuropsychol* 1994;6:501-522.
84. Holmquist LA, Wanlass RL. A multidimensional approach towards malingering detection. *Arch Clin Neuropsychol* 2002;17:143-156.
85. Rogers R, Harrell EH, Liff CD. Feigning neuropsychological impairment: a critical review of methodological and clinical considerations, *Clin Psychol Review* 1993;3:255–274.
86. Bianchini KJ, Mathias CW, Greve KW. Symptom validity testing: a critical review. *Clin Neuropsychol* 2001;15:19-45.
87. Powell MR, Gfeller, JD, Hendricks BL, Sharland M. Detecting symptom- and test-coached simulators with the Test of Memory Malingering *Archiv Clin Neuropsychol* 2004;19:693-702.
88. Simonsen JC. Validation of sincerity of effort. *J Back Musculoskel Rehab*, 1996;6:289-295.
89. Lechner DE, Bradbury SF, Bradley LA. Detecting sincerity of effort: a summary of methods and approaches. *Phys Ther* 1998;78(8):867-888.
-

90. Sbordone RJ, Purish AD. Hazards of blind analysis of neuropsychological test data in assessing cognitive disability: The role of confounding factors. *Neurorehab* 1996;7:15-26.
 91. Larrabee GJ. *Assessment of malingered neuropsychological deficits*. USA: Oxford University Press; 2007.
 92. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959;56:81-105.
 93. Innes E, Straker L. Strategies when conducting work-related assessments. *Work* 2002; 19:149-65.
 94. Innes E, Straker L. Workplace assessments and functional capacity evaluations: current beliefs of therapists in Australia. *Work* 2003;20:225-36.
 95. Travis J. Cross-disciplinary competency standards for work-related assessments: Communicating the requirements for effective professional practice. *Work* 2002;19:269-80.
 96. Greve KW, Bianchini KJ. Setting empirical cut-offs on psychometric indicators of negative response bias: a methodological commentary with recommendations. *Arch Clin Neuropsychol* 2004;19:533-41.
 97. Willison J, Tombaugh TN. Detecting simulation of attention deficits using reaction time tests. *Arch Clin Neuropsychol* 2006;21(1):41-52.
-

98. Etherton JL, Bianchini KJ, Ciota ME, Heinly MT, Greve KW. Pain, malingering and the WAIS-II Working Memory Index. *Spine J* 2006;6:61-71.
 99. Samuel RZ, Mittenberg W. Determination of malingering in disability evaluations, *Primary Psychiatry* 2005;12:60-68.
 100. Greve KW, Ord JS, Bianchini KJ, Curtis KL, Prevalence of malingering in patients with chronic pain referred to a psychological evaluation in a medico-legal context. *Arch Phys Med Rehabil* 2009;90:1117-1126.
 101. Bianchini KJ, Curtis KL, Greve KW. Compensation and malingering in traumatic brain injury: a dose-response relationship? *Clin Neuropsychol* 2006;20(4):831-847.
 102. Burns N, Grove SK. *The practice of nursing research: conduct, critique, and utilization*. 5th edition. Missouri: Elsevier Saunders; 2005.
 103. American Educational Research Association, American Psychological Association and National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 1999.
 104. Law M, *Measurement in occupational therapy: scientific criteria for evaluation*. *Can J Occup Ther* 1987;54:133-138.
 105. Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. *Psychol Assess* 199;7(3): 238-247.
-

-
106. Theron FH, Hugo FJ, Spangenberg JJ, Hemp F, Maritz JS, Emsley RA. Validation of psychometric scales for malingering in a student sample. *SA Journ Psychiatry* 2001;7:96-102.
107. De Villiers N, Hugo FJ, Theron FH, Hemp F, Spangenberg JJ, Emsley RA. Psychometric tests for malingering in persons with a diagnosis of depression. *SA Journ Psychiatry* 2001;7:103-109.
108. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
109. Greve KW, Ord JS, Bianchini KJ, Curtis KL. Prevalence of malingering in patients with chronic pain referred to psychologic evaluation in a medico-legal context. *Arch Phys Med Rehabil* 2009;90:1117-1127.
110. Chafetz MD, Abrahams JP, Kohlmaier J. Malingering on the Social Security Disability Consultative Exam: A new rating scale. *Arch Clin Neuropsychol* 2007;22:1-14.
111. Schrader H, Obelieniene D, Bovim G, Surkiene D, Mickeviciene D, Miseviciene I, et al. Natural evolution of late whiplash syndrome outside the medicolegal context. *Lancet* 1996;347:1207-1211.
112. Cassidy JD, Carroll LJ, Coté P, Lemstra M, Berglund A, Nygren A. Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash injury. *New England J Med* 2000;342:1179-1186.
-

