

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

The Genesis and Development of the Groot
Schoor Neurocognitive Screening Battery:

A neurocognitive screening tool for the South African context

Ross Balchin

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

UNIVERSITY OF CAPE TOWN

July 2008

Declaration

This study was conducted from 2004 to 2008 under the supervision of Professor Mark Solms.

I hereby declare that this is my own work and has not been presented for a degree at another university.

Ross Balchin

July 2008

University of Cape Town

Acknowledgements

I would like to sincerely thank the following people:

My mentor and supervisor, Professor Mark Solms for his invaluable insights and excellent supervision throughout the study.

My parents, for their constant love and support.

Eleni Pantellis for her work in designing and compiling the GSNSB Prototype.

Jill Mosdell, Dylan Lopich, Warren King, Nazanin Madadi and Jessica Oliver for their valuable help with collecting data and assisting with analysis.

Nolu Tyam, Anthea Adams and Goodwell Fihla, for their tireless translation and interpreting work, and for serving so effectively on the panel of cultural and language experts.

Dr Oz Ameen, for his support, and for his assistance in sourcing patients.

The financial support of the University of Cape Town through its KW Johnston Bequest and Harry Crossley scholarship awards.

All the participants — patients and controls — whose willingness to be tested made the study possible.

Table of Contents

Chapter One	Introduction	Page
	Epidemiology	3
	Background to Neuropsychology in South Africa	8
	The Groote Schuur Neurocognitive Screening Battery and its Background	22
	Aim of the Study	31
Chapter Two	Evaluating the Efficacy of the Mini-Mental State Examination in South Africa	32
	Studies Using the Mini-Mental State Examination	36
	Limitations of the Mini-Mental State Examination	37
	Rationale	40
	Methodology	42
	Results	46
	Discussion	70
Chapter Three	Translating the Groote Schuur Neurocognitive Screening Battery	79
	Translation of the Groote Schuur Neurocognitive Screening Battery	79
	Rationale	82
	Chronological Order of the Phases of the Groote Schuur Neurocognitive Screening Battery's translation Process	83

Chapter Four	Creation of Neurocognitive Tests for the South African Context	98
	Culture: Its Influence on Neurocognitive Tests	101
	Education: Its Influence on Neurocognitive Tests	103
	Memory Function	105
	Language Function	112
	Spatial Cognition	117
	Executive Function	119
	Rationale	122
	Methodology	123
	Results	148
	Discussion	171
Chapter Five	Re-piloting Three Neurocognitive Tests	185
	Rationale	185
	Methodology	186
	Results	195
	Discussion	207
Chapter Six	Establishing the Reliability and Validity of the Groote Schuur Neurocognitive Screening Battery	213
	Validation in Relation to Neuropsychological Evaluation	213
	Reliability in Relation to Neuropsychological Evaluation	220
	Rationale	221
	Methodology	222

	Results	232
	Discussion	269
Chapter Seven	The Implementation of Post-validation changes to The Groote Schuur Neurocognitive Screening Battery	294
	Rationale	294
	Conclusion	306
References		308
<u>Volume Two</u>		
Appendix A	GSNSB Prototype	337
Appendix B	GSNSB English Version	338
Appendix C	GSNSB Afrikaans Version	339
Appendix D	GSNSB isiXhosa Version	340
Appendix E	GSNSB Post-validation Modifications	341
Appendix F	Mini-Mental State Examination	342
Appendix G	Age and Education Weighted Norms Table for the Mini-Mental State Examination	344
Appendix H	Screening Sheet	345
Appendix I	Consent Form	346
Appendix J	Patient Information Sheet	347
Appendix K	Scoring Sheet	348

Abstract

The multicultural, multilingual composition of South Africa's population poses many complex challenges for healthcare professionals working within this environment. This applies especially to mental healthcare. The burden of providing neurocognitive screening in South Africa currently falls largely upon the widely used, yet outdated and diagnostically limited, Mini-Mental State Examination. Neuropsychology is a fledgling field in South Africa, and specialist expertise is scarce. In general, the neurocognitive tests that are available for use in the South African context were created and normed in Europe and North America, and are thus culturally biased when used on South Africa's population. Tests such as the Mini-Mental State Examination struggle to carry their diagnostic responsibility in this setting, which severely compromises their clinical utility. Consequently, there is an urgent need in South Africa for a clinically efficient, diagnostically useful neurocognitive screening tool, which can serve as 'transferable technology'.

This study sought to address this need by developing and validating the Groote Schuur Neurocognitive Screening Battery Prototype, a theory-driven tool created in response to requests by Groote Schuur Hospital neurologists for a suitable replacement for the Mini-Mental State Examination. The design of this battery constitutes a 'middle ground' between the hypothetico-deductive and psychometric approaches to neurocognitive assessment. However, the Prototype had a number of limitations: it was not available in languages other than English, it contained many culturally biased tests and its validity and reliability had yet to be formally established.

To address these limitations, the study met five major objectives. Firstly, a detailed appraisal of the efficacy of the Mini-Mental State Examination in South Africa was undertaken.

Secondly, the battery was translated into both Afrikaans and isiXhosa. Thirdly, the cultural bias inherent in the neurocognitive tests being utilised in the Prototype was addressed by developing nine new, culturally appropriate tests for inclusion in the Groote Schuur Neurocognitive Screening Battery prior to its validation. This process drew on cultural and neuropsychological expertise and used pilot studies on healthy individuals to test the appropriateness of these new tests. Both qualitative and quantitative analyses were undertaken in order to offer converging lines of evidence as to the effectiveness of the process adopted.

Fourthly, once satisfactory tests had been created, the reliability and validity of the Groote Schuur Neurocognitive Screening Battery was formally investigated. This process, drawing again on both qualitative and quantitative lines of evidence, involved the investigation of the battery's ability to differentiate between cases with various lesion-sites and controls in a theoretically meaningful manner.

Findings indicate that the Groote Schuur Neurocognitive Screening Battery as a product of the development undertaken during this study displays both good reliability and validity and that it is also culturally appropriate for use in the South African context.

Fifthly, the final contribution of this study was to implement further improvements in the decision-tree design of the battery based on the outcomes of the reliability and validity study.

CHAPTER ONE: INTRODUCTION

Of all the measures currently being used in industry, clinical and counselling practice, and in educational and forensic contexts, most have not been thoroughly researched for bias, very few cross-cultural studies have been published on their use here, and very few are available in a variety of South African languages.

(Foxcroft & Roodt, 2005, p. 254)

Healthcare professionals working in South Africa are faced with many unique and complex challenges, not the least of which is the fact that 11 official languages are spoken in this multicultural society. The legacy of Apartheid has also left large numbers of the population — especially among the current adult generation — poorly educated. Poverty is widespread and the occurrence of diseases and accidents causing brain dysfunction is also extremely high, as is shown by the review of available epidemiological data below. The high prevalence of these diseases and pathologies underscores the need for neurocognitive screening and assessment in South Africa.

Epidemiology

The available epidemiological literature in South Africa highlights some of the main areas where neurocognitive assessment is required. Up-to-date statistics, however, are not readily available; Table 1.1 (see p. 7) provides the latest data of the pathologies most frequently seen at Groote Schuur Hospital, where this study was conducted. Groote Schuur is a fairly typical state hospital in South Africa and is representative of other state hospitals in the country in that the patient population served are of largely very low socio-economic status, speak a variety of languages, and come from diverse cultural backgrounds.

Table 1.1

Epidemiology of neurocognitive dysfunction

<i>Cause of brain pathology</i>	<i>Total cases South Africa</i>	<i>Western Cape</i>
Stroke (2001)	300/100 000 (Limpopo Province only)	Statistics unavailable
HIV/AIDS (year 2006)	6 100 000 (estimated)	Statistics unavailable
Road Traffic Accidents (persons <i>seriously</i> injured - 1998)		
Total:	36 246	Statistics unavailable
Drivers of all vehicles:	10 875	
Passengers:	14 733	
Bicycles:	752	
Pedestrians:	9886	
Tuberculosis (year 2002)		
all cases:	224 286	39 650
pulmonary:	182 690	29 840
Epilepsy (year 2000)		
Internationally:	5/1000 people (estimated) (South Africa: unknown, but likely to be higher)	Statistics unavailable
Dementia	Statistics unavailable	Statistics unavailable

In South Africa a wide range of debilitating diseases, and the high rates of road traffic accidents and violent crime, affect many members of the population. Among these diseases are: HIV/AIDS, tuberculosis (TB), strokes of various aetiology, dementias of various aetiology, and epilepsy (Bryer, Eastman, Kies, Lee Pan, Philcox & Silber, 2000). All of these devastating disease processes have a significant neurological component, accompanied by varying cognitive sequelae (Bryer et al., 2000). Additionally, incidence rates of head trauma are extremely high in South Africa, with violent crime, motor vehicle, motorcycle, bicycle and pedestrian accidents contributing to these statistics.

Stroke

Given that the Neurology Department at Groote Schuur Hospital has a large Stroke Unit, many stroke patients are referred for neuropsychological assessment. Studies on the epidemiology of stroke in South Africa are, however, rare (Fritz, 1997). Where prevalence rates are available, they have been gleaned from studies on sections of the population or from burden of disease estimates of the population as a whole. “Stroke is the second commonest cause of death worldwide, with two-thirds of these deaths occurring in developing regions of the world like Sub-Saharan Africa” (Connor & Bryer, 2006, p.195). “From [the] ‘Initial burden of disease estimates for South Africa, 2000’ stroke was found to be the fourth commonest cause of death accounting for 6% of all deaths in 2000” (Connor & Bryer, 2006, p.196). In the Western Cape, where Groote Schuur Hospital is situated, stroke accounted for 8 percent of deaths in the year 2000 (Bradshaw, et al., 2000).

Prevalence rates of stroke in South Africa are extremely difficult to determine, with the South African Stroke Prevention Initiative (SASPI) having recently conducted the first prevalence study for South Africa in the Limpopo Province (Connor & Bryer, 2006). The crude incidence from this study was 300 strokes per 100 000 individuals (Connor & Bryer, 2006). “No community-based incidence studies have ever been done in South Africa or in Southern Africa because of the difficulty of performing them” (Connor & Bryer, 2006, p. 197). A study conducted more recently estimated that 60 people die from cerebrovascular accidents (CVAs) daily in South Africa (Heart and Stroke Foundation South Africa, 2007).

HIV/AIDS

“HIV/AIDS was thrust upon a country that, in its new birth of democracy, was addressing several challenges, which included redressing the imbalances of its past. South Africa, with

an estimated population of 47.4 million, has an adult literacy rate of 82% and urbanization rate of 49%. In 12 years, HIV prevalence in 15–49-year-olds rose from less than 1% to about 20%” (UNAIDS table of country-specific HIV/AIDS estimates and data, 2004). The United Nations (UN) estimated that more than 6 million South Africans were infected with HIV at the end of the year 2006 (provided in the UNAIDS ‘2006 Report on the global AIDS epidemic’).

Traumatic Brain Injury (TBI)

Violence and Road Traffic Accidents (RTA) are a major cause of brain injury in South Africa. According to Gilbert and Tollman (2007), violence accounts for 45 to 55 percent of all Traumatic Brain Injuries (TBI) in South Africa, with transport-related causes contributing 20 to 25 percent, the remaining 20 percent being from incidental causes. It is reported that 84.1 percent of traumatic head injuries occurring in Cape Town (where Groote Schuur Hospital is situated) in 1995, involved families whose monthly income was less than R1000 (\$200) (Gilbert & Tollman, 2007; Medical Research Council, 1999). Statistics South Africa stopped publishing survey data regarding Road Traffic Accidents in 1998, while the Arrive Alive division of the Government’s Department of Transport publish annual regional and national data, but only for the occurrence of vehicle fatalities, not injuries. The latest available figures are shown in Table 1.1 (Statistics South Africa, 2004).

Tuberculosis

The reported cases of all forms of tuberculosis (TB), both nationally and in the Western Cape, shed light on the prevalence and severity of this disease. The Department of Health provides data on the prevalence rates of both tuberculosis and pulmonary tuberculosis as a

subtype. Although specific data on tubercular meningitis, a variant of tuberculosis, is not provided, the prevalence of this serious disease may be gleaned from the other data.

Epilepsy

According to Bryer et al. (2000), epilepsy is a common disorder, the estimated prevalence rate of active epilepsy (that is, a seizure occurring within the past 2 years) being about 5/1000 persons. Prevalence data for South Africa are not currently available, but it is thought that the prevalence rate in this country is greater.

Dementia

Many types of dementia are regularly encountered in our patient population. Among these are: Alzheimer's Disease, Vascular Dementia, Fronto-temporal Dementia, Pick's Disease, Alcoholic Dementia, Parkinsonian dementia and, rarely, Jakob-Creutzfeld Disease. Each dementing process has its own distinct pattern of cognitive sequelae. There are no available prevalence rates for the various types of dementia in South Africa. Additionally, many of these dementia types have multiple variants and it is often only after autopsy that a final diagnosis can actually be made.

Currently, the vast majority of South Africans affected by neurocognitive dysfunction resulting from the above-mentioned pathologies do not benefit from any form of formal neuropsychological assessment. For example, the HIV/AIDS epidemic highlights the urgent need for neuropsychological assessment, bearing in mind the fact that HIV can have significant neurocognitive sequelae. The information provided by neuropsychological evaluation of these patients would not only help track the progress of the disease in the individual concerned, but also make a meaningful contribution in determining the treatment

required and the mode of treatment administration and compliance monitoring. However, before the benefits of such assessments can be seen, appropriate neuropsychological tests, tailor-made for South Africa's multicultural population, are needed.

Background of Neuropsychology in South Africa

Neuropsychology is a fledgling field in South Africa. It was only in 2006 that the Health Professions Council of South Africa agreed to recognise neuropsychology as a registerable professional category, alongside and on an equal footing with the existing categories, such as clinical psychology. This was the culmination of a long and arduous journey of discussion, negotiation and debate, which began as early as the 1980s. Consequently, the practice of neuropsychology in South Africa has, to date, fallen under the domain of clinical psychology, largely following psychometric lines of approach to clinical practice.

In 2002, the University of Cape Town (UCT) became the first tertiary institution in South Africa to offer specialised training in neuropsychology, separate from clinical psychology. Under the tutorage of Professor Mark Solms, the Neuropsychology division of the Psychology Department — located in the Neurology Department — began training neuropsychologists at Groote Schuur Hospital (GSH). Neuropsychologists from this Department are trained predominantly in the qualitative (hypothetico-deductive) approach to clinical assessment.

It is in this setting, within the historical, cultural, linguistic and socio-economic context outlined above, that this research project was conceptualised. The key objective of this research was to provide the country with a much-needed, clinically effective and culturally specific, neurocognitive screening tool. The development of a tool such as this involved

addressing the complex interaction of education, culture and linguistics on neurocognitive testing in the clinical setting. It also involved taking these variables into account when adapting many of the 'bedside' neurocognitive tests included in what is now called the Groote Schuur Neurocognitive Screening Battery (GSNSB). Before examining the genesis and development of the GSNSB, an overview of the different approaches to neuropsychological assessment is given.

Three Distinct Paradigms in Neuropsychological Assessment

Neurocognitive testing in South Africa takes one of three forms. The first of these, *screening*, is usually used early on in the patient's treatment, in order to provide a gross estimate of cognitive abilities. If deficits are observed, the patient is then referred for neuropsychological assessment (Mitrushina, Boone, Razani & D'Elia, 2005; Nestor & Hodges, 2001). The patient would then undergo one of two forms of neuropsychological assessment, either a *psychometric assessment* using a standardised battery of tests, or a clinical neuropsychological assessment using the *qualitative/hypothetico-deductive* approach.

Neurocognitive Screening

Neurocognitive screening has a unique and specific role in the clinical setting. Unlike the hypothetico-deductive and psychometric assessment approaches, which are of final diagnostic value, screening provides a basic introductory investigation as to whether a patient does or does not display cognitive impairments. Screening measures can generally be used by non-specialists, provided they have been trained to administer them. On the other hand, diagnostic assessment tools are the preserve of specialists, as they are comprehensive and in-depth and can only be administered by qualified professionals (Luiz, Stroud & Jansen, 2005). Screening thus serves as a general, global measure of the presence or absence of cognitive

deficit, as an initial starting point, without going on further to explore possible deficits in more detail. Screening tools lack the detail, complexity and theoretical underpinnings to be of final diagnostic value.

There is a wide range of screening tools; some cover a broad array of functions, while others have a more narrow scope or specificity (Lezak, Howieson & Loring, 2004). It is often difficult to define which criteria a screening tool should be based on, as most deficits do not affect all patients. If a screening tool is too narrow in the range of deficits assessed, a number of deficits might go undetected (false negative errors). Alternatively, if the tool assesses a broader array of deficits, which may not be specific to any particular pathology, a large number of false positives may result (Lezak et al., 2004).

As this is usually the first form of testing that a patient receives on arrival at the hospital, patients are usually screened while still presenting with acute deficits. “Although [screening] is relatively quick and produces a numerical score that can be compared on consecutive assessments, a more extensive assessment is appropriate in specialist practice” (Nestor & Hodges, 2001, as cited in Mellers, 2004, p. 58). If the screening tool is used for the purpose for which it was intended, then: “a combination of tests, including some that are sensitive to specific impairment, some to general impairment, and others that tend to draw out diagnostic signs, will make the best diagnostic discriminations” (Lezak et al., 2004, p. 150).