113. Trimble M. Somatoform disorders – A medico-legal guide. [s.l.]: Cambridge University Press; 2004.
114. Loeser JD, Henderlite SE, Conrad DA. Incentive effects of workers' compensation benefits: A literature synthesis. *Med Care Res Rev* 1995;52(1):34-59.
115. Rohling ML, Binder LM, Langhinrichsen-Rohling J. Money matters: a meta-analytical review of the association between financial compensation and the experience and treatment of chronic pain. *Health Psychol* 1995;14:537-547. In: Aronoff GM, Mandel S, Genovese E, Maitz EA, Dorto AJ, Klimek EH, Staats TE. Evaluating malingering in contested injury or illness. *Pain Practice* 2000;7(2):178-204.
116. Meyers JE, Volbrecht ME. A validation of multiple malingering detection methods in a large clinical sample. *Arch Clin Neuropsychol* 2003;18:261-276.
117. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd edition. New York: Oxford University Press; 1996.
118. Protection of Information Bill (No. 6 of 2010). [Online] 2010 [cited 2010, November 24]. Available: <http://www.info.gov.za/view/DownloadFileAction?id=118894>.
119. US Social Security Administration. Policy interpretation ruling [Online] 1996 [accessed 2010, July 25]; Available: http://www.ssa.gov/OP_Home/rulings/di/01/SSR96-07-di-01.html
120. Polaski v Heckler, 1984 Court of Appeals USA (A). Available: http://www.law.cornell.edu/socsec/course/readings/739_f2d_1320.htm
-

121. Hugo FJ, Hemp F. Malingering in clinical practice with specific reference to psychiatry and psychology. SA Med J 2002;92(5):54.
122. Hugo FJ, Thornton HB, Emsley RA, Theron FH, De Villiers N, Hemp F. The prevalence of malingering in persons seeking compensation on the basis of neuropsychiatric symptoms. SA Journ Psychiatry 2001;7(4):110-114.
123. Sullivan K, Lange R, Dawes S. Malingering base rates and detection methods in Australia. J Forensic Neuropsychol 2005;4(4):49-70.
-

APPENDIX A
DATA CAPTURE SHEET

claimant number	gender	age	diagnosis	Slick criteria A (Incentive)	Slick criteria B (Testing)	Slick criteria C (Self Reports)	Slick criteria D (Alternative factors)	Slick classification	CAT criteria A (ADL)	CAT criteria B (Medication)	CAT criteria C (Treatment)	CAT criteria D (Collateral - med)	CAT criteria E (Allegations, statements and evidence)	CAT score	CAT classification																																																																																																																																																																																																																																																
			Where:																																																																																																																																																																																																																																																												
			<table border="1"> <tr> <td>gender</td> <td>male</td> <td>female</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>age</td> <td>21-30</td> <td>31-40</td> <td>41-50</td> <td>51-60</td> <td>>60</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>diagnosis</td> <td>pain</td> <td>fibromyalgia</td> <td>depression</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Slick criteria A (Incentive)</td> <td>Monthly benefit</td> <td>Lump sum</td> <td>Utigation</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Slick criteria B (Testing)</td> <td>RB</td> <td>TD + KP</td> <td>TD + OBS</td> <td>TD + COLL</td> <td>TD + HIST</td> <td>NA</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Slick criteria C (Self Reports)</td> <td>EXAG</td> <td>SR + KP</td> <td>SR + OBS</td> <td>SR + COLL</td> <td>SR + HIST</td> <td>NA</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Slick criteria D (Alternative factors)</td> <td>Yes</td> <td>No</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Slick classification</td> <td>Probable</td> <td>Possible</td> <td>Not</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT criteria A (ADL)</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT criteria B (Medication)</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT criteria C (Treatment)</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT criteria D (Collateral - med)</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT criteria E (Allegations, statements and evidence)</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT score</td> <td>1</td> <td>10</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CAT classification</td> <td>Not</td> <td>Partial</td> <td>Fully</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>													gender	male	female														age	21-30	31-40	41-50	51-60	>60											diagnosis	pain	fibromyalgia	depression													Slick criteria A (Incentive)	Monthly benefit	Lump sum	Utigation													Slick criteria B (Testing)	RB	TD + KP	TD + OBS	TD + COLL	TD + HIST	NA										Slick criteria C (Self Reports)	EXAG	SR + KP	SR + OBS	SR + COLL	SR + HIST	NA										Slick criteria D (Alternative factors)	Yes	No														Slick classification	Probable	Possible	Not													CAT criteria A (ADL)	Not	Partial	Fully													CAT criteria B (Medication)	Not	Partial	Fully													CAT criteria C (Treatment)	Not	Partial	Fully													CAT criteria D (Collateral - med)	Not	Partial	Fully													CAT criteria E (Allegations, statements and evidence)	Not	Partial	Fully													CAT score	1	10														CAT classification	Not	Partial	Fully												
gender	male	female																																																																																																																																																																																																																																																													
age	21-30	31-40	41-50	51-60	>60																																																																																																																																																																																																																																																										
diagnosis	pain	fibromyalgia	depression																																																																																																																																																																																																																																																												
Slick criteria A (Incentive)	Monthly benefit	Lump sum	Utigation																																																																																																																																																																																																																																																												
Slick criteria B (Testing)	RB	TD + KP	TD + OBS	TD + COLL	TD + HIST	NA																																																																																																																																																																																																																																																									
Slick criteria C (Self Reports)	EXAG	SR + KP	SR + OBS	SR + COLL	SR + HIST	NA																																																																																																																																																																																																																																																									
Slick criteria D (Alternative factors)	Yes	No																																																																																																																																																																																																																																																													
Slick classification	Probable	Possible	Not																																																																																																																																																																																																																																																												
CAT criteria A (ADL)	Not	Partial	Fully																																																																																																																																																																																																																																																												
CAT criteria B (Medication)	Not	Partial	Fully																																																																																																																																																																																																																																																												
CAT criteria C (Treatment)	Not	Partial	Fully																																																																																																																																																																																																																																																												
CAT criteria D (Collateral - med)	Not	Partial	Fully																																																																																																																																																																																																																																																												
CAT criteria E (Allegations, statements and evidence)	Not	Partial	Fully																																																																																																																																																																																																																																																												
CAT score	1	10																																																																																																																																																																																																																																																													
CAT classification	Not	Partial	Fully																																																																																																																																																																																																																																																												
			Legend:																																																																																																																																																																																																																																																												
			<table> <tr> <td>RB</td> <td>Response Bias</td> <td>OBS</td> <td>Observation</td> </tr> <tr> <td>EXAG</td> <td>Exaggeration</td> <td>COLL</td> <td>Collateral</td> </tr> <tr> <td>TD</td> <td>Test Data</td> <td>HIST</td> <td>History</td> </tr> <tr> <td>SR</td> <td>Self Report</td> <td>NA</td> <td>Not Applicable</td> </tr> </table>													RB	Response Bias	OBS	Observation	EXAG	Exaggeration	COLL	Collateral	TD	Test Data	HIST	History	SR	Self Report	NA	Not Applicable																																																																																																																																																																																																																																
RB	Response Bias	OBS	Observation																																																																																																																																																																																																																																																												
EXAG	Exaggeration	COLL	Collateral																																																																																																																																																																																																																																																												
TD	Test Data	HIST	History																																																																																																																																																																																																																																																												
SR	Self Report	NA	Not Applicable																																																																																																																																																																																																																																																												

APPENDIX B
SCORING CRITERIA

Adapted Slick criteria (as proposed by Aronoff et al, 2000)