In reality, most screening tools focus on one specific domain of neurocognitive function, and the majority do not have an adequate understanding of nosology incorporated into their design. “A key point in the screening debate is the suitability of currently available screening instruments: few screens have been validated in the populations for which they are intended

to be used, many have low accuracy for mild levels of impairment, and there are often demographic biases in score distributions. Although no single instrument for cognitive screening is suitable for global use, clinician surveys indicate that the Mini-Mental State Examination (MMSE) is overwhelmingly ubiquitous in practice” (Cullen, O’Neill, Evans, Coen & Lawlor, 2007, p. 790).

A recent survey conducted by Cullen et al. (2007) into the effectiveness of 39 internationally available screening tools — using a semi-structured pro-forma including reliability statistics, sample types, validity statistics by type of diagnosis and pertinent comments or criticisms contained in individual papers — concluded that only a few tools had particular strengths and that no tool was all encompassing in terms of its diagnostic value. This survey also concluded that future research should focus on refining existing tools and that such endeavours should concentrate on incorporating into this task a better understanding of symptom profiles in different diagnoses (Cullen et al., 2007).

One of the most widely used screening tools in South Africa is the Mini-Mental State Examination, a detailed account of which is provided in Chapter Two.

Luria and the Hypothetico-deductive Clinical Approach

Luria’s approach to neuropsychology rests on an understanding of neuroanatomical correlates of normal mental functions and how these relate to the specific deficits observed in the patient. This theory of the functional organisation of the brain is understood through viewing “higher cortical processes as complex, dynamically localised, functional systems that are affected differently with lesions of different parts of the cerebral hemispheres” (Luria, 1966, p. 586). In other words, the brain and its organisation are viewed as a complex functional

system, comprising an array of more basic functions, all acting in concert. From this perspective, man's "perception and action, his memory, speech and thinking, makes use of a highly complex system of concertedly working zones of cerebral cortex" (Luria, 1973, p. 341).

In terms of Luria's theory of brain function, 'functional system' refers to the interaction between various tissues, rather than just a single tissue (Luria, 1973; Solms & Turnbull, 2002). Just as the function of digestion involves multiple individual functions (such as the work of the bowel, stomach and intestine, all acting in concert), so too neurocognitive function operates in a similar manner, with the interaction of various neuroanatomical 'centres' throughout the brain contributing towards mental functioning (Hebben & Milberg, 2002; Luria, 1973; Solms & Turnbull, 2002; Walsh & Darby, 1999). From this point of view, "the task of neuroscience is therefore not to localise 'centres', but, rather, to identify the components of the various complex systems that interact to generate the mental functions. Luria called this task 'dynamic localization' " (Solms & Turnbull, 2002, p. 64).

At the heart of Luria's approach to understanding these functional systems lies the importance of qualitative observation and the need to generate hypotheses to test neurocognitive functions. "Luria's examinations consisted of sequences of observations organised into various decision trees reflecting the function that was being analyzed" (Hebben & Milberg, 2002, p. 154). This essentially involves investigating multiple possible determinants of failure, and eliminating as many explanations as possible (Walsh & Darby, 1999). Investigating brain functions using this approach is based on the principle of multiple determination whereby, "Behavioural deficits are defined in terms of impaired test

performance. But impaired test performance may be a final common pathway for expression of quite diverse types of impairment” (Walsh & Darby, 1999, p. 388).

Luria’s approach dictates that the neuropsychologist’s key role is to ‘qualify the symptom’ through the generation and testing of hypotheses about the complex underlying functional nature of the patient’s deficits (Luria, 1973). Here ‘qualification’ refers to focusing on *how* the patient failed any given measure, rather than simply the fact of whether they were able to pass or fail (Luria & Majowski, 1977). A careful analysis of the patient’s deficits is needed (Luria & Majowski, 1977). The neuropsychologist therefore requires a theoretical knowledge of the functional organisation of the brain in order to generate appropriate hypotheses. The testing of these hypotheses leads to the identification of a collectively meaningful group of symptoms that point to a definite underlying lesion-site (Luria & Majowski, 1977).

The hypothetico-deductive paradigm of dynamic neuropsychology provides an efficient and cost-effective clinical tool by means of which a comprehensive neurocognitive assessment can be performed. The approach involves a clinical/‘bedside’ assessment of the patient using a flexible range of tests designed to assess specific domains of cognitive function, which can be linked to neuroanatomical correlates within the brain. In this way, rather than merely reporting test scores (as is the case with psychometric practice), qualitative information is gathered that informs and directs the clinical assessment, resulting in a description and explanation of the patient’s individual deficits.

It is for the most part within this hypothetico-deductive/qualitative paradigm of assessment that clinical neuropsychology is practised at Groote Schuur Hospital. This is reflected in the battery described below.

The Psychometric Approach

In contrast to the hypothetico-deductive approach, the psychometric (normative) approach to neuropsychology is primarily concerned with measuring cognitive deficits (quantifying degrees of deviation from the norm), and utilises standardised test scores derived from normally distributed populations. One of the main purposes of this approach is to demonstrate statistically how a patient's test performance compares to that of "a standardised population (the group of individuals tested for the purpose of obtaining normative data on the test)" (Lezak et al., 2004, p.141). With this approach, "[p]redictions about the site of the lesion and its nature (diffuse or focal, static or changing) are based on statistically identified relationships between test scores" (Luria & Majovski, 1977, p. 961). One of the strengths of psychometric testing is that it allows a patient's individual test performances to be compared directly and precisely to other performances by that particular patient, as well as to the performances of other patients (Lezak et al., 2004). Another major benefit of this approach is that it allows for a quantification of the patient's test performance at a particular point in time that can then be measured against the patient's test performances in the future, allowing for any decline in performance to be empirically demonstrated (Anastasi & Urbina, 1997; Crawford, 2004; Glozman, 1999; Russell, Russell & Hill, 2005). The psychometric approach is particularly useful when the neuropsychologist is required to measure potential changes in cognitive function over time (Crawford, 2004).

A shortcoming of this approach is that "a test response is not a score; scores, where applicable, are abstractions designed to facilitate intra-individual and inter-individual comparisons ... to reason — or do research — only in terms of scores or score patterns is to do violence to the nature of the raw material. The scores do not communicate the full

response” (Schafer, as cited in Walsh, 1987, p. 335). In other words, despite individuals achieving comparable test scores, different possible determinants of failure could be responsible for the performance observed; the ‘sameness’ of the scores is therefore partially misleading (Walsh, 1991).

A further danger is that *if* the score is taken at face value, then the patient will be assumed to have a deficit of the corresponding neurocognitive function described in the test manual (Walsh, 1991). This highlights how the diagnostic use of psychometric tests in the clinical setting can be limited, if test scores are not interpreted meaningfully. Rather than explaining the symptoms, many neuropsychologists just measure the functions of the patient against a normative performance. Hence, psychometric tests, rather than providing a diagnosis, may sometimes act as the starting point (raising a clinical question) for further neurocognitive assessment. The psychometric approach can be expensive, the imported test batteries themselves being costly, and often time-consuming to administer and score.

Psychometric practice has traditionally involved *testing* as opposed to *assessment*. An important distinction exists between these two terms. Testing, which refers to a quantitative representation of the patient’s performance on various measures, allows for comparisons with normative data. It is also possible to compare the patient’s performance on different tests, as well as to make comparisons over time (Mitrushina et al, 2005).

On the other hand, in addition to psychometric evaluation of a patient in terms of normative data, *assessment* allows for qualitative behavioural observations regarding the patient’s test performance, focusing on the nature of dysfunctions observed and the underlying cognitive mechanisms involved (Mitrushina et al., 2005).

This psychometric approach is widely used in South Africa and a few test batteries have been normed for this population (or parts thereof). However, normed tests are extremely difficult to obtain and norms do not exist for many population groups. This is due to the large diversity of cultures in this country, varying levels of acculturation and the low education level attained by many of the population (many are illiterate). In addition, many communities are difficult to access, given that they often live in rural areas or in sprawling informal settlements near to cities, where the required infrastructure to conduct such tests is lacking . The resources required to conduct large-scale normative studies (such as money and expertise, including language proficiency) are also scarce.

Approaches to Neuropsychological Batteries

A battery of tests is understood not simply as a group of tests, but rather as a specifically assembled constellation of tests used in combination as part of an assessment method, with the purpose of reaching a neuropsychological conclusion (Russell et al., 2005; Walsh, 1985). Test batteries allow for patterns of neurocognitive deficits to be observed through the combination of tests assembled (Russell et al., 2005). There is an important distinction between a *screening* battery and an *assessment* battery. Assessment batteries are used for clinical assessments by trained clinical psychologists and neuropsychologists, whereas screening batteries are used to provide a brief preliminary investigation of overall cognitive function, and are utilised by a broader array of non-specialists, such as general practitioners, for example. This is a vital distinction in terms of the present study, as the field of neuropsychology has established stringent criteria regarding the development and validation of assessment batteries, yet few guidelines for screening batteries.

There are two primary approaches used in utilising assessment batteries — the *fixed battery* and the *flexible battery* approaches. The fixed battery approach dictates that a comprehensive battery of tests be used to assess all patients in a rigid order (Mitrushina et al., 2005). This standardised use of batteries allows for comparisons between individuals to be made, but does not usually investigate the possible multiple determinants of failure on any given test (Hebben & Milberg, 2002; Walsh & Darby, 1999). This approach relies on the use of standardised, well-normed tests.

For the flexible approach the hypotheses drive the process: the clinician reviews all available information about the patient and then selects relevant tests (Mitrushina et al., 2005). This allows for patient-specific investigations to be conducted, using tests in conjunction with behavioural observations and a clinical interview, to probe the influences underlying a test performance.

There are two primary subtypes of the flexible approach. Firstly, there are those batteries that use psychometrically scored tests in a flexible way, such as the Boston Process Approach. Critics of this approach say that it lacks supporting norms and the detailed standard methods required to adequately evaluate reliability and validity in conventional ways (Hebben & Milberg, 2002). Secondly, there are those batteries that use the aforementioned qualitative approach in a relatively standardised way, such as Luria's Neuropsychological Investigation (LNI).

Screening Batteries

A number of test batteries in addition to the Mini-Mental State Examination (MMSE) have been designed for both screening (of mental status) and neuropsychological assessment

purposes. Significantly, the vast majority are psychometric in nature and, while some of the latter have tried to incorporate qualitative observation and measures, it is almost always from this psychometric departure point. Almost all screening batteries focus on producing a final summation score, which is the end product of the assessment (Lezak et al., 2004). Some well-known screening tools include: the CAMCOG, which is the cognitive section of the CAMDEX (The Cambridge Mental Disorders in the Elderly Examination) (Huppert, Brayne, Gill, Paykel & Beardsall, 1995); the Mini-Mental State Examination (Folstein, Folstein & Mc Hugh, 1975); the Information Memory Concentration (IMC) test (Blessed, Tomlinson & Roth, 1968); and the Mattis (Mattis, 1976) Dementia Rating Scale (DRS) (Salmon & Hodges, 2003). Of these, the CAMCOG and the Mattis Dementia Rating scale are not suited to 'bedside' clinical assessment due to their length (Nestor & Hodges, 2003). Another well-known screening tool is the Modified Mini-Mental State Examination (3MS), which was produced by Teng and Chui (1987) (see Chapter Two).

Some screening tools are developed for general use. For example, the internationally used Neurobehavioral Cognitive Status Examination (now called the Cognistat) was designed using a summarised profile of scores, rather than the typical screening approach of using one single summation score (Lezak et al., 2004). "A 'screen and metric' approach is used in which an initial item at a near normal level of difficulty is tested first; patients who fail this general screening item are given easier tasks within that domain in an effort to establish floor level and to identify gradations of impairment" (Lezak et al., 2004). The tool covers brief assessment of: language, memory, orientation, level of consciousness, attention and calculation (Kiernan, Mueller, Langston & VanDyke, 1987). This tool is often used for the screening of dementia, as well as with psychiatric patients, but it is cautioned that such

screening tools lack specificity and sensitivity and are unable to provide detail about the domains of function examined (Milberg, 1996).

Another example of a screening battery for general use is the Neuropsychological Screening Battery for Hispanics (NeSBHIS), designed by Ponton, Satz, Herrera (1996) in order to provide appropriate assessment for Hispanics in the United States of America. The design incorporates 11 tests, covering memory function, language (including the FAS Test, and a Spanish version of the Boston Naming Test) and psychomotor function, reasoning and mental control (Lezak et al., 2004). This screening battery was normed on 300 Spanish speakers; stratified norms were developed, as the tests displayed good correlations with the variables of age and education (Ponton et al., 1996). This example illustrates how the psychometric approach is typically used in the context of screening tools.

Others are designed to assess one specific domain of cognitive function. For example, the Aphasia Screening Test, which takes about 30 minutes to administer, was designed by Halstead and Wepman (1959) to screen all aphasic disabilities. This screening battery comprises 51 items and is scored by coding the patient's errors onto a diagnostic profile (Lezak et al., 2004).

Assessment Batteries

Among the best known batteries used for neuropsychological testing/assessment are the Benton Neuropsychological Investigation, the Wechsler Memory Scale (WMS), the Boston Process Approach, the Halstead-Reitan Battery, the Luria-Nebraska Neuropsychological Battery (LNNB), the Delis-Kaplan Executive Function System (D-KEFS) and Luria's Neuropsychological Investigation (Kolb & Whishaw, 2003). Few of these batteries were

based on theory in their construction or use and, at the time of their creation, Alexander Luria was one of only a few clinicians who had attempted to construct a general theory of the cognitive function of the brain (Kolb & Whishaw, 2003).

In the light of this, it can be seen that there is a spectrum of designs within this array of test batteries. At the one end are the standardised batteries, such as the Luria-Nebraska Neuropsychological Battery and the Halstead-Reitan Battery, which are designed with fixed criteria for 'organicity' and little consideration of the functional organisation of brain systems (Kolb & Whishaw, 2003). These batteries are relatively easy to administer and score and require minimal knowledge of the functional organisation of the brain. The batteries at the other end of the spectrum, which are more qualitative than quantitative, are individualised and need to be administered and interpreted by clinicians with good theoretical knowledge. The individual patient's etiology, along with observations of his/her test performance, governs the testing to be carried out (Kolb & Whishaw, 2003).

An example of the latter type of battery is Luria's Neuropsychological Investigation. For this battery, Christensen formalised and compiled a manual of instructions, test cards and text drawn from Luria's original techniques and materials (Christensen, 1979; Lezak et al., 2004). This represents the qualitative equivalent of a psychometric battery, and is designed for use in a flexible manner. Here, sequences of decision-trees are used, organised according to specific functions, and detailed instructions are provided as to how to administer the tasks. The grading of these qualitative decision-trees takes the form of a 3-point scale, which involves deciding on the magnitude of each symptom and allocating it to one of three categories: 'none', 'mild' or 'marked' (Glozman, 1999; Luria, 1999). A number of tasks from Luria's

Neuropsychological Investigation, such as the 'Red/Green' Test and the 'Fist/Side/Palm' Test, are incorporated into the GSNSB (see Appendix A).

More recently, Glozman (1999) has developed a six-point scoring system as a way of capturing results from Luria's Neuropsychological Investigation. Here, zero represents intact function, with the overall aim being to produce a quantitative expression of the patient's pattern of deficits. As with Luria's original procedure, this scoring system retains the qualification of the symptom in the investigation of different functional domains.

In an attempt to capture Luria's concepts in empirical form, Charles Golden tried to standardise and norm the Luria Neuropsychological Investigation in creating the Luria-Nebraska Neuropsychological Battery. This tool utilises scales representing functions and left versus right hemisphere lesion locations, and was normed by evaluating combinations of these scales. However, this battery only contains some suggestions for qualitative observation, and is therefore to all intents and purposes quantitative in nature (Hebben & Milberg, 2002). As a result, this battery should probably be classified as a fixed battery rather than a middle ground between two contrasting approaches. Critics of the Luria-Nebraska Neuropsychological Battery say that it does not do justice to Luria's concepts, as equating functions to a single scale loses the descriptive variations that result from assessing performances qualitatively.

Finally, some batteries, such as the Boston Process Approach, fall between the two ends of the spectrum. These batteries incorporate aspects from both these perspectives into their design; a test is formally scored and compared with norms while, at the same time, qualitative observation is also factored into the analysis. Milberg and Hebben (2006) suggest that the

Boston Process Approach is in reality more of a general paradigm for the assessment of brain damage that focuses on observation, problem-solving abilities and description of deficits, than an actual test battery, despite the fact that specific tests are associated with it.

The Groote Schuur Neurocognitive Screening Battery and its Background

Five years ago, the Neuropsychology division at Groote Schuur Hospital began compiling relevant neurocognitive tests into a comprehensive, easy-to-use screening battery, along with its own instructions, decision-trees and scoring procedures (see Appendix A for the GSNSB Prototype). Before the practice of neuropsychology became established at Groote Schuur Hospital, the Neurology Department had been relying on the Mini-Mental State Examination (MMSE), administered by neurologists, as the only form of screening of their patients' mental status.

The MMSE is a widely used thirty-item screening test. It offers a quick screening for the presence of cognitive deficit, and requires a mere five to ten minutes to administer. The MMSE is a theory-blind battery — that is, its conceptualisation and construction do not rest on a model of the brain's functional organisation. The MMSE assesses the patient's orientation, language, reading, constructional abilities, recall, attention and calculation in a brief and simplistic manner (see Appendix F).

As is discussed later, this study (along with the experiences of our colleagues in Neurology) has found major weaknesses in using the MMSE in the South African context (see Chapter Two). The MMSE is not comprehensive enough and does not provide sufficient information relating to cognitive deficit, being able to demonstrate only very superficially, and in a confusing way, the presence or absence of some cognitive problems. Furthermore,

understanding what scores to use as reference points, and how these should be interpreted, is extremely difficult, especially in a multicultural setting. In addition, because the MMSE screening test is not theory driven (that is, not linked to, or structured around, a theoretical model about the functional organisation of the brain), it lacks clinical effectiveness when viewed from the hypothetico-deductive approach. The test results cannot be linked to specific syndromes or pathology, and are hence difficult to interpret meaningfully.

Given these shortcomings of the MMSE, the neurologists at Groote Schuur Hospital were looking for a more purposeful, comprehensive *screening* tool which, in the absence of neuropsychologists, they themselves could administer and interpret. They therefore approached the Neuropsychology division and asked if a better form of bedside screening tool could be designed to meet their specific purposes. Nell (2000) has described this type of adoption of expertise from specialists in a particular field to others lacking training or expertise in that particular domain as ‘transferable technology’; this has been more precisely defined as: “one that can effectively be devolved by fully trained personnel to those with a lesser level of training” (Holtzman, Evans, Kennedy & Iscoe, 1987, as cited in Nell, 2000, p. 108).

Transferable technology can not only be used clinically but also for educational purposes. The dire need for neuropsychological services and expertise in the South African context creates an ideal setting for such transfer to take place, yet to date this need has not been met. What was required was a cost-effective, yet comprehensive screening tool, which was quick and easy to administer, especially given the constraints on personnel and financial resources at Groote Schuur Hospital, as in South Africa in general. This was the basis for creation of the GSNSB Prototype (see Appendix A).

Rather than simply creating a larger version of the MMSE, the mindset behind the creation of this new screening battery was very different. What was envisaged was a test battery that would provide the neurologists not only with a comprehensive screening tool, but also with a range of tests that could eventually serve a purpose within a thorough hypothetico-deductive clinical assessment.

In light of this, a number of key considerations had to be taken into account in the construction of the GSNSB Prototype. What was particularly important was to select items with widely acknowledged clinical utility and reliability in terms of diagnosing organic brain dysfunction. Furthermore, it was essential to bear in mind the importance of assessing both general and specific cognitive functions in different functional domains when selecting items for inclusion (Black & Strub, 1994).

To achieve this objective the approach was to design a theory-driven battery, utilising neuropsychological knowledge of the functional architecture of the brain to both select the relevant tests and to structure the GSNSB into theoretically logical sections. Luria's theory of the functional organisation of the brain was used to develop a battery comprising sections representing the primary cognitive domains of the brain, based on clinico-anatomical correlates. Following on this systematic approach, a series of decision-trees was incorporated into the structure of the GSNSB, in order to detect the presence or absence of cognitive deficit in a specific region of the brain in a theoretically coherent and meaningful manner. The aim was that the GSNSB should be administered in a flexible, selective way, depending on the specific question about each patient's clinical presentation. Therefore, the whole tool would not necessarily need to be administered — the clinician would be able to administer

the sections relevant to the clinical question he/she wished to examine. In other words, this screening tool would combine the fixed and flexible approaches adopted in clinical assessment batteries. From the flexible approach, the assessor using the GSNSB can administer the sections of the GSNSB in various ways depending on the pathology he/she wishes to explore, and the hypothesis testing is built-in in the form of decision-trees. As with the fixed approach, the tests to be used for each domain of cognitive function are predetermined.

The 'scores' used in the GSNSB are not psychometric, but instead serve to record an answer to a yes/no question: is there presence or absence of given cognitive deficit in a specific domain of cognitive function? In clinical practice, this may be seen as a starting point for guiding the investigation and informing the final clinical diagnosis, which is the end product of a series of such answers to a hypothetico-deductively generated series of questions. This approach was incorporated into the design of the GSNSB's scoring system, using cut-off scores to 'quantify' the qualitative observations regarding test performance, and is similar to Grozman's (1999) work in developing Luria's assessment method, which adopted a six-point cut-off score scale.

An important distinction between Grozman's work on Luria's assessment method is that it involved a clinical assessment tool, not a screening tool. Additionally, with the GSNSB, the relative weighting of the cut-off scores is not uniform, with certain tests being worth more than others on account of their stronger validity and reliability for certain deficits. The goal behind this combined qualitative/quantitative approach was: "to evaluate quantitatively the magnitude of each symptom and the severity of disturbances in each functional area and to include a summarized score for the cognitive disturbances present in a given patient"

(Glozman, 1999, p. 24). Furthermore, the incorporation of these scores in the GSNSB allows for re-test comparisons to be made.

Because the GSNSB was created primarily to replace the MMSE, it was necessary to develop a scoring procedure that would serve the purpose of differentiating between the presence or absence of cognitive deficit (of a pathological condition versus no pathology) for each decision-tree procedure. To achieve this objective, the scores developed are allocated by grading the patients' performances on a range of simple neuropsychological tests, with more complex tests being incorporated in the GSNSB as 'optional' supplementary tests. As Hebben and Milberg (2002, p. 91) observe with reference to assessment batteries, as opposed to screening batteries, a "comprehensive test battery contains measures of both higher and lower cognitive domains in order to identify the point of processing at which functions break down. In addition, the clinician must assemble a test battery that permits assessment of the same cognitive domain with multiple measures to explore the reliability of the deficit". The above-mentioned grading is based upon interpreting the qualitative performances of the patient, depending on the specific test in use. Scores are allocated for determining either the presence or absence of a pathological condition/cognitive deficit, thereby answering a yes/no question.

There are numerous examples of this approach to scoring in the GSNSB Prototype and its later form (see Appendix A and/or Appendix B). The scoring system serves to determine the presence or absence of a clinically meaningful deficit within each functional domain of the brain — for guiding the cognitive assessment, and as a point of reference upon which a clinical diagnosis can be based. As Lezak (1995, p. 737) puts it, "[t]he examiner's experience and training provide the standards for evaluating much of the patient's responses and

behaviour...the data obtained [with this form of testing] are impressionistic and tend to be coarse-grained, compared with the fine scaling of psychometric tests". The structuring of the GSNSB in a theoretically coherent manner allows for its use in a hypothetico-deductive way.

The GSNSB may therefore be seen as constituting a 'middle path' between the hypothetico-deductive and normative approaches. This 'middle path' has been forged from the practice of the above-mentioned hypothetico-deductive qualitative approach. The GSNSB is only intended as a screening tool — it utilises scores, as the normative approach does, yet the allocation of these scores is based upon hypothetico-deductive reasoning, and the scores serve a different purpose (answering a yes/no question of whether cognitive deficit is present or absent) to that used in the normative approach.

The initial compilation of the GSNSB (the Prototype) was given to Groote Schuur's neurologists in 2003 — they were pleased with this initial attempt. At the time, Professor Mark Solms also presented a series of seminars to instruct the neurologists how to use the GSNSB Prototype, to provide basic neuropsychological knowledge on the domains of cognitive function in the brain, and to explain the qualitative features of some common pathological conditions. This basic knowledge is important for anyone using the GSNSB, and is key to effective 'transferable technology' (Nell, 2000).

Early Problems Identified with the Groote Schuur Neurocognitive Screening Battery

Prototype

All of the neurocognitive 'bedside' tests chosen for inclusion in the GSNSB Prototype were widely recognised and internationally established, and were being used on an ongoing basis in the daily clinical assessments of the neuropsychologists at Groote Schuur Hospital.

However, through clinical experience gained from working with a wide range of patients in the South African context, some major shortcomings relating to the cultural appropriateness of some of the tests and test items became evident (a thorough account of this cultural bias and inappropriateness is presented in Chapter Four). Specifically, many of the tests displayed inherent cultural bias, which rendered them clinically ineffective in the South African context.

A second problem identified with the GSNSB Prototype pertained to copyright issues, some of the tests being the intellectual property of the organisations that produced them. A third problem was that the GSNSB Prototype only existed in English, whereas large portions of the Western Cape population speak either Afrikaans or isiXhosa as first languages. A fourth, crucial problem was the fact that the GSNSB Prototype had not been validated, and hence its clinical effectiveness and reliability had yet to be formally demonstrated.

Changes Required to the Grootte Schuur Neurocognitive Screening Battery Prototype

The present study was designed to specifically address these four problems inherent in the GSNSB Prototype and to develop this much-needed clinical tool to a point where it could be meaningfully utilised. It was evident that the problematic tests would need to be replaced with versions adapted for the South African context. This process would require researching, implementing and piloting the necessary changes to eliminate the inherent cultural inappropriateness. It would also be necessary to translate the GSNSB Prototype into both Afrikaans and isiXhosa. Once these objectives had been satisfactorily implemented and piloted, it would be necessary to validate the GSNSB in its entirety (having first incorporated the new changes made to the tests). This final part of the process would necessitate demonstration of the validity and reliability of GSNSB, including its ability to distinguish

between the presence or absence of a pathological condition; to distinguish between the cognitive functions of the various anatomical structures of the brain; and its ability to do so consistently over repeated trials with different examiners.

The GSNSB always required validation, yet through extensive clinical experience at Groote Schuur Hospital (and the work that had already gone into creating the GSNSB Prototype), it had been realised that it was important to first ensure that the content of all the tests contained in the GSNSB took a form best suited to the assessment of South African patients. By refining and then validating a culturally appropriate, diagnostically meaningful screening tool, the success of the endeavour would have far-reaching implications for the emerging field of neuropsychology in South Africa, both in terms of clinical practice and research.

The first task to be accomplished with this study was therefore to develop culturally fair equivalents to the relevant tests in the GSNSB Prototype. Before the creation of the GSNSB Prototype, the changes that we had made to the test items had been implemented on an ad hoc basis within our neuropsychology division, using a trial-and-error approach during clinical assessments to overcome some of their cultural inappropriateness. This had involved informally changing various items in some of the existing tests while the tests were being administered. This process aimed at making the tests more culturally fair and appropriate, while at the same time serving to clarify the clinician's interpretations of the test performances. Working in this way had proved effective and the author and his colleagues had already begun to get a feel for what would work and what would not. Formal changes, however, were required in order to validate the GSNSB.

The test adaptations would need to be designed to retain the initial scoring structure and the core thematic components of the original tests, while the content would need to be changed in an attempt to make it culturally relevant for the South African context. New visual tests would have to be created, using an artist to complete the final versions, while new verbal tests would first be required in English, before experienced translators could be consulted to develop Afrikaans and isiXhosa versions. All test changes would require evaluation by a panel of professionals including neuropsychologists, translators and back-translators, and interpreters and cultural experts. This would ensure that the GSNSB contained the most clinically effective (for South Africa) neuropsychological tests when the validation process began.

Aim of the Study

The aim of this study was to address the key problems identified with the initial GSNSB in order to advance the development of this screening tool. The study had five primary objectives:

- i) To investigate the efficacy of the Mini-Mental State Examination in the South African context (Chapter Two).
- ii) To accurately translate the GSNSB into both Afrikaans and isiXhosa (Chapter Three).
- iii) To develop the nine new neurocognitive tests (i.e. new versions of existing tests) required to address the inherent cultural and educational bias of some of the clinically ineffective tests included in the initial draft of the GSNSB (Chapters Four and Five).
- iv) To demonstrate the reliability and validity of the GSNSB as a whole (Chapter Six).
- v) To implement further improvements in the design of the GSNSB based on the outcomes of the reliability and validity study (Chapter Seven).

CHAPTER TWO: EVALUATING THE EFFICACY OF THE MINI-MENTAL STATE EXAMINATION IN THE SOUTH AFRICAN CONTEXT

The Mini-Mental State Examination (originally called the ‘Mini-Mental State’ and known in short as the MMSE) is a short, 11-item screening tool designed by Folstein, Folstein and McHuge, made available in 1975 (see Appendix F). The tool was created to offer a quick and easy-to-administer cognitive screening battery as an alternative to the lengthy tests currently in use at this time, such as the standard WAIS (Wechsler Adult Intelligence Scale) and Withers and Hinton’s test. The MMSE was initially intended to aid in the differential diagnosis of psychiatric and neurological patients (Lezak et al., 2004). Folstein et al. originally investigated their new test’s reliability and validity on a sample of 206 patients with dementia syndromes, affective disorder, schizophrenia, affective disorder with cognitive impairment (‘pseudodementia’), personality disorders and mania, as well as on 63 normal controls. The results showed that none of the control participants scored below 24. This particular study found the MMSE to be both valid and reliable in differentiating between patients with dementia, depression with cognitive impairment, and depression (Folstein et al., 1975). The study also found the patients in the sample with schizophrenia, dementia, delirium and affective disorder to consistently score below a key score of 20 out of 30. The authors claimed the tool was thorough in assessing the cognitive domain of mental functions, and able to qualitatively estimate the severity of cognitive impairment (Folstein et al., 1975).

The MMSE assesses a limited range of cognitive functions, is scored out of 30 points, and takes just five to ten minutes to administer (Lezak et al., 2004). Specifically, the test’s items cover the following cognitive functions simply and quickly: orientation, memory, attention

and calculation, language (naming, repetition, comprehension, reading and writing), and visuospatial construction. This breakdown of functions can be divided into two sections. The first of these covers orientation, memory and attention and calculation, and requires only verbal responses from the patient. The second section, encompassing the language and construction tests, requires writing, drawing, and verbal responses at various points (Simard, 1998).

Folstein et al. (1975) designed the MMSE for questions relating to the cognitive aspects of mental function rather than those concerning mood and form of thinking. Spreen and Strauss (1991, p. 17) also highlight how, “[w]hen adults generate very few responses on the Wechsler Adult Intelligence Scale (WAIS), making it difficult to rank them according to the extent of their cognitive deficits, tests such as the Mattis Dementia Rating Scale (Mattis, 1976) or the Mini-Mental State Exam ... may be preferred in order to provide gross estimates of cognitive functioning”. Of these two tools, the former offers a larger range of tests than the MMSE (Spreen & Strauss, 1991). However, the Dementia Rating Scale, and batteries such as the CAMCOG, are not suitable for ‘bedside’ clinical practice (Nestor & Hodges, 2003). Initially designed to offer a brief standardised assessment of mental status in psychiatric patients, the MMSE was intended to help distinguish functional disorders from organic ones. In more recent times, the MMSE has mainly been used in the context of the cognitive deficits associated with neurodegenerative disorders, in detecting them, and recording their progression (Salmon & Hodges, 2003).

The MMSE is the most widely used screening tool for dementia, and is used both on its own or as a component of other assessment protocols, such as the CERAD (Consortium to Establish a Registry for Alzheimer’s Disease) Battery (Lezak et al., 2004). In addition to its

broad clinical use, the MMSE is also widely used for research purposes, especially in dementia research, as it is well known, is an easy tool to administer, and lends itself well to use in comparing patient cohorts across studies (Salmon & Hodges, 2003).

The central purpose behind a screening tool such as the MMSE is to identify a person who falls below a certain cut-off score, who is then considered a 'subject' (Harvey, 2003). The key cut-off score for the MMSE is usually 24, with a performance below this cut-off being an indication of cognitive impairment or dementia (Mitrushina et al., 2005). The choice of the cut-off score of 24 was based on the original Folstein et al. (1975) validation and reliability study, which revealed that none of the 63 elderly control participants investigated achieved a score lower than this. There is some contention in the literature as to what exactly the optimum cut-off values should be, with many studies arguing various levels of test specificity and sensitivity at different values, in different clinical populations. In clinical practice at Groote Schuur Hospital, for example, the neurologists work primarily from a cut-off score of 23.

Some authors (see section below) have examined the test characteristics of specificity and sensitivity and have proposed using cut-off scores of between 25 and 27 instead, as a way of increasing both these characteristics (Salmon & Hodges, 2003). For example, a study conducted by Monsch, Foldi, Ermini-Funfschilling and Berres (1995) investigating the optimal cut-off score for the MMSE in detecting dementia, using 70 patients with diagnosed dementia of the Alzheimer's Type (DAT) and 50 normal controls, concluded that the optimal cut-off score for detection was 26, with 74 percent sensitivity and 100 percent specificity at this level. Studies by Feher and Martin (1992), and Kukull, Larson and Teri (1994) both reported that the MMSE displayed good specificity, but reduced sensitivity. The former

study's results produced a specificity value of 0.9 and a sensitivity value of 0.69 (Feher & Martin, 1992). Similarly, the Kukull et al. (1994) study's results found a specificity value of 0.96 and a sensitivity value of 0.63. The cut-off value of 24, however, has become the established standard cut-off score for the MMSE, and was therefore adopted in the present research.

Mini-Mental State Examination test scores have also been found in various studies (see section below) to be strongly influenced by education level and age, with lower scores occurring as age increases, and higher scores occurring as education level increases (Lezak et al., 2004). To address the issues of the influence of education and age on MMSE performance, Crum, Anthony, Bassett and Folstein (1993) designed a study as a way of establishing population-based norms for the test. Assessing 18 056 adult participants in America using the MMSE, this study found that both age and education impacted on performance (Crum et al., 1993). As a guideline for clinicians, the results from this study were compiled into an 'Age and Education Weighted Norms Table for Mini-Mental State Examination' (see Appendix G). This table incorporates a summary of scores for specific age and education levels, in mean, median and percentile distributions, allowing the clinician to look up a predicted normal score for the specific age and level of education of his/her patient. The authors cautioned in the study that the MMSE should be used only to identify current cognitive problems and not as a diagnostic tool (Crum et al., 1993).

A similar, more recent study, conducted by Bravo and Hebert (1997) in a Canadian population, also established useful age and education reference values for the MMSE. This study also determined that age and education level significantly influenced performance on the test. Based on a sample of 7 754 elderly normal controls aged over 65, the results showed

that, for the five age groups and four levels of education, the test scores decreased with age and increased with level of education (Bravo & Hebert, 1997).

For the purposes of the present study, the Crum et al. (1993) norms were used, as they are more comprehensive, covering all age-groups, as opposed to the Bravo and Herbert (1997) norms, which only comprise norms for individuals aged 65 and over.

Studies Using the Mini-Mental State Examination

The MMSE has been widely used and cited in the field of medicine. A survey conducted by Nilsson (2007), which examined the number of citations of the MMSE in commercially available databases, found the MMSE to be one of the most cited papers ever in the field of medicine, with 19 721 citations between 1977 and 2007. The test has been widely used as an indicator of severity in dementia research, especially Alzheimer's disease, and has also been extensively used with psychiatric patients. Nathan, Wilkinson, Stammers and Low (2001) highlight the fact that the MMSE is accurately able to discern the presence of dementia in its early stages. Banos and Franklin (2002) attest to the usefulness of the MMSE in the context of ascertaining mental status simply, provided that the test is limited to this role and not used for diagnostic purposes. The test has also been used in a number of studies investigating Alzheimer's disease and other dementias. For example, a recent study conducted by Park, Pavlik, Rountree, Darby and Doody (2007) incorporated the MMSE — which was proved to be accurate in identifying Alzheimer's disease — when investigating the question of whether functional decline is necessary for a diagnosis of mild Alzheimer's disease. The results suggested that it is not necessary.

A recent study by Kraybill, Larson, Tsuang, Teri, McCormick, Bowen, Kukull, Leverenz, and Cherrier (2005) utilised the MMSE in conjunction with a range of other tests in the investigation of the neuropsychological profiles of dementia patients with Alzheimer's disease, Lewy Body pathology or a coexistence of the two. The results specific to the MMSE showed it is effective in identifying dementia, and that a reduction in test performance over time is greatest with the patients with Alzheimer's disease and Lewy Body's in comparison with other tests (Kraybill et al., 2005).

Limitations of the Mini-Mental State Examination

The general findings of studies examining the limitations of the MMSE are encapsulated by Tombaugh, McDowell, Kristjansson and Hubley (1996, p. 48):

Tombaugh and McIntyre (1992) concluded that the MMSE possessed moderate to high reliability coefficients, demonstrated high levels of sensitivity for cognitive deficits in patients suffering from moderate to severe Alzheimer's disease, and reflected the cognitive decline typical of dementia patients. Criticisms of the MMSE included (a) its failure to discriminate between people with mild dementia and those who are not demented, (b) a limited ability to detect impairment caused by focal lesions, particularly those in the right hemisphere, (c) overly simple language items that reduce sensitivity to mild linguistic deficits, and (d) a large number of false-positive errors because of its bias against individuals with low education.

Numerous authors have found specific weaknesses with the MMSE. Tombaugh and McIntyre (1992) found that the MMSE is extremely sensitive to level of education and age. Further studies have shown how MMSE performances are affected by ethnicity, culture and level of acculturation. For example, in a study examining the cultural relevance of the MMSE in 140 American Indians, Jervis, Beals, Fickenscher, and Arciniegas (2007) recently found that 11

percent of participants performed two or more standard deviations below the predicted norms for the test. A study by Espino, Lichtenstein, Palmer and Hazuda (2001) examined a sample of 827 Mexican Americans and European Americans aged 65 and over in a population-based study to ascertain differences in correlates on the MMSE. The results showed that the Mexican Americans were more than twice as likely to have an MMSE score below 24, relative to the European Americans (Espino et al., 2001).

Additionally, numerous authors have highlighted problems concerning the clinical limitations of the MMSE. Specifically, the MMSE has been viewed as being poor both in detecting mild forms of dementia and in differentiating between different types of dementia. The MMSE has also been accused of lacking tests of executive function. For example, Juby, Tench and Baker (2002) examined the charts of 68 patients who had been tested using the executive interview and clock drawing test for executive dysfunction. Their results revealed that 32 patients achieved a normal MMSE score despite 22 having been diagnosed with executive impairment according to the executive interview. Simard (1998) also highlighted how the MMSE lacks tests of executive function, citing as key omissions tests of ability to abstract or judge social situations.

Simard (1998) also added how the MMSE often lacks the sensitivity to detect the early signs of dementia, which can result in 'ceiling effects'. Documented weaknesses of the MMSE include the ceiling effects it produces, the limited number of cognitive abilities it assesses, the limited number of test items available, and its limited range of scores per item (Rapp, Espeland, Hogan, Jones & Dugan, 2003). In describing the clinical use of the MMSE, Nestor and Hodges (2003, p. 33), comment that, based on clinical experiences: "as the test is heavily weighted toward verbal tasks, a patient with significant disturbance of right hemisphere

function may well score 29/30. Likewise, patients with significant frontal pathology frequently score flawlessly; even patients with early Alzheimer's disease may score 27/30 (only failing the recall items) and yet show profound deficits on specific memory tests. The simplicity of the language tasks and the lack of executive components make it particularly insensitive to frontotemporal dementia”.

In 2005, Lopez, Charter, Mostafavi, Nibut and Smith conducted a study examining the psychometric properties of the MMSE on 412 elderly medical patients and found that many of the items were too easy. Finally, a study by Feher, Mahurin, Doody and Cooke (1992) that examined the ‘subtests’ of the MMSE in a sample of 76 neuropsychology referred patients found that four out of the five language items were insensitive to impairment, while three out of the five items did not correlate significantly when compared to scores from neuropsychological tests.

The Modified Mini-Mental State 3MS

The *Modified Mini-Mental State* (3MS) was produced by Teng and Chui (1987) in an attempt to rectify some of the above-mentioned problems with the MMSE. Specifically, four additional items were included in the original MMSE and the score procedure was modified from a total out of 30 to a total score allocation of 100. The administration order of the tests was also altered (Lezak et al., 2004). “The 3MS is designed to sample a broader variety of cognitive functions, cover a wide range of difficulty levels, and enhance reliability and validity. The 3MS retains the brevity, ease of administration, and objective scoring of the MMS but broadens the range of possible scores” (Teng & Chui, 1987, p. 314). The four items added included verbal fluency, similarities, and recognition and cued recall (Teng & Chui, 1987).

Only a relatively small number of studies have examined the 3MS in relation to the MMSE. Among these, Teng, Chui and Gong (1990) found that the 3MS displayed more sensitivity and specificity towards dementia, and was more reliable than the MMSE. A study conducted by Tombaugh et al. (1996) used a sample of 525 community-dwelling individuals in order to compare the MMSE and the 3MS psychometrically, while at the same time deriving norms for the tests. The study's results indicated that although the 3MS and the MMSE were comparable in terms of their performance, the verbal fluency test was a good addition to the 3MS in terms of its sensitivity (Tombaugh et al., 1996). Finally, a revision to the 3MS, the 3MS-R, was recently devised with officially published norms (Lezak et al., 2004). To the author's knowledge, there is no literature available on the use of the 3MS in South Africa.

Rationale

The aim of this aspect of the research was to evaluate the efficacy of the MMSE screening tool in the South African clinical context. Given that the GSNSB had been designed to replace the MMSE, it was necessary to formally investigate the efficacy of the MMSE in the light of its apparent inability as a meaningful diagnostic screening tool. The identification of this need for a formal investigation stemmed, in the main, from two sources, both of which served to inform this present research. The first of these sources was the Groote Schuur Hospital neurologists, whose loss of clinical confidence in the MMSE had led them to request the neuropsychologists to develop an alternative screening tool. This loss of clinical confidence was a consequence of many years of experiencing first hand the diagnostic limitations of the MMSE in the South African context.

The second source was the neuropsychology team at Groote Schuur Hospital, whose clinical experiences in taking detailed patient histories, reviewing patient MMSE performances, and

diagnosing neurocognitive deficits, had caused them, too, to doubt the accuracy and validity of MMSE scores. This stemmed from three frequently occurring scenarios with our referrals, especially those referrals that queried dementia. The first of these was that, although many patients obtained a good MMSE score, a formal neurocognitive assessment would later find them to have demonstrable cognitive impairment. Alternatively, patients would perform poorly on the MMSE despite the fact that no neurocognitive deficit could later be formally demonstrated. A third scenario was that the patient would perform poorly on the MMSE but the particular items that they failed were not indicative of any particular picture of neurocognitive deficit. Consequently, the Groote Schuur Hospital neuropsychologists were also in support of a revised approach to neurocognitive screening — one that would prove unique in the South African context.

The importance of having a diagnostically valid screening tool cannot be overemphasised. Without accurate diagnostic tools, the provision of a proper clinical service to patients is severely undermined. Given that screening is most often the first form of cognitive assessment received by a patient upon admission to the hospital, if this screening is devoid of accuracy, the patient may never go on to receive the specialist treatment they require, and neurocognitive problems may go undetected. If accurate neurocognitive diagnoses are not made, the negative consequences to the patient and his/her family in terms of employment, relationships, coping strategies and finances can be severe. In the absence of widespread specialist neuropsychological expertise, a screening tool is required that allows a non-specialist to obtain meaningful insights into neurocognitive deficits, thereby bridging the void in the absence of a specialist. Such a tool can be referred to as ‘transferable technology’ (Nell, 2000).

Methodology

Sample

A sample of ten patients was drawn upon for this aspect of the research. These patients were all archived referrals from the Neuropsychology division at Groote Schuur Hospital, and had therefore previously undergone full neurocognitive assessments by the neuropsychologists (that is, a neuropsychological diagnosis had been made). All these patients had completed an MMSE prior to their referral. Given that the MMSE is primarily a dementia screening tool, all referrals included in this sample were for queried dementia.

The retrospective sampling from the archives used a random sampling procedure. This involved including the first ten patients found to have both a complete record of the neuropsychological assessment, along with a breakdown of which MMSE items the patient had failed, and a total MMSE score.

Materials

Along with the MMSE (see Appendix F), the Age and Education Weighted Norms Table for the Mini-Mental State Examination (see Appendix G) was also consulted (Crum et al., 1993). This table offers population-based norms for the MMSE, categorised by age and education levels, and was used to predict what the patients should have scored for their respective ages and levels of education. The other materials used included each patient's hospital folder containing his/her complete medical records; the tests used by the neuropsychologists in their clinical assessments of these patients; and the reports written by the neuropsychologist following assessment of each patient. The neuropsychologists' reports were housed in the Neuropsychology division, while the patients' medical folders were obtained from the hospital archives.

The assessments conducted by the neuropsychologists included the investigation of memory function, executive function, spatial cognition and language function. The tests used included the Rey Complex Figure, the Babcock Story and the 4 Hidden Objects Test for the testing of memory. Testing of executive function involved the following tests: the 18 Book Problem, the COWAT (or FAS Test), the Fist/Side/Palm Test, the Red/Green Test, the Babcock Story, the Rhythm Tapping Test, the 20 Questions Test, Proverbs and Cognitive Estimation. The Digit Span Test was used to test working memory.

Tests used for spatial cognition included the Cube Analysis Test, the Rey Complex Figure, and the Scene Drawing Test, while anosognosia was assessed through verbal questioning. Finally, the tests used to assess language function included the Boston Naming Test and the Token Test. Language repetition was assessed using sentences of increasing length, while the patients' ability to read and write was also formally assessed.

All of the above-mentioned tests were used as part of the investigation into the efficacy of the MMSE as a screening tool. Altogether, this comprehensive array of tests greatly facilitated in the qualitative analysis of the MMSE results.

Design

A case study design was used to thoroughly investigate the perceived shortcomings of the MMSE as identified initially by the resident neurologists and neuropsychologists. This design involved the qualitative analysis of each case. Using the MMSE score as a departure point, the analysis involved comparing each patient's MMSE score and its breakdown with the clinical diagnosis subsequently made by the neuropsychologists, thereby adopting the

hypothetico-deductive line of clinical reasoning and investigation to explore different aspects of the MMSE. The primary advantage of this design was that it allowed for the richest possible qualitative data to be gathered by using a comprehensive form of clinical assessment to critique a screening approach which, by definition, is not as exhaustive.

Data Analysis

To qualitatively analyse the data, each patient's MMSE score, once obtained, was broken down to identify on which specific items, if any, points had been lost. Once these items had been recorded, the diagnostic outcome of the neuropsychologists' assessment of the patient was consulted for comparison with the MMSE performance. At this point, the patient's predicted MMSE score for his/her age and level of education was also calculated using the Age and Education Weighted Norms Table.

Once these three key facets of the data had been gathered for each of the 10 cases, the qualitative analysis involved investigation of one of four possible scenarios. Firstly, the patient had scored below his/her predicted score on the MMSE for his/her age and level of education, but in areas of the test inconsistent with the areas of primary deficit established by the neuropsychologists. Secondly, the patient had achieved a normal predicted MMSE score for his/her age and level of education, when the neuropsychologist had actually demonstrated neurocognitive deficits. Thirdly, the patient had scored below his/her predicted score on the MMSE for his/her age and level of education, when the neuropsychologists had found him/her to be neurocognitively intact. Fourthly, the patient's MMSE performance for his/her age and level of education was consistent with the findings of the neuropsychologists. In order to shed light on these possible scenarios, the exact line of clinical inquiry (clinical reasoning) that the neuropsychologist had followed in reaching a final diagnosis (including

his/her selective use of tests, and the theoretical rationale behind this usage) was closely scrutinised.

This qualitative analysis was used with two primary objectives. The first of these was to identify accurately exactly which aspects of the MMSE, if any, were responsible for producing misleading and inaccurate diagnoses. The second objective was to establish, from a hypothetico-deductive perspective, exactly why these specific aspects had been problematic. The results of this detailed analysis later served to inform and direct the development of the GSNSB, to avoid potential pitfalls and build in alternative decision-tree structures.

Procedure

The patients' records in the Neuropsychology division's archives were examined on the Groote Schuur Hospital premises and none were removed at any stage. This ensured that the records remained secure and confidential and that none were lost. These archived patient records contained the referral letter with the referral question, a summary of the patient's past medical history, and a full breakdown of the neuropsychologists' assessment, including which tests were used, and their conclusion and clinical impression of the patient. The patients' hospital folders containing all their general medical records were ordered from the hospital's computer system and were delivered to the office, following the same procedure that occurs when patients' folders are requested for referral purposes. The information extracted from these hospital folders included the patients' background medical histories, their scan reports and their prescribed medications, etc.

The study adopted the usual procedure when using the Age and Education Weighted Norms Table (see Appendix G). The patient's 'Age' category is chosen from the left-hand column of the table, read from top to bottom. The age categories are provided in four-year intervals, starting from the age of 18 up to the age of 85 and over, e.g. Age: 18–24, 25–29, 30–34, 35–39, etc. Once the relevant age of the specific patient has been selected, the table is then read across from left to right, to find the appropriate educational level for that specific patient. The 'Education' categories are also provided in intervals, starting with zero to four years of education, through to 13 or more years (which equates to a tertiary education), that is, Education level: 0–4, 5–8, 9–12, 13 or more. Once the patient's 'age category' and 'educational level' have been cross referenced on the table, a corresponding mean (out of 30) is provided, which is the population-based norm, e.g. the mean equals 27 for a 18–24 year old with 5–8 years of education (Crum et al., 1993).

Results

For each case presented below, the diagnosis made by the neuropsychologists is provided in conjunction with the summary of the patient's medical history. A detailed outline of the neuropsychologists' clinical investigation — including the neurocognitive tests used and the clinical reasoning underlying their analysis — is also given. Finally, the result of the qualitative investigation of the patient's MMSE score is reported.

Case 1: Mr S

Diagnosis and Case History

Mr S, a 67 year old with a grade 8 education, had been referred to the neuropsychologists for queried dementia/the possibility of a psychiatric condition. He was subsequently diagnosed with Lewy Body Dementia by the neuropsychologists. He scored 30 out of 30 on the MMSE,

while the mean score for his age and level of education was 26 according to the Age and Education Weighted Norms Table (Crum et al., 1993).

Prior to his referral, Mr S had an 11-month history of abnormal behaviour, memory problems, and visual and gustatory hallucinations. The first episode comprised five days of visual hallucinations. The family reported odd behaviour, nonsensical speech, frequent falling, postural dizziness and gustatory hallucinations. He described this episode clearly, including several complex hallucinations, which were detailed and dreamlike, but odd and bizarre — the hallucinations were not upsetting to him, nor did they seem to bother him. In total, four such episodes were reported in the first year. Mr S realised that what he had seen could not really have happened. He was unable to recall the next three episodes, but his wife's account described a hallucinatory cousin visiting him, constant falls, speaking nonsense and slurred speech. When an episode occurred, Mr S reported strong gustatory hallucinations involving bitter taste, a thick sensation in his tongue, and seeing strange paint on the wall. On one occasion his wife noted body jerking and urinary incontinence. Moderate alcohol usage was reported, as was stress when his wife became severely ill.

On formal assessment, Mr S was orientated to person, place and time and was able to hold a meaningful discussion regarding the upcoming elections — consequently, he was considered lucid and not delirious. When tested using the Boston Naming Test, no naming difficulties were demonstrated. Given that the reported hallucinations might have been a consequence of problems with higher visual functions, the Boston Naming Test was also used to test visual problems, but no higher visual problems were detected. The Rey Complex Figure was used to test visuospatial construction ability, which was found to be intact, ruling out constructional apraxia. Mr S's recall of the Rey Complex Figure was also good, so difficulties with visual

memory were also excluded. He was also found to have excellent verbal memory on the Babcock Story with both his immediate and delayed recalls. The neuropsychologists thus concluded that Mr S did not have an axial amnesia.

Mr S's executive functioning and abnormal behaviour were also investigated, in light of the fact that certain possible causes of his hallucinations had already been excluded. A number of executive difficulties were immediately demonstrated, including poor performances on the Colour Word Interference Test, the Trails 'A' Test, which he took a long time to complete, and the Digit Span Test, where a digit span of four was all that could be achieved. Mr S presented with concrete thought, which was clearly demonstrated with the 20 Questions and Proverbs tests. Mr S coped with simple concrete problems, category fluency, and the tapping task, but was unable to understand the 18 Book Problem. On the FAS Test for testing generativity, he was very slow, only scoring 11 in total.

The neuropsychologists decided that Mr S's hallucinations were a consequence of organic hallucinosis given that they were visual, detailed, bizarre and had not troubled him. The combination of neurocognitive executive dysfunction and organic hallucinations was consistent with the clinical presentation of Lewy Body Dementia.

Qualitative Analysis of the MMSE

When analysing the MMSE in light of Mr S's case, a number of interesting findings emerged. Firstly, the MMSE was unable to detect the executive dysfunction evident in the neuropsychologists' neurocognitive assessment, primarily because it lacks items testing the dorsolateral frontal convexity (that is, tasks such as, or equivalent to, the 20 Questions Test, Proverbs, 18 Book Problem, or the Fist/Side/Palm Test, which test for difficulties such as

planning problems, set shifting problems, and problems with abstract thought). Additionally, working memory is not adequately tested in the MMSE. This is because the 'Registration' task (Item 3) asks for the repetition of only three items, which is not taxing enough to test the average of seven units for an intact working memory. Another concern is that the MMSE lacks any timed items, so it cannot be taken into account how long a patient takes to respond on the tests. This was why Mr S's problems with generativity were missed. However, tests like the FAS have a time specification of one minute per letter, allowing the examiner to record such problems.

Because the MMSE lacks tests to elicit abstract thought processes, it was unable to suggest that Mr S had concrete ideation. The MMSE also failed to detect Mr S's obvious attention problem. Clearly the MMSE lacks sufficient detail, as the item for attention (Item 4) is no more a test for attention than any of the other tests. Overall, this analysis revealed that the MMSE was unable to elicit any of the central neurocognitive features of this case's Lewy Body Dementia, evident from his score of 30 out of 30.

Case 2: Mrs O

Diagnosis and Case History

Mrs O, a 58 year old with a tertiary education, was referred to the neuropsychologists as she complained of memory loss for words and phrases. She was subsequently diagnosed with semantic dementia. She scored 30 out of 30 on the MMSE, while the mean score given by the Age and Education Weighted Norms Table for her age and level of education was 29 (Crum et al., 1993). Mrs O reported that her memory loss for words and phrases had occurred for the past year, with deterioration in vocabulary and comprehension being noted, along with frequent inappropriate responses, repetitions, and an inability to follow logical arguments.

Mrs O experienced difficulties with verbs and sentence construction, and increasingly avoided situations she had previously sought, reporting feelings of depression over the loss of her linguistic ability. It was recorded that she got excited over simple things and consequently emitted childlike responses. The neuropsychologists also noted that she admitted to significant alcohol use, as well as to becoming lost in the suburb where she worked, rationalising this by saying that she did not know Cape Town well. She also tried to explain away her lack of knowledge of current events.

Formal neurocognitive assessment of Mrs O had begun with the assessment of her language comprehension. History taking had revealed a number of comprehension difficulties, especially with lengthy instructions and multi-staged questions. On occasion, she even struggled to answer simpler questions, answering off the topic. It was also noted that she had word-finding difficulties, circumlocution, and paraphasias in her spontaneous speech, and that she sometimes used inappropriate word substitutions. When using the Token Test to test her verbal comprehension, her comprehension improved as the task was made less complicated by requiring her to respond verbally. Mrs O performed better on all the tasks requiring nonverbal responses.

Mrs O was orientated to person, place and time. She did, however, display a problem with current events. When confrontation naming was assessed using the Boston Naming Test, severe anomia was found, with a striking loss of semantic concepts, demonstrated when she could not describe the use of objects depicted in the test. On further assessment of language function, she performed normally on both repetition and reading tasks, but her writing was found to be impaired (although better than her spontaneous speech). Testing of memory

found Mrs O's axial memory intact, but her digit span was five, which indicated a mild deficit of verbal working memory. The testing also found her visuospatial ability preserved and she was not found to be agnosic.

Qualitative Analysis of the MMSE

Firstly, Mrs O's naming problems were not detected by the MMSE. Analysis of this finding showed that it stemmed from the lack of complexity in the naming task (Item 6), which requires the patient to name but two items, 'pencil' and a 'watch'. Both these items are very common and familiar and, in addition, far more items of varying complexity would be required to adequately demonstrate problems with naming. Item 9 requires the patient to follow a three-stage command without a verbal response being required, as does the Token Test used by the neuropsychologists. However, the MMSE has no proper comprehension task requiring verbal response. The Serial Sevens Test (Item 4), like many of the other test items, requires a verbal response, but it is not a primary language task and is moreover listed under the 'Attention and Calculation' item, not under 'Language'.

Mrs O's writing problem was also missed by the MMSE, despite the fact that writing is tested in Item 10. This seems to have resulted from the fact that no specific requirement for sentence length is given; if the patient only produces a short sentence, a writing impairment can go unnoticed due to the fact that the more complex the task, the more likely it is that a deficit will be elicited. Finally, Mrs O passed three of the language tests on the MMSE which, according to the findings of the neuropsychologists, she should have failed. The clear lack of sensitivity and detail present in the majority of the MMSE's items meant that, despite almost half of its items being dedicated to the assessment of language (five primary language

tasks out of 11 items in total), the severe language difficulties present in this case went undetected.

Case 3: Mrs P

Diagnosis and Case History

Mrs P was a 47 year old with 12 years of education. She had been referred to the neuropsychologists for queried depression, memory problems and possible dementia, and had been diagnosed with fronto-temporal dementia. She scored 30 out of 30 on the MMSE, while the mean score for her age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993). Psychiatric problems were present in her family history, and Mrs P herself had a long-standing history of psychiatric problems and alcohol abuse. She had experienced one anxiety attack, at the age of 22. A CT head scan showed major frontal atrophy, but she had previously been thought to be psychotic. She was diagnosed with major depression and a personality disorder, and had had a steady deterioration over three years. She was put on antidepressants. Mrs P became very religious and had paranoid delusions. She also experienced problems with activities of daily living, refusing to clean or cook, and wanted to live in the hospital ward.

Assessment of Mrs P found her to be appropriate in her behaviour, and no thought disorder or psychotic features were evident. She denied having experienced any delusions and hallucinations, but complained of severe headaches. The three-year history of deterioration described included radical personality changes, reduced personal hygiene, and spending the majority of the day (up to 16 hours) in bed. It also included her failing to do any washing, and walking many kilometres until her feet blistered. It was also discovered that she experienced severe anxiety when her husband went out, which resulted in her walking aimlessly. She was

previously well-presented, sociable, and talkative. Mrs P only really spoke when spoken to, and described herself as depressed because she never wanted to do anything.

Formal neurocognitive testing found Mrs P to be orientated to person, place and time. Her language was reduced and she had some difficulty with finding words in spontaneous speech. On confrontation naming using the Boston Naming Test, she was found to be mildly anomic. On testing of executive function, a problem with planning was noted; her copy of the Rey Complex Figure was mildly impaired, as she failed to draw the overall shape, despite paying careful attention, noting mistakes and repeatedly starting it over. Her immediate and delayed recalls were consistent with her copy. Her digit span was recorded as being six. Her performances on the Fist/Side/Palm Test, the Stroop Test, and the FAS Test were unremarkable. Mrs P's 20-Questions Test performance was poor and she presented as very concrete. On the trail-making task her performance was slow for both trials. Overall, the neuropsychologists observed that Mrs P made a determined effort to follow the rules on all tasks.

The neuropsychologists' diagnosis of Fronto-temporal dementia was made on the basis of her deficits in the executive sphere, while her memory and other cognitive functions remained intact. Mrs P's history, along with the reported deterioration, supported this diagnosis. The executive impairments demonstrated were consistent with a mesial-frontal cognitive picture rather than an orbital/basal one — these impairments being her problems with generativity and planning, her striking lack of ability with abstract thought, and her flattened affect and inert movement.

Qualitative Analysis of the MMSE

On analysis, the MMSE failed to detect Mrs P's cognitive impairment — evident in her score of 30 out of 30 — for a number of key reasons. Firstly, as with the previous case, the MMSE missed Mrs P's naming problem on account of its lack of detail and sensitivity, in particular Item 6. The MMSE also missed Mrs P's obvious lack of ability to think abstractly, evident from her concrete thought processes. This is due to the fact that the MMSE lacks equivalent test items to the 'Proverbs', '20 Questions' or 'Similarities' Tests, which are essential in detecting concrete thinking. Mrs P's other executive problems with generativity and planning were also missed by the MMSE because it does not contain timed items that require the patient to self-generate his/her own concepts and ideas within a specified time. Furthermore, the MMSE does not contain items that require planning as a central component.

Case 4: Mr M

Diagnosis and Case History

Mr M, aged 44 with a grade 10 education, was referred to the neuropsychologists with diagnosed autosomal dominant spastic tetra-paresis, and was described as having profound cognitive dysfunction. A frontal-type dementia was therefore queried. Following neurocognitive assessment, Mr M was diagnosed as having a Depressive Pseudodementia (that is, severe depression accounted for his poor performance). This diagnosis was made, even though he showed poor executive function with respect to memory, motor sequencing and generativity. He scored 23 out of 30 on the MMSE, while the mean score for his age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993).

It was discovered that Mr M was often home alone, left to thinking about his life and the past. As a result, he often got emotional, becoming distracted and forgetful, and experiencing severe headaches. He had not felt as though he was depressed, despite admitting he was 'sad' and 'lonely'. Mr M reported a tragic history: his mother had died when he was young; he struggled to talk about this, stating that it was 'extremely personal'. In addition, one of his brothers had been stabbed and the other had died from meningitis, as had his sister. As a result of these experiences, Mr M frequently worried.

Mr M's medical history revealed that he was first admitted to Groote Schuur Hospital in 2001. At the time, he reported having weak feet and knees, and stiffness in his right leg, which was progressing, requiring the use of crutches to walk.

Neurological assessment at the time demonstrated weakness, frequent falls, incontinence, gait difficulties, frontal lobe disinhibition and cognitive dysfunction. A further investigation, using a CT head scan, showed hypodensity in his right corona radiata. In 2005 Mr M was given a disability grant. Around this time, further symptoms developed and he began slurring his speech and speaking slowly, often repeating himself, while his overall neurological condition was reported to be deteriorating further. The frontal signs he had included: perseveration, labile mood and pout reflex, and a seven-year history of progressive weakness of spasticity of the lower limb.

Formal neurocognitive assessment of Mr M showed he was orientated to person, place and time, and had a digit span of five. Assessment of memory, using the 4-Hidden Objects Test, found his immediate recall to be good, but his delayed recall was poor. He had, however, benefitted from prompting (a sign of executive impairment). When given the Babcock Story,

Mr M's performance was poor, although he benefitted from prompting, and there was no confabulation. The neuropsychologists thus concluded that with regard to executive function, his mesial functioning was intact. Further testing using the Rey Complex Figure was halted, as Mr M could not manage the complexity of the task, given the poor strategy he had used. As an alternative, the neuropsychologists gave him a simplified version of the figure, for which he produced a good copy. However, his memory performance with both the immediate and delayed recalls of this simplified figure was poor. For the remaining tests administered to him, Mr M did poorly on the Fist/Side/Palm Test, although this might have been accounted for by the spasticity in his hand. He did well on the Tapping Test, the Red/Green Test, as well as on the calculation and estimation tasks given to him. His 18-Book Problem answer was concrete. Finally, on confrontation naming, he performed normally when administered the Boston Naming Test.

Qualitative Analysis of the MMSE

When examining the MMSE in the light of Mr M's case, a number of discrepancies were found between his MMSE performance and outcome of the neuropsychologists' neurocognitive testing. Firstly, Mr M lost two points for 'Orientation' on the MMSE, while the neuropsychologists had found him to be fully orientated. He also lost one point for 'Calculation and Attention' on the MMSE, while the neuropsychologists had found these functions to be intact. Mr M also lost three points for the 'Recall' item in the MMSE (his most significant loss of points), and one further point for his design copy (Item 11).

According to the widely accepted MMSE cut-off score for dementia of 24, Mr M's score of 23 out of 30 classified him as meeting the diagnostic criteria for dementia (Lezak et al., 2004; Mitrushina et al., 2005).

The MMSE showed that Mr M had memory problems, as his loss of three points on the 'Recall' item (Item 5) suggests. However, the neuropsychologists' assessment revealed that his 'memory' problem was in fact a problem with retrieval on an executive basis, and not as a result of an axial amnesia. Here, analysis of the MMSE showed that although it could detect a memory deficit, it lacks items that can differentiate a memory-encoding problem from a memory-retrieval problem.

The final point lost on the MMSE was for the design copy (Item 11). In the light of the neuropsychologists' assessment, this lost point could well be explained by the spasticity in his hand, making handwriting difficult for him. Here, the analysis revealed that the MMSE does not factor primary non-cognitive deficits into test performance, thereby potentially missing other possible determinants that might account for test item failure.

Case 5: Mr D

Diagnosis and Case History

Mr D, aged 69 and with a grade 12 education, was referred to the neuropsychologists for a queried dementia following a diagnosis of Parkinson's disease. They found his memory to be intact, with no indication of a dementing process, but were, however, able to demonstrate specific executive problems to do with planning, generativity and initiative. Mr D scored 29 out of 30 on the MMSE, while the mean score for his age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993). Mr D had been previously diagnosed with Dopamine Dysregulation Syndrome, with features including anxiety, panic attacks, irritability and exaggerated rebound depression, and dysphoria. He had also been diagnosed, 18 years prior to his referral to Neuropsychology, with idiopathic Parkinson's disease.

Mr D, a former smoker, was found to have been self-administering large doses of Carbidopa for some months as a result of experiencing too many 'off' periods (freezing), reporting that he 'craved' the drug. Following this, he began experiencing hallucinations (without delusions), increased sexual drive, insomnia, dyskinetic movement and episodes of falling. As his excessive dosages were gradually reduced, his symptoms began to improve.

On formal neurocognitive assessment, Mr D was found to be orientated to person, place and time. He was hypophonic (that is, low speech volume), but his long-term memory when describing his history was good and his description of his medical condition was correct. He was talkative, cooperative, used appropriate language and was generally in good spirits. A difficulty with his working memory was noted when he produced a digit span of five. On testing of memory function, he performed perfectly on the 4 Hidden Objects Test, but struggled initially on the Babcock Story, recalling only four items on the first trial, with some minor confabulations (for example, he said the man's hands in the story were bleeding), and then eight items on the second trial. Mr D was able to answer all the prompts given to him correctly.

When executive function was formally assessed, his performances on the Tapping Test, the Similarities Test and the Proverbs Test were all considered adequate, although he was slow on all tasks. Based on the referral question, the neuropsychologists had tested for signs of a sub-cortical or Alzheimer's dementia, specifically focusing on tests of memory and executive function, and concluded that Mr D was not dementing, although he did have an adynamic-type dysexecutive picture.

Qualitative Analysis of the MMSE

According to the MMSE guidelines, Mr D's score of 29 out of 30 meant he was neurocognitively normal and 'above average' for his age and level of education (Crum et al., 1993). On the other hand, the neuropsychologists had demonstrated that although he was not dementing, he presented with an adynamic dysexecutive neurocognitive picture, consistent with that seen with a subtype of Parkinson's disease. Again, the MMSE missed evident neurocognitive problems, specifically the adynamia present in this case and the working memory impairment. The reason for this was that the MMSE lacks items to test the executive functions of the frontal subcortex (deep white matter), which would require items that look for asponenaity, adynamia and impersistence. Again, such items would need to be timed in order to examine the patient's ability to generate and initiate a response.

Case 6: Mr H

Diagnosis and Case History

Mr H was referred to the neuropsychologists as a result of his complaining of memory problems, and to query possible dementia. Following their assessment, he was diagnosed with exhibiting the early stages of a fronto-temporal dementia. Mr H was aged 63, with a grade nine education. He scored 25 out of 30 on the MMSE, while the mean score for his age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993). On assessment, it was observed that he displayed socially inappropriate behaviour (laughter), a labile mood, everyday forgetfulness, and suffered from incontinence. No hallucinations were reported, although he did have vague paranoia, especially towards burglars.

On formal neurocognitive assessment, Mr H was fully orientated and had a normal working memory. On tests of his executive function, problems became evident. He was impulsive, lacked insight and was extremely concrete in thought. In addition, he did not correct himself, as demonstrated with both the 20 Questions Test and the Similarities Test. These deficits of executive function led to the conclusion that the clinical picture was consistent with that of the early stages of a fronto-temporal type dementia.

Qualitative Analysis of the MMSE

In the case of Mr H, his MMSE score classified him as 'below average' for his age and level of education, but it did not reveal him to be demented, as he scored above the 24-point cut-off. When analysing how the MMSE missed the fact that he was dementing, a number of points arose. Firstly, the neuropsychologists' diagnosis rested on the fact that severe executive impairment was demonstrated, while on the MMSE, Mr H lost points in other cognitive domains outside of executive function. Specifically, he lost one point for 'Orientation' (getting the month wrong), three points for 'Attention and Calculation' and one point for 'Recall'. The only similarity between the MMSE and the neuropsychologists' testing was that both found a problem with his orientation. Although it was highly probable that he failed these specific MMSE items on account of his executive problems, there is no way of demonstrating this, as the MMSE is not theory driven, and therefore one can fail any of its items for many possible reasons. In this case too there were no MMSE items to test impulse control or abstract thought, as the Similarities Test, the Cognitive Estimation, and the 20-Questions Test had done in the neuropsychologists' appraisal.

Case 7: Mr MK

Diagnosis and Case History

Mr MK was aged 74, with a tertiary education. He was referred to Neuropsychology from Geriatrics for an investigation of his neurocognitive function as a result of queried possible vascular dementia or Alzheimer's disease. Following investigation, he was diagnosed with a severe working memory problem and executive dysfunction, consistent with a mild to moderate dysexecutive syndrome of the dorsolateral convexity. Mr MK scored 29 out of 30 on the MMSE, while the mean score for his age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993). It was reported that he had memory problems, irritability and acute listlessness. There was also a family history of Alzheimer's disease. Medical tests were negative for diabetes, heart disease, thyroid disease, calcium, B12, folate and syphilis.

A CT head scan revealed generalised cortical and cerebellar atrophy, as well as exvacuo dilatation, while a SPECT scan suggested he had a compromised vascular supply and neural changes, which were consistent with small infarctions of the cortex or generalised atrophy. Mr MK also had an extensive medical and surgical history, including deep vein thrombosis (DVT), peripheral vascular disease (PVD), and a transurethral prostatectomy. He reported problems with his memory and was worried, especially given his family history of Alzheimer's disease, a disease from which, at the time, his sister was suffering. He was not sure when his memory problems had started, but he thought they began about three years prior to the assessment. He did not report any problems with his mood, but he experienced word-finding difficulties.

On formal neurocognitive assessment, Mr MK was fully orientated and was found not to be amnesic: he gave a coherent history, demonstrating intact remote memory function, and he was able to perform adequately on both the Babcock Story and Rey Complex Figure recalls, without any confabulation on the verbal recalls. On testing of executive function, his generativity as tested with the FAS Test was good. His digit span, however, was only five. On the Fist/Side/Palm Test he had mild difficulty executing the task. Finally, while he displayed good judgment, he was found to be concrete and impulsive, as demonstrated by the Proverbs and Colour Word Interference tests respectively.

Qualitative Analysis of the MMSE

The MMSE failed to detect any of Mr MK's executive impairments. In contrast to the neuropsychologists' assessment, the MMSE did not reveal a working memory problem. His digit span was recorded at five, the normal human working memory capacity being seven units (Solms & Turnbull, 2002). On analysis, the MMSE's only (equivalent) 'assessment' of working memory is Item 3 ('Registration'), but this item only assesses three units of information, which is not complex enough to elicit even a severe deficit of working memory. To further compound this problem, Item 5 of the MMSE ('Recall') is actually a test of long-term memory, not working memory. This is because the intermediate item, Item 4 ('Attention and Calculation'), acts as a distraction task following Item 3, resulting in the Item 5 recall task requiring retrieval from long-term memory, not short-term/working memory.

Case 8: Mr E

Diagnosis and Case History

Mr E was 54 years old with a grade 8 education. He was referred to the neuropsychologists for possible expressive and receptive aphasia, Alzheimer's disease, or multi-infarct dementia.

He was diagnosed with a severe dysexecutive syndrome of the frontal-subcortical type (including adynamia) and dysarthria (disturbance of articulation). He scored 20 out of 30 on the MMSE, the mean score for his age and level of education being 27 according to the Age and Education Weighted Norms Table (Crum et al., 1993). Mr E presented with hypertension and Transient Ischemic Attacks (TIAs), and took Fluoxetine for depression. He reported a 12-month history of difficulties involving language production, naming, memory and getting dressed. Mr E's wife also reported that he experienced episodes of nocturnal confusion, despite his sleeping well otherwise. A CT head scan revealed severe generalised atrophy and peri-ventricular white matter changes.

On formal neurocognitive assessment, the neuropsychologists found Mr E to be fully orientated and severely dysarthric. He did not show any signs of aphasia and was not anomic, although he was slightly apraxic. When memory was formally tested, it was found to be intact, both in respect to his performance on the 4 Hidden Objects Test and his recall of a simplified Rey Complex Figure. Mr E did, however, have a severe working memory problem, only managing a digit span of three.

Testing of Mr E's executive functioning revealed that he was severely adynamic, evident in that he only scored five in total on the FAS Test. The combination of his severe dysarthria and adynamia made his speech problematic. Furthermore, Mr E had other signs of executive impairment. He was disinhibited on the Red/Green Test and struggled with the Fist/Side/Palm and problem-solving tasks too. He also failed to correct himself and showed poor self-verification on all the tests given to him. He showed poor complex reasoning on the Problem-Solving Task and displayed ideational perseveration.

Overall, the neuropsychologists concluded that Mr E's clear dysexecutive function of the subcortical type — evident from his ideational perseveration, adynamia, disinhibition, poor motor sequencing, poor self-verification and difficulty with complex reasoning — was indicative of multi-infarct dementia when seen in the context of his intact episodic memory, and his history of episodes of nocturnal confusion, hypertension and Transient Ischemic Attacks.

Qualitative Analysis of the MMSE

When analysing Mr E's MMSE performance, a number of interesting points were identified. In this particular case, the MMSE had detected that the patient was dementing, indicated by his score of 20 out of 30. However, a number of discrepancies were noted between the MMSE items that Mr E failed and the findings of the neuropsychologists' testing. Firstly, he lost a point in orientation (name of ward) on the MMSE, whereas the neuropsychologists had found him to be fully orientated. Likewise, he lost two points on the MMSE's 'Recall' task and a point on the 'Registration' task, while the neuropsychological assessment found his memory to be fully intact. Similarly, the neuropsychologists had found Mr E's language function to be fully intact, while he lost two points on the 'Language' items of the MMSE — one point for repetition and one for writing. The MMSE also lacks any item specifically designed to detect dysarthria. Mr E's severe dysarthria might explain his lost point on the MMSE repetition item (Item 8), but the MMSE does not allow any interpretation, with the result that a primary motor problem could easily be mistaken for a language deficit, which was the most likely scenario in this particular case.

In the light of the facts that firstly, all the striking neurocognitive deficits of Mr E were in the domain of executive function, and secondly, that the MMSE lacks any items testing executive

function, it must be surmised that the MMSE's detection of a dementia was due to Mr E's losing points on account of his poor working memory. But because the MMSE allows no room for interpretation, one cannot isolate a reason for failure of an item from multiple determinants of failure. Again, in this case analysis, it was clear that the MMSE lacks both an adequate assessment of working memory, as well as timed items, and items testing complex reasoning. These items are essential in demonstrating an adynamic frontal picture, and are mandatory requirements in an assessment tool that claims to screen for dementia.

Case 9: Mrs D

Diagnosis and Case History

Mrs D was aged 79, with a grade 7 education. She was referred to Neuropsychology with queried Alzheimer's dementia. Following neurocognitive assessment, she was diagnosed with axial memory impairment consistent with dementia of the Alzheimer's type. It was also decided that the possibility of a vascular overlay could not be excluded. Mrs D scored 16 out of 30 on the MMSE, while the mean score for her age and level of education was 25 according to the Age and Education Weighted Norms Table (Crum et al., 1993).

Mrs D presented with complaints of memory problems that had persisted for some time. Medically, she had diagnosed diabetes and hypertension. Her son reported that her memory had been problematic for about five years, and that she had also become topographically disorientated in previously familiar places. Mrs D was also nocturnally confused; once, at 3am, she was reported to have woken and started taking the curtains down. The memory problems reported included losing money, forgetting conversations and not remembering to take her medication. However, no difficulties with language or word finding were reported.

On formal testing of her neurocognitive functioning, the neuropsychologists found Mrs D to be orientated to time and person, but not place. Her working memory was slightly impaired, as her digit span was scored at five. On testing of verbal and visual memory, she performed poorly. On the 4 Hidden Objects Test, she made a number of errors, while on the Rey Complex Figure both her immediate and delayed recalls were extremely poor. Despite this, Mrs D's attention during the tasks was good. When her language function was assessed, it was found to be normal. She struggled on visuospatial tasks and on constructional tasks, most evident from her Rey Complex Figure copy.

On assessment of her executive functioning, Mrs D was found to experience difficulty with complexity, although her simple problem-solving ability was intact. She struggled with the Fist/Side/Palm Test of motor sequencing. On the FAS, a generativity problem was noted, as her verbal fluency was poor, with a total of just 13 over the three letter trials. Examination of her performances on the executive tests showed Mrs D to be clearly impulsive and displaying ideational perseveration, and it was deemed that she had a mild dysexecutive syndrome. The neuropsychologists, however, determined that her executive problems could not account for her poor memory.

Qualitative Analysis of the MMSE

When analysing the MMSE in this case, it was first noted that there was a discrepancy between her orientation performance as noted by the MMSE and the neuropsychological assessment. She lost five points for orientation to both 'time' and 'place' on the MMSE, while she was found to be disorientated only to 'place' according to the neuropsychologists. The MMSE did not assess her orientation to 'person', which is a key aspect in assessing orientation, as the MMSE lacks this component of assessment. Mrs D lost four points for

‘Attention and Calculation’ on the MMSE and she scored full marks on its ‘Registration’ task, which was in keeping with the findings of the neuropsychologists’ assessment.

The neuropsychologists found Mrs D to have severe memory problems on the basis of their clinical assessment. However, the MMSE missed the severity of these problems, as was evident in the fact that she lost only one point for the ‘Recall’ item (Item 5). Analysis of this finding suggested that the item dedicated to the assessment of memory in the MMSE (that is, Item 5) was too simplistic to adequately probe memory function and, in addition to this, there is no delayed recall task present in the MMSE.

A further discrepancy between the neuropsychologists’ assessment and the MMSE lay in the observation that Mrs D lost, in total, three points on the ‘Language’ items of the MMSE — specifically, one point for naming objects (Item 6), and two points for the 3-step command (Item 7) — whereas the neuropsychologists had demonstrated her language function to be intact. Mrs D failed the design copying item (Item 11) on the MMSE, which was most likely due to the constructional difficulties the neuropsychologists demonstrated. However, a more detailed evaluation of constructional abilities is required in the MMSE, as one could fail this solitary item for any number of possible reasons. This observation is valid for the majority of the MMSE items. Finally, the analysis revealed that the MMSE has no items dedicated to identifying the visuospatial difficulties that Mrs D was experiencing. It also lacks the items necessary for testing executive function, including, in this case, ideational perseveration, concreteness, motor sequencing problems and impulsivity.

Case 10: Mrs S

Diagnosis and Case History

Mrs S, aged 56 and with a grade 9 education, had shown frontal lobe release signs, and was referred for a neurocognitive assessment for queried problems with her short-term memory and attention. She was also queried as having constructional apraxia, possible anxiety or depression, and hypoxic damage. Following their assessment, the neuropsychologists found Mrs S's higher cortical functions normal. Mrs S was Afrikaans speaking and scored 23 out of 30 on the MMSE, while the mean score for her age and level of education was 28 according to the Age and Education Weighted Norms Table (Crum et al., 1993). It was reported that Mrs S had undergone steady cognitive decline, that she had problems with her attention, and was showing signs of depression and anxiety. From her medical history, it was discovered that she had had a long stay in an intensive care unit, after which time she was diagnosed with a peripheral neuropathy, with an intention tremor, frontal release signs, and problems with her working memory and attention. A CT head scan showed mild cerebral atrophy, in keeping with her age.

On formal neurocognitive assessment, the neuropsychologists had found Mrs S's executive functions to be normal and she produced a digit span of six for working memory. The assessment of her language function found it to be normal. Likewise, tests of memory found her memory function, both verbal and visual, to be fully intact. Testing of gnosis found no deficit. Finally, testing of Mrs S's constructional ability found her not to have any constructional apraxia.

Qualitative Analysis of the MMSE

When analysing Mrs S's MMSE performance, it was first noted that her total score of 23 out of 30 categorised her as meeting the diagnostic criteria for dementia, as score of 24 or below is taken to mean that a dementing process is present (Lezak et al., 2004). This performance was in stark contrast to the neuropsychologists' appraisal, which found her higher cortical functions to be intact. Mrs S lost points on the MMSE in a number of areas: two points for the 'Orientation' item (Items 1 and 2), three points for the 'Attention and Calculation' item (Item 4), one point for 'writing' (Item 10 of the 'Language' section), and one for the design copy (Item 11).

A possible explanation for this major discrepancy between the MMSE's outcome and the neuropsychologists' findings is that Mrs S performed poorly on the MMSE because her first-language was Afrikaans and the MMSE was administered in English. This analysis revealed that the MMSE's lack of theoretical underpinnings in informing its design means that one is left to guess the possible cause of a patient's failure on any particular item. This observation also highlighted that the MMSE does not allow for the recording of other variables that might influence/account for the test performance, for example the patient's first language or primary deficits (such as visual or auditory problems).

Discussion

The importance of this aspect of the research, and the implications thereof, should not be underestimated. Firstly, these findings were a crucial factor in informing the ongoing design and development of the GSNSB, and in guiding the implementation of the decision-tree approach — vital if the shortcomings of the MMSE approach were not to be repeated. Secondly, the MMSE is an extremely widely used screening tool in South Africa and many clinicians, in both the public and private medical settings, rely on it, rightly or wrongly, to make clinical judgments and decisions. It is therefore of fundamental importance that South Africans be provided with a diagnostically meaningful, purposeful screening tool — a task that cannot be achieved unless the limitations of its predecessor are identified, understood and rectified. In the light of this, a thorough appraisal of the effectiveness of the incumbent screening tool in the South African context was first required.

In total, ten clinical cases, all referrals from the Neuropsychology division's archives, were reviewed and qualitatively analysed during this aspect of the research (see Table 2.1 below for a summary of the cases reviewed).

Table 2.1

Case summaries

Case	MMSE Score	Neuropsychologists' Impression
Case 1	30/30	Lewy Body Dementia
Case 2	30/30	Semantic Dementia
Case 3	30/30	Fronto-temporal Dementia
Case 4	23/30	Depressive Pseudodementia
Case 5	29/30	Executive problems of planning, initiative and generativity
Case 6	25/30	Early stages of Fronto-temporal dementia
Case 7	29/30	A dysexecutive syndrome of the dorsolateral convexity
Case 8	20/30	A dysexecutive syndrome of the frontal-subcortical type
Case 9	16/30	Dementia of the Alzheimer's type
Case 10	23/30	Higher cortical functions normal

When reviewing the results of the qualitative case analyses, the first striking finding was the number of cases where the MMSE had failed to detect the patients' executive impairment. This finding is consistent with the views of Nestor and Hodges (2003), Simard (1998), and Juby et al. (2002) in supporting the notion that the MMSE's lack of executive tests proves problematic in detecting dementia with frontal involvement. This was clearly illustrated by the number of patients who were diagnosed as dementing by the neuropsychologists, yet managed to perform well on the MMSE (three such patients scored 30 out of 30). In all but two of the ten cases (cases 4 and 10), the MMSE missed clear executive dysfunction. In the

two cases where this finding did not hold true, the obvious explanation was that these were patients who did not have executive dysfunction — one being a depressive pseudodementia (Case 4), while the other (Case 10) was a patient who did not have any neurocognitive deficits at all.

Further analysis of the findings highlighted the specific aspects of executive function that the MMSE missed. Firstly, the MMSE failed to adequately assess working memory, clearly evident in cases 5, 7, 8 and 9. This is due to the ‘Registration’ task (Item 3) not being detailed enough to test working memory sufficiently, as it uses only three items, rather than the mandatory seven items required to test working memory adequately (Solms & Turnbull, 2002). Additionally, this ‘Registration’ task is not even specified as a test of ‘working memory’, but rather as the preliminary learning task for the ‘Recall’ test (Item 5). The further aspects of executive function that were not assessed by the MMSE included generativity, abstract thought, inattention, perseveration, ability to problem solve, adynamia, planning and motor sequencing — these aspects of executive function, and the ability to elicit them, were all key to informing the diagnosis in the neuropsychologists’ assessments.

In order to assess the above-mentioned executive functions, the MMSE would firstly have required timed items with time-limits, in order that the patient’s lack of generativity, indicative of possible adynamia or perseveration, could be tested for. The lack of an item equivalent to the Fist/Side/Palm Test meant that motor sequencing could not be examined, while the lack of items such as the Similarities, Proverbs and 18 Book Problem meant that abstract thought and the ability to problem solve could not be examined.

The findings from the case studies also revealed the MMSE's limitations in assessing other neurocognitive functions, beyond the above-mentioned lack of executive tests. It was evident that the assessment of 'Language' function in the MMSE is inadequate in certain areas.

Firstly, the assessment of naming, as noted in Case 3 for example, is inadequate due to the fact that only two items, a 'pencil' and a 'watch', require naming (Item 6 in the MMSE). The problem here is twofold. Firstly, these two items are too familiar to adequately tax naming ability, and secondly, in order to test naming ability, far more items are required to adequately detect an anomia. This latter point is clearly illustrated by the stance adopted by the neuropsychologists in assessing naming — they require a large number of items (as seen in the Boston Naming Test) in order to sufficiently probe naming ability and to exclude other possible causes of failure on the naming task. This task simply cannot be achieved with only two items dedicated to the assessment of naming.

This problem of lack of detail in its 'Language' items also applies to the assessment of writing ability (Item 10). Case 2 clearly illustrated that failure to specify the length of the sentence the patient has to write means that a writing impairment can easily be missed if the patient produces only a short sentence on command. These findings specific to problems of the language assessment in the MMSE support the findings of Tombaugh and McIntyre (1992) and, more recently, Lopez et al. (2005) who regard the language items in the MMSE as being overly simplistic. These findings are also in accordance with the Feher et al. (1992) finding that the majority of the language items in the MMSE lack sensitivity in detecting impairment.

The final problematic area identified by the case study analyses was the MMSE's assessment of memory function. As was the case with its assessment of 'Language', the MMSE lacks

sensitivity and detail. Most telling is its lack of a sufficiently detailed test of immediate recall after distraction, and also the total absence of any ‘delayed recall’ item. Using the neuropsychologists’ assessment as a benchmark for analysis, one sees that even when testing memory retrieval in its simplistic form, the test they rely on (the 4 Hidden Objects Test) is still more complex than the mere three-item recall offered in the MMSE. In support of these findings regarding the inadequate assessment of memory, Nestor and Hodges (2003) state their surprise that this tool, so widely used in the field of Alzheimer’s disease, is so lacking in this domain.

A further problem with the MMSE, as is seen clearly in the analysis of Case 4, is that it lacks the items required to differentiate between an impairment of memory retrieval and an impairment of memory encoding (two processes that are entirely different). In order for the MMSE to make such a differentiation, multiple-choice questions would require inclusion in addition to the ‘recall’ task, in order to determine whether or not the patient could benefit from prompting. If a patient is able to benefit from prompting, in the context of a poor initial recall, it is a sign of an impairment of executive function (retrieval) rather than that of hippocampal function (encoding) (Butters & Miliotis, 1993). This lack of ability to adequately appraise memory function is especially worrying when one considers that the primary historical use of the MMSE was as a screening tool for dementia of the Alzheimer’s type.

A further interesting finding of the present study was that due to the ambiguity caused by the lack of clear administration instructions and guidelines in the MMSE, many of the Groote Schuur Hospital neurologists interviewed during the course of the study reported that they present three items visually to the patient for the ‘Registration’ task, rather than just saying

aloud the names of three items without visual presentation (which is actually what is required). The result of this is that often Item 3 is not even a working memory task at all, as the patient sees three objects visually all the time during this task, meaning that at no point does he/she have to 'hold' the objects in short term/working memory.

At this junction, it is important to examine the findings of the above-mentioned analyses in light of the origins of the MMSE as a screening tool. The original creation of the MMSE as screening tool in the 1970s occurred at a time when the neuropsychological understanding of neurocognitive disorders and the dementias was not as advanced as present-day knowledge. Specifically, knowledge of the cognitive functions of the frontal lobes, and the varieties of dementia, was at the time far from what it is today. This point was highlighted by Luria in the early 1970s: “[t]he functional organization of the frontal lobes and of their individual zones has been inadequately studied, and we can therefore examine only some of the most general features of those alternative forms of the frontal syndrome which arise in lesions of different parts of the frontal region” (Luria, 1973, p. 221).

During the 1970s, Alzheimer’s disease was the only adequately researched and understood dementia. Modern neuroscience now understands that many varieties of dementia exist, a number of which have executive dysfunction as a key clinical feature, which can manifest at various stages of a dementing illness, depending on which specific dementia is present. Understanding the complex nature of the workings of the frontal lobes and executive function in any neuropsychological account of brain disease is vital, given the hierarchical way in which executive functions ultimately govern and mediate in all neurocognitive functioning (Damasio, 1979; Kolb & Whishaw, 2003; Luria, 1973; Solms & Turnbull, 2002; Walsh, 1987, 1999).

Key to identifying, neuropsychologically speaking, which particular dementia a patient might have, is knowledge of the order (of the pattern) at which specific neurocognitive deficits typically declare themselves for each of the different possible types of dementia. For a pertinent example, in Alzheimer's disease axial amnesia is typically the first neurocognitive deficit to declare itself when, initially at least, other neurocognitive functions typically remain largely intact. Alternatively, in the example of Pick's disease, executive dysfunction is typically the first neurocognitive sign detected (Hodges & Miller, 2003). Therefore, the point at which executive functions become affected in the course of the disease progress provides vital diagnostic information as to which dementia is present.

Given the key role executive function now plays in the classification and diagnosis of the dementias, it is imperative that a modern screening tool has items that test executive function. The MMSE is not a modern screening tool. Given that in the early stages of certain dementias executive dysfunction often precedes other neurocognitive deficits, it is of the utmost importance that a screening tool be able to test executive function. In light of the findings of this study, the MMSE might be able to measure the severity of a dementia in certain instances, but it is clearly incapable of detecting the early stages of any dementia that initially presents with executive dysfunction. A primary reason for this finding is that the MMSE was designed in an era where the assessment of executive function in the context of dementia did not play a major role.

Another major shortcoming identified in this study was the MMSE's lack of theoretical underpinnings. A complete understanding of neuropsychology incorporates the belief that multiple determinants of failure on any particular test item are possible (Walsh & Darby,

1999). In the case of the MMSE, it is clear that the design of the tool did not take into account how the different functions of the brain operate or how they are organised. In other words, the MMSE is not a theory-driven tool. The consequence of this is that a patient might fail any of the items for many different reasons, yet there is no way of isolating the specific determinant of failure. This is because the items are not ordered in a logical way nor do they systematically exclude possible causes of failure by asking specific questions or by applying theoretical know-how.

A clear illustration of the above-mentioned point is the example of working memory, evident in cases 5, 7, 8 and 9. If the MMSE had been based on theoretical knowledge, the 'Registration' task (Item 3) would have tested seven 'units', not three, as it is theoretically understood that a normal human working memory span should be seven. A further example is the lack of items testing executive function. In summary, there is far too much room left for interpretation with the MMSE and it lacks the theory to systematically test and exclude other possible causes of failure of a test item. This study highlights the urgent need for a theory-driven neurocognitive screening tool for South Africa. The only way to solve this inherent problem is to adopt a decision-tree approach (see Chapter Six), to exclude other possible determinants of failure on each item. This approach rests firmly on an understanding of how the patient's performance on a neurocognitive test relates to the underlying brain mechanism and specific pathology present.

In conclusion, the central finding of this case-based study was that the MMSE is demonstrably unreliable and unsuitable as a screening tool in the current South African context. The primary reason underlying this conclusion is that the MMSE is outdated, having been designed at a time when the neuropsychological understanding of dementia was limited

relative to present-day knowledge. Additionally, the MMSE's lack of theoretical underpinnings, combined with the total absence of items testing executive function, make it unsuitable to accurately diagnose the presence of dementia in the current clinical context.

Furthermore, the intended purpose behind the design of the MMSE was largely limited to the screening of dementia and cognitive impairment in psychiatric patients, while the current South African clinical context requires a tool that can screen a wider range of neurocognitive functions, including dementia. The lack of resources in South Africa necessitates a screening tool that can provide meaningful insights into neurocognitive functioning in the absence of a specialist neuropsychologist. Given this need, the MMSE is clearly beyond its diagnostic capabilities when used in this context. The MMSE should, at best, be used as the source of multiple hypotheses about a patient.

CHAPTER THREE: TRANSLATING THE GROOTE SCHUUR NEUROCOGNITIVE SCREENING BATTERY

Translation of the Groote Schuur Neurocognitive Screening Battery

Language is considered by many to be one of the single most significant moderators of test performance, given that the performance on a test could be a result of language difficulties if it was not administered in the patient's home language, rather than the result of the function being tested (Griessel, 2005; Grieve, 2005; Nell, 2000). As outlined in Chapter One, a major shortcoming of the Groote Schuur Neurocognitive Screening Battery (GSNSB) Prototype was that it existed only in English. This proved problematic as many of the patients seen in the Western Cape region are either Afrikaans or isiXhosa first-language speakers, with varying proficiencies in English as a second or even as a third or fourth language. Edwards and Louw (1997, p. 27) highlight the current problem in South Africa:

[I]n the development of psychological tests, focus has largely been on the white population. Many tests have been standardised in English and Afrikaans. However, they are designed for use by first language speakers. When they are used by people answering in their second language, results can be misleading. As for the many South Africans who cannot speak English or Afrikaans, hardly any tests are available at all.

This remains a persistent problem within neuropsychology, especially since the one or two tests that have been translated are psychometric batteries (such as the Wechsler Adult Intelligence Scale) not specific to neuropsychological testing, as opposed to individually used 'bedside' assessment tests (see Chapter Four). There is a serious shortage of neuropsychological tests in languages other than English (Artiola i Fortuny & Mullaney,

1998; Mitrushina et al., 2005). The lack of translated neuropsychological tests internationally can be explained by the fact that they are usually required for minority groups such as, for example, Hispanics in North America. However, in South Africa the opposite, unique, situation is true, as it is the majority groups who are not catered for. Griessel (2005, p. 88) highlights that in South Africa (where there are 11 official languages), “for the foreseeable future the majority of measures will still be available in English and Afrikaans as it will take a number of years to adapt and translate measures into different African languages”.

Through the clinical experiences of the author and his colleagues, it was frequently observed that, when assessing a patient in his/her second or third language, there was considerable uncertainty as to whether he/she performed poorly on the basis of his/her language or on the basis of deficit. This task required an unrealistic degree of clinical judgement. “Ideally, the neuropsychologist-examiner should be fluent in the client’s home language. Under conditions of mass migration across language borders, in Africa and eastern Europe, and in settings of linguistic diversity, this is often impossible” (Nell, 2000, p. 148).

A second problem occurring in this clinical context was that interpreters were often required to facilitate the clinical assessments of patients who did not speak English — often nursing staff had to be relied upon to perform this crucial task. This scenario proved time and again to be difficult and cumbersome, producing only limited success. A number of reasons lie behind these difficulties in using an interpreter. The interpreter can potentially either omit key content during the process, or might add in additional information, resulting in an invalidation of the outcome of tests being administered (Nell, 2000; Swartz, 1998). Further problems with the use of interpreters is that they might condense or summarise what the patient said into what they thought he/she meant, failing to accurately reflect the clinician’s

and patient's actual views. Finally, errors of substitution might be made whereby the interpreter replaces what was said with something that was not actually mentioned. Role exchange might also occur whereby the interpreter takes on the clinician's role and asks questions of his/her own (Swartz, 1998). Many of the subtle nuances inherent in both a skilful hypothetico-deductive clinical assessment and in the disorders seen in the patients are lost in translation when an interpreter is relied upon.

In addition to the pitfalls identified above, the objective of producing an effective tool, worthy of being deemed 'transferable technology', required that the GSNSB could cater for all patients in the Western Cape's population. Unfortunately, taking logistical and time constraints into account, the GSNSB could not be translated into all 11 of South Africa's official languages, thereby extending its use to non-English speakers in other provinces. This has been earmarked as a future goal, beyond the scope of this current research.

It is hoped that the provision of a structured screening tool in both Afrikaans and isiXhosa will help to eliminate some of the problems associated with using an interpreter. Such a tool will provide a set, comprehensive structure and protocol to screening assessments, while at the same time reducing the reliance on clinical judgement necessitated when using an interpreter. Furthermore, because the GSNSB will be administered primarily by clinicians (that is, not specialist neuropsychologists), it is more likely that these individuals will be able to speak isiXhosa and/or Afrikaans (bearing in mind the handful of neuropsychologists available), thereby further reducing the need for interpreters. As Swartz (1998, pp. 45–46) has so succinctly put it, "the ideal situation for mental health interaction is one in which there is no need for an interpreter".

Rationale

One of the fundamental objectives of this research project was to provide South African patients with neurocognitive testing in their first language. In order to achieve this crucial goal, it was necessary to translate the GSNSB Prototype into both Afrikaans and isiXhosa versions due to the specific demographics and history of the Western Cape region of South Africa. Where applicable, it was also necessary to translate some of the neurocognitive tests contained in the GSNSB as part of this overall process. It was vital that an exhaustive translation approach be adopted in order to ensure the success of this screening tool.

Historically, along with English, Afrikaans and isiXhosa have, and continue to be, the predominant languages spoken in the Western Cape region. While English and Afrikaans are spoken throughout South Africa, the isiXhosa language originates in the Eastern Cape Province of South Africa. During the Apartheid era, many rural people from this region migrated west from their 'homelands' to search for work in Cape Town and its environs, where they were forced to dwell in marginalised areas, allocated by the government of the time (Joyce, 1989). Consequently, many isiXhosa people now live permanently in this region, evidence of which is readily apparent when one examines the makeup of the patient population of Groote Schuur Hospital. It was within this multicultural setting that many of the clinical observations that informed this present research were made. The need to provide the uncatered-for majority of South African's population with effective neuropsychological screening in their home language is both compelling and urgent.

Chronological Order of the Phases of the Groote Schuur Neurocognitive Screening Battery's Translation Process

This doctoral study began with the GSNSB Prototype, which contained the original constellation of bedside tests assembled from abroad, and existed only in English. It must be remembered that the final form of the GSNSB created by this research, with its adapted tests, instructions, and translated versions — as validated by this research (see Chapter Six) — was only completed just prior to the ‘validation phase’ of this project. To reach the final point in the GSNSB’s translation into Afrikaans and isiXhosa, a number of phases of development were required. The chronological order of these various phases carried out in the translation process (as described in this chapter) was as follows. See Table 3.1 for a summary of the phases.

Table 3.1

Phases undertaken during the translation process

Phase One
Translation of GSNSB Prototype and adapted tests into Afrikaans and isiXhosa
Phase Two
Further translation work done to GSNSB and tests following initial pilot study
Phase Three
Proof-reading all three language versions of the entire GSNSB

Phase One: Translation of the Grootte Schuur Neurocognitive Screening Battery Prototype and Relevant Tests

There were a number of challenges in translating the GSNSB Prototype. At the heart of these was the need to assemble a translation team who were not only highly proficient in their respective languages, but who also understood the subtleties and nuances of the terminology, so important in the clinical context. In other words, these individuals needed to translate the GSNSB, and to understand the medical context in which it was to be used and also how the terminology used (some of which was highly specialised) equated to the clinical application. Various authors (Artiola i Fortuny & Mullaney, 1997, 1998; Loewenstein, Arguelles, Arguelles & Linn-Fuentes, 1994; Van der Vijver & Hambleton, 1996) have stressed the importance of translators having in-depth knowledge of both the relevant culture and the language into which the test is being translated. This ensures that translations are conceptually equivalent as well as being linguistically so. This often makes the task extremely challenging, as idioms, expressions and other figurative language often vary between cultural/language groups. A further complication is the “fact that a combination of languages, referred to as a ‘township patois’, is commonly used in the residential areas surrounding cities and a pure version of one language is seldom spoken” (Grieve, 2005, p. 231). The need for cultural equivalence throughout the test development and translation process is stressed by Kanjee (2005) and Teng (1996). Finally, the translators needed to have time available to work closely with the author in translating and to help with the test adaptations and data collection at a later stage.

Brickman, Cabo and Manly (2006, p. 95) highlight the importance of adopting stringent criteria when translating neuropsychological tests:

It is common practice in dealing with non-English speaking populations to simply translate English language measures into a new language. Even if translation and back-translation methodology is used, individuals who perform the translation are often not neuropsychologists, not well educated in the target language, and not generally balanced bilinguals. Ideally, the test designers should be fluent in the language of the target population and have knowledge of how that language may differ by geographic region.

The importance of these requirements could not be overestimated: if specific meanings of terms or instructions were to be lost in translation, the clinical utility of the GSNSB would be severely jeopardised.

Van der Vijver and Hambleton (1996) have compiled a comprehensive set of guidelines to meet these stringent demands, highlighting three types of potential bias to avoid during the translation process: construct, method and item bias. Construct bias refers to construct measured by an instrument that is different across cultures.

Construct bias is more likely to occur when an existing instrument is translated than when an instrument is simultaneously developed for different languages. In the latter case it is easier to avoid ethnocentric tendencies and to remove words and concepts in a source language that are not common in the two languages and cultures. A successful avoidance of ethnocentric tendencies in instruments may require a multicultural, multilingual team with an expertise in the construct under study.

(Van der Vijver & Hambleton, 1996, p. 90)

Method bias refers to bias in the testing procedure stemming from intergroup differences in familiarity of the items, appropriate ways to respond, and motivation, etc. This is avoided by

having high test-retest reliability and through the collection of qualitative data and observations regarding issues affecting the testing procedure (Van der Vijver & Hambleton, 1996).

Item bias refers to the bias with specific test items caused by poor wording of items and inaccurate translation, and can be avoided by carefully examining and back-translating items, as well as through a variety of statistical procedures (Van der Vijver & Hambleton, 1996).

Care was taken throughout the translation process to ensure that these biases were avoided, both in the translation of the body of the GSNSB and in translating the relevant individual tests (highlighted in Chapter Four). The team of translators was also assembled with these objectives in mind.

After months of searching and careful planning, the team of translators chosen to meet the research challenges was selected from staff of the South African Languages Department at the University of Cape Town — from the Afrikaans and isiXhosa departments respectively. The reasons for this choice were twofold. Firstly, these individuals had experience in lecturing on and tutoring their respective languages and cultures, and all were balanced bilinguals (Brickman et al., 2006). Secondly, and of equal importance, they had also all taught medical students within the clinical context, and all had extensive experience working as interpreters for clinicians with patients at Groote Schuur Hospital. This repertoire of skills not only made them suitable for translating the GSNSB, but also enabled them to collaborate both as language and cultural experts during the test-adaptation process (see Chapter Four), and as interpreters when assessing participants in the other aspects of the study (see Chapters Five and Six).

The continuity which this team provided by being able to work on various aspects of this research had many advantages and served to strengthen the overall validity of the research. Not only did it allow for one set of individuals to be trained in the usage and workings of the battery, but furthermore, it enabled these individuals to gain experience and insights in various aspects of the research. This meant that the experience that they gained during the translation and test adaptation processes could be utilised in their role as interpreters, thereby adding to the qualitative input and feedback they were able to provide. Observing first-hand how the patients/participants in this study responded to the tests and translations they had advised on gave a richer tapestry of knowledge and insight into the cross-cultural challenges that the research was attempting to address.

In total, the initial translation of the GSNSB Prototype and the relevant newly adapted tests took just over seven months to complete. This process involved a number of key stages and required weekly meetings with the translators, as well as consultations with a panel of neuropsychologists when required. The first stage, which took approximately two months, involved briefing the translators and introducing them to the GSNSB's content. This process was carried out separately for the isiXhosa and Afrikaans translators. In total, two isiXhosa and two Afrikaans translators were employed.

The translators were first familiarised with the nature and intended purposes of the research, and then with the material contained in the GSNSB. Next, each translator was given half of the GSNSB to translate — that is, the two isiXhosa translators each received half the GSNSB to translate into isiXhosa, and each of the Afrikaans translators received half of the battery to translate into Afrikaans. Once each translator had translated his/her respective half, it was

then given to his/her colleague to back-translate; this cross-referencing approach added thoroughness to the process, and also allowed it to be completed within the timeframe and resources available. The teams of translators then met with the author to discuss concerns and any inconsistencies with the respective versions of the GSNSB (see Appendix C and Appendix D for the final Afrikaans and isiXhosa versions).

The importance of adopting a thorough approach to the translation, incorporating the back-translation method, is frequently emphasised (Brickman et al., 2006; Brislin, 1986; Kanjee, 2005; Puente & Ardila, 2000; Swartz, 1998). The idea of having a panel to review the success of the translations, as adopted by this study, is highly recommended in order to resolve differences between old and new versions and to resolve any inconsistencies (Brislin, 1986; Kanjee, 2005; Puente & Ardila, 2000). A similar approach was effectively utilised in the South African medical context by Mkoka, Vaughan, Wylie, Yelland and Jelsma (2003) when translating an international instrument (the European Quality of Life 5-Dimensions (EQ-5D)) into isiXhosa.

Translation of the adapted tests

Apart from the translation of the GSNSB Prototype itself, the translators were also given the individual tests requiring translation in preparation for the pilot study. This process took roughly three months to complete and ran concurrently with the translation of the GSNSB Prototype. These tests comprised those whose composition was language based and included the original Babcock Story, the newly created Township Fire Story, the original Anna Thompson Story, the newly created Mary Selo Story, and the COWAT (FAS Test) (see Chapter Four for details of these tests). The translators worked together in translating the individual tests in the presence of the author, who monitored the process and facilitated in

deciding which terms to use from the respective languages as the most suitable translations of the English.

As was the case with the main body of the GSNSB Prototype, once the new replacement tests had been conceptualised and created, they too were translated from English into both Afrikaans and isiXhosa. This process provided new challenges and proved difficult for some of the tests. With both the Mary Selo Story and the Township Fire Story, the problem was that when translated, both the isiXhosa and Afrikaans versions were slightly longer due to the unique phraseology and sentence structure/composition (the morphological differences) of the respective languages. Consequently, when it came to dividing the stories up into their 21 scored units of information, they had 22 and 23 units of information respectively. This problem was initially missed and was only discovered after the mini pilot had been conducted, highlighting the value of including this mini pilot (see Chapter Four). This problem took a while to solve and was not easy to rectify. Eventually, the phrasing of the English version of the Township Fire Story had to be subtly changed to accommodate certain phrases in the Afrikaans and isiXhosa versions. Again, this example highlights the subtle nuances across languages that made this study so interesting and challenging.

Translation of the Naming Test also proved a challenge as it had to be ensured that each language's vocabulary included the specific terms required to name each test item accurately. As with the example of the 'supernatural' creatures (for example, the *unicorn*) mentioned in Chapter Four, certain items had to be excluded as suitable isiXhosa and/or Afrikaans equivalents did not exist, or because they carried subtly different meaning. Here, the cultural experts' knowledge of the indigenous cultures proved invaluable.

Especially relevant to the translation of the stories was the need to ensure that the terms used in both the Afrikaans and isiXhosa versions did not subtly alter the meaning of the story. For example, in the translation it was important to retain the specific meaning of the word ‘shack’ as used in the English Township Fire Story rather than to use the loose translation ‘house’, as these are two very distinct types of structure. Moreover, both are mentioned in the story, so the same word could not appear twice anyway (due to the scoring procedure). Here, ‘shack’ refers to an informal home made of waste and iron sheeting. Consequently, the Afrikaans word *plakkershut* was used for ‘shack’ instead of the word *huise*, which was used for ‘houses’. Equivalently, for the isiXhosa, the word *etyotyombeni* was used instead of the word *ezindlwini*, which was used for ‘houses’.

Another example was the word ‘township’ in the story, which refers to an informal settlement of people, typically common in the marginalised areas of South African cities. Such townships are also often referred to as a ‘location’, similar to a ‘shanty town’ in other countries. Here it was important to retain the meaning of an informal settlement in the story, as the type of housing in an area where shacks are found differs drastically from suburbs, where modern western-style houses exist. Therefore, for the Afrikaans version the word ‘township’ was retained, as this is also formally used in Afrikaans, and the other Afrikaans words for township — *dorpie* and *stadsgebied* — both refer to more formal towns. In the isiXhosa version, the word *elokishini* was preferred to the word *edolophini* for the same reason.

General difficulties that emerged during the translation process are elaborated on in the section below.

Difficulties Encountered During the Initial Translation Process

A number of key difficulties occurred during the initial translation process. The first problem raised by the translators to be solved by the author was that the vocabularies of their respective languages lacked specific terminology for much of the neuropsychological jargon. Grieves (2005, p. 231) comments, with specific reference to the South African context that, “[s]ome languages do not have the concepts and expressions required by measures and an equivalent form of the measure cannot be translated”. This problem was especially true of isiXhosa and was compounded by the fact that no isiXhosa medical dictionaries currently exist in completed form (prototypes are still in development). In any event, it is unlikely that they would include exhaustive accounts of neuropsychological jargon once complete (in fact, the possibility of including such terminology in the future was discussed with the translators as a possible collaboration stemming from the present study). This problem was a real concern, given that much of the GSNSB’s instructions and clinical definitions are dependent on such terms.

This lack of medical terminology and concepts in isiXhosa also highlighted the recurring theme of ‘acculturation’, which had been encountered in many aspects of this research. The term acculturation refers to the process of the integration of individuals from their own into a new culture, as well as exposure to one’s own culture during childhood, and includes the learning of manners and style, the habits, beliefs and the values of the prevailing culture (Grieve, 2005; Uzzell, 2007). This process occurs at different rates and influences different aspects of behaviour to varying extents. “Levels of acculturation are defined by the degree to which cultural values, beliefs, and practices are incorporated by members from another culture” (Uzzell, 2007, p. 5). In South Africa, this process is all the more diverse and complex because of the multicultural composition of the population. One typical scenario in

South Africa is the move of rural people to the cities, where their traditional beliefs are left behind as they become exposed to the more western way of life (Grieve, 2005; Shuttleworth-Jordan, 1996). In this study, the issues surrounding medical terminology and concepts illustrate how the acculturation process involving the traditional isiXhosa culture and western medicine is still in flux, and that the challenges present in South Africa's cross-cultural medical settings are both very real and persuasive. The fact that neuropsychology is an emerging field in South Africa made tackling the challenges encountered during this groundbreaking research all the more intriguing.

After numerous consultations with the translators, the terms mentioned below were found not to have suitable isiXhosa equivalents as direct matches. Where possible, some possible substitutions for these terms were tried, but the author decided that these did not carry the exact clinical meaning as the original English. It was therefore decided by the time of the fourth draft of the GSNSB Prototype's translation into isiXhosa to retain the original English terms in these instances. The terms that could not be translated into isiXhosa include: *3-D Analysis, right hemisphere, frontal, prefrontal cortex, pre-motor, memory* (when referring to the clinical use of the term), *retrograde, anterograde, diencephalic, mesial temporal, paraphasia, pure motor aphasia (oral apraxia), pure word deafness, pure alexia, pure agraphia, spatial perception and cognition, spatial acalculia, anosognosia, somatoparaphrenia, misoplegia, deep white matter, orbital/basal, utilisation behaviour and dorsolateral.*

In contrast, the only words for which the Afrikaans translators could not find suitable Afrikaans terms were: *retrograde* and *anterograde* (both used when referring to amnesia). These terms were retained in English in the final Afrikaans GSNSB. The major reason

identified for this difference between the availability of isiXhosa and Afrikaans terms was that there are a number of Afrikaans medical dictionaries available. In addition, many of the Afrikaans terms had been accurately Anglicised, consistent with the original English terms. There are a number of examples of anglicising: mesial to *mesiale* in Afrikaans, paraphasia to *parafasie*, trans-motor to *trans-motories*, trans-sensory to *trans-sensories*, anosognosia to *anosognosie*, anosodiaphoria to *anosodiaforie*, misoplegia to *misoplegie*, somatoparaphrenia to *somatoparafrenie*, mesial to *mesiaal*, orbital to *orbitaal*, dorsolateral to *dorsolateraal*, premotor to *premotories*, and prefrontal to *prefrontaal*.

A further problem encountered with the Afrikaans translations was that some of the words had been incorrectly translated the first time due to the complexity of the terminology and the subtleties of meaning. For example, *siekteleer* had been used as a translation of ‘nosology’. However, the direct translation into English of *siekteleer* is ‘pathology’, not ‘nosology’, which has a very different meaning. It was therefore decided to use the Anglicised Afrikaans word *nosologie* instead. A second example was the use of the Afrikaans word *syferblindheid* as a translation for ‘spatial acalculia’, which literally translates to ‘number blindness’. Given that ‘number blindness’ is clearly not correct, the Afrikaans term *ruimtelike rekenafasie* was instead used, which translates the term precisely.

Once these terms had been substituted, the panel of neuropsychological experts was consulted to ensure that they were satisfied that the translated terms still retained their precise clinical meaning when each term was back-translated.

Another issue encountered with the Afrikaans translation — a product of acculturation — was that the Afrikaans dialect spoken in rural areas is subtly different from Afrikaans spoken

in urban settings, in the sense that urban Afrikaans is more contaminated with English. In other words, urban Afrikaans is often spoken with English word substitutions and many traditional Afrikaans words are not frequently used (personal communication with Anthea Adams). As a consequence of this key observation, extra care was taken to ensure that the Afrikaans translation of the GSNSB could be understood by both rural and urban speakers.

The second major difficulty raised by the translators when translating the body of the GSNSB Prototype was that many of the phrases used in their respective languages carried different idiomatic meanings to the English. Mitrushina et al. (2005) highlight this as a potential difficulty when translating neuropsychological tests, given that the use of idioms and expressions differs across cultural groups.

The solution to this problem required the author to sit with the panels of translators once their first drafts were complete to determine which possible phraseology would carry the same English meaning. The success of this task was of vital importance — the GSNSB's instructions needed to be precise, guiding the administrator of the tool to an accurate appraisal of the patient's neurocognitive deficits in the absence of a specialist neuropsychologist. Any ambiguity in the meaning of the administration instructions for the tool would seriously compromise its clinical accuracy. One of the many pertinent examples of differences in idiomatic meaning can be found in the 'Language Function' section in the isiXhosa translation of the GSNSB. Here, the heading 'Assessment of Naming' was originally translated as *Imvavanyo zoThiyo-magama*. This was found to be problematic because in isiXhosa culture and vocabulary 'naming' is used only in the context of naming persons, the only way to translate this directly being *zoThiyo* (that is, to name a specific person). This would have confused both isiXhosa patients and clinicians administering the

GSNSB, as they would have assumed that the task required the patient to give human names (such as Peter or Siphon, for example) to the items pointed out at the bedside (for example, the pillow, sheet, spectacles). It was decided that the only way to overcome this problem of idiomatic meaning would be to use a broader, more descriptive term indicating that the patient was to specify the name of the item itself as opposed to giving it a human name. Therefore, the more appropriate isiXhosa word *zeZibizo* was used instead.

Phase Two: Further Modifications and Translation Work

Based on the findings of the initial pilot study, six of the nine newly developed tests were approved to be immediately included in the GSNSB without further modification. However, the Naming Test, Township Fire Story and 3-D Analysis Test all required further modification and re-piloting prior to their inclusion ahead of the validation. As a consequence of this process, more work was required in translating the changes made to these tests post-piloting in both Afrikaans and isiXhosa. For example, from the Township Fire Story, the word *week* was changed to *year*, and the final parts of the Story were changed from “...*while trying to put out the flames and rescue their possessions from the fire. It took eight hours before the fire was extinguished*” to “...*while trying to save their possessions. In rescuing a child who was trapped in a shack, a woman broke her arm*”.

Phase Three: Proof-reading and Formatting the GSNSB

Before the GSNSB was ready for use as a workable tool, final preparatory work was needed. Here, a crucial aspect was the proof-reading of the new English, Afrikaans and isiXhosa versions (see Appendices B, C and D for these final proof-read versions of the GSNSB). For this process, which lasted approximately five months, the translators worked together in proof-reading their respective languages. Having two translators working together to proof-

read the isiXhosa and Afrikaans versions of the GSNSB strengthened the validity of the process. In addition, an English editor was brought in to edit and proof-read the English version as, up to this point, this had not yet been done formally.

In conjunction with the proof-reading process, the author also carried out extensive work on the formatting of the GSNSB. The initial GSNSB was created in a rough form in the Excel computer program. As a result, many formatting changes were required before the battery could be printed and utilised. The page breaks, page numbering and margins all had to be adjusted as a result of the extra space required to accommodate the new material. Many of the scoring blocks forming part of the decision-trees required re-sizing as a result of the changes in space. It was also necessary to add new headings. For example, the Naming Test was not part of the GSNSB Prototype, and a new heading was thus required in the 'Memory' section, along with additional instructions pertaining to this test. All the newly developed tests were added to the GSNSB, and their new names replaced those of the original tests.

Along with these heading and name changes, alterations to certain of the cut-off scores, in accordance with the recommendations of the re-piloting study (see Chapter Five), were required. It was also necessary to adjust the test instructions that corresponded with the changes to the cut-off scores. Specifically, the FAS Test, the 3-D Analysis Test and the Township Fire Story had their cut-off scores and instructions changed. To accommodate the Township Fire Story, the questions pertaining to the story as part of the 'Mesial' subsection of the 'Executive Function' section needed changing. As the original questions referred to the content of the 'Babcock Story', new questions referring specifically to the content of the new story had to be included. For example, the questions from the GSNSB Prototype: "*Was there*

a fire in the story?” and *“Who put the boy under the bridge?”* were changed to *“Was there a flood in the story?”* and *“Who put the child in the shack?”*

Along with the new tests, new appendices were added at this point to accommodate those tests that did not form part of the core body of the GSNSB. The tests added as appendices were: the Township Fire Story, the