Criteria for Malingering		Detail	Example
A	Presence of substantial external incentive		
B	Evidence from testing (1. Definite negative response bias) 2. Probable response bias 3. Discrepancy between test data and known patterns of brain functioning 4. Discrepancy between test data and observed behaviour 5. Discrepancy between test data and reliable collateral reports 6. Discrepancy between test data and documented background history	2. Performance on one or more indices designed to measure exaggeration or fabrication is consistent with feigning 3. Markedly discrepant from accepted models of normal; discrepancy must be consistent with attempt to exaggerate or fabricate. 4. Performances on two or more tests within same domain are discrepant with observed level of function 5. Performance on two or more tests within same domain are discrepant with day-to-day level of function described at least one collateral informant 6. Poor performance on two or more standardised tests of function or symptoms within specific domain that is inconsistent with medical history	2. Grip strength: REG / FHPT / CV / PACT or neuropsychological test results or BDI exceeding 40 3. Concentration increases over time during depression 4. Hip ROM vs bended lift vs squat (vs ADL toilet) or two memory test and recall of instruction 5. Gait analysis vs gait during FCE vs reported gait at home vs surveillance 6. Component testing (ROM, MS, sensation or BDI, HADS)
C	Evidence from self-report 1. Self-reported history is discrepant with documented history 2. Self-reported symptoms are discrepant with known patterns of brain/neuro/psyche functioning 3. Self-reported symptoms are discrepant with behavioural observations 4. Self-reported symptoms are discrepant with information obtained from collateral informants 5. Evidence from exaggerated or fabricated psychological dysfunction	1. Marked difference 2. Number, pattern, severity of symptoms 3. Observed behaviour as opposed to test 4. Symptoms, history or observed behaviour 5. Focus is on exaggerated response of psychological adjustment or distress	1. Denies pre-existing back problem 2. No memory for child's name with simple depression 3. Inability to bend 4. Treating doctor/therapist 5. Catastrophising
D	Behaviour meeting necessary criteria from groups B, or C are not fully accounted for by psychiatric, neurologic, or developmental factors		Strong indication of sick role
II. Probable	Two or more types of probable evidence of intent from B criteria (B2–B6) or one B criterion (B2–B6) and one or more C criteria		
III. Possible	Evidence from self-report (one or more of criteria C1-C5).		

Credibility Assessment Tool

			Not consistent with the objective evidence and/or expected outcome/severity (0 points)	Partially consistent with the objective evidence and/or expected outcome/severity (1 point)	Fully consistent with the objective evidence and/or expected outcome/severity (2 points)
A	Impact of symptoms or condition on ADL	self-reported and/or self-administered questionnaires sustained and regular performance			
B	Type, dosage, effectiveness, and side effects of medications	need for large doses or multiple medication, addictive behaviour, pattern of increasing use which is well monitored, alternative medicine, side-effects, cessation of any medication and self-medicating behaviour			
C	Treatment sought and received (incl rehab)	professional help, compliance, attempts to treat the condition or find relief by attempting multiple treatments, frequency of visits, cost, risk, success rate, "reasonable man"			
D	Opinions about function given by other treating and examining sources in the file	<ul style="list-style-type: none"> The opinions of practitioners who have examined the patient are given greater weight than the opinions of those who have not Treating sources are given greater weight rather than providers of one-time examinations. A source that provides supporting evidence to substantiate the opinion about functional ability should be given more weight than should a source that does not have supporting evidence. Opinions most consistent with the preponderance of evidence are given greater weight. The opinion of a specialist in the field may be given greater weight than would that of a generalist, even if the length of treatment by the specialist was much less. Furthermore, the opinion of a physician who is more familiar with the demands and tasks in the workplace is likely to be given greater weight than would the opinion of a physician who is unaware of such demands. 			
E	(Inconsistencies or conflicts in the) allegations, statements, or medical evidence in the file	<ul style="list-style-type: none"> The degree to which the allegations are consistent with the objective evidence. The history given at different examinations. The consistency of the history of the injury/illness, the onset and duration of symptoms, and the functional effects on ADL as reported to various medical professionals. 			
		Total credibility score of 0-3= Not credible Total credibility score of 4-7= Partially credible Total credibility score of 8-10= Fully credible		Total credibility score= _____ (0-10)	

APPENDIX C
INFORMED CONSENT

PARTICIPANT INFORMATION LEAFLET AND CONSENT FORM

TITLE OF THE RESEARCH PROJECT:

A CRITICAL REVIEW OF THE VALIDITY OF THE CREDIBILITY ASSESSMENT TOOL (CAT) AND ITS' APPLICATION TO THE SCREENING OF SUSPECTED MALINGERING

REFERENCE NUMBER: N08/02/045

PRINCIPAL INVESTIGATOR: KAREN S. THEUNISSEN

ADDRESS: P O BOX 31216
TOKAI, 7966

CONTACT NUMBER: 083 6633348

You are being invited to take part in a research project. Please take some time to read the information presented here, which will explain the details of this project. It is very important that you are fully satisfied that you clearly understand what this research entails and how you could be involved. Also, your participation is **entirely voluntary** and you are free to decline to participate. You are also free to withdraw from the study at any point, even if you do agree to take part.

This study has been approved by the Health Research Ethics Committee (HREC) at Stellenbosch University and will be conducted according to the ethical guidelines and principles of the international Declaration of Helsinki, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research.

You can contact the Health Research Ethics Committee at 021-938 9207 if you have any concerns or complaints. You will receive a copy of this information and consent form for your own records.

What is this research study all about?

- Effective disability assessments are required within the insurance industry, moreso given the prevalence of malingering and its relevance to pain and/or depression. Although there have been a number of methods proposed to ascertain the sincerity of effort, there has been limited focus on developing a tool which could be used for the purposes of multidisciplinary assessment.
- This study will involve a paper-based analysis of the researcher's own reports and accompanying referral information from 2005 to 2009 of all cases for pain and/or depression.
- Of note is that there is no impact on the outcome of the payment of the claim as retrospective data is used and the research findings will not be documented in the insurer's claim file. No personal identifiers will be recorded on the data capture sheet and information will be analysed collectively, thereby insuring anonymity of all cases. A colleague with similar professional experience will perform a peer check of a random sample of completed analyses which will also be kept anonymous.

Why have you been invited to participate?

All insurance companies have been asked to participate.

What will your responsibilities be?

As the medical reports are legally owned by the insurer, written informed consent will be required for ethical reasons.

Given that the researcher was forwarded available medical information on file upon request for assessment, and possess her assessment report, no further information will be required from the insurer.

Will you benefit from taking part in this research?

There is future benefit in the research in that the application of the tool can lead to a more scientific approach to disability assessment, which in turn will minimise future financial risks to the insurer.

Declaration by participant

By signing below, I agree to take part in the above-mentioned research study.

I declare that:

- I have read this information and consent form.
- I have had a chance to ask questions and all my questions have been adequately answered.
- I understand that taking part in this study is **voluntary** and I have not been pressurised to take part.
- I may choose to leave the study at any time and will not be penalised or prejudiced in any way.

Signed at (*place*) on (*date*) 2010.

.....

Signature of participant

Signature of witness

Declaration by investigator

I, Karen S. Theunissen, declare that:

- I explained the information in this document to
- I encouraged him/her to ask questions and took adequate time to answer them.
- I am satisfied that he/she adequately understands all aspects of the research, as discussed above.

Signed at (*place*) on (*date*) 2010.

.....
Signature of investigator

Signature of witness

APPENDIX D
ETHICS APPROVAL



UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvermoet • your knowledge partner

18 November 2009

MAILED

Mrs KG Brink Theunissen
Department of Occupational Therapy
2nd Floor, Teaching building
Stellenbosch University
Tygerberg campus
7505

Dear Mrs Brink Theunissen

"A critical review of the credibility assessment tool and its application to the screening of mailgating."

ETHICS REFERENCE NO: N08/02/046

RE : APPROVAL

At a meeting of the Health Research Ethics Committee that was held on 5 March 2008, the above project was approved on condition that further information is submitted.

This information was supplied and the project was finally approved on 18 November 2009 for a period of one year from this date. This project is therefore now registered and you can proceed with the work.

Please quote the above-mentioned project number in ALL future correspondence.

Please note that a progress report (obtainable on the website of our Division: www.sun.ac.za/irds) should be submitted to the Committee before the year has expired. The Committee will then consider the continuation of the project for a further year (if necessary). Annually a number of projects may be selected randomly and subjected to an external audit. Translations of the consent document in the languages applicable to the study participants should be submitted.

Federal Wide Assurance Number: 00001372

Institutional Review Board (IRB) Number: IRB0005239

The Health Research Ethics Committee complies with the SA National Health Act No.61 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 Part 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes 2004 (Department of Health).

Approval date: 18 November 2010

Expiry date: 18 November 2011

03 June 2010 10:49

Page 1 of 2



Fakulteit Gesondheidswetenskappe - Faculty of Health Sciences





UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvenoot • your knowledge partner.

Yours faithfully

MRS EL ROHLAND

RESEARCH DEVELOPMENT AND SUPPORT

Tel: 021 938 9677 / E-mail: elr@sun.ac.za

Fax: 021 931 3352

03 June 2010 10:49

Page 2 of 2



Fakulteit Gesondheidswetenskappe · Faculty of Health Sciences



APPENDIX E
ORIGINALITY REPORT

Turnitin Originality Report

A Critical Review Of The Validity Of The Credibility Assessment Tool (CAT) And Its'
Application To The Screening Of Suspected Malingering by KS BRINK



From Research Assignment pre final (Advanced Occupational Therapy (MS03094) - Gevorderde
Arbeidsterapie 2010)

- Processed on 10-11-10 14:34 SAST
- ID: 152641729
- Word Count: 21410

Similarity Index

31%

Similarity by Source

Internet Sources:

22%

Publications:

19%

Student Papers:

11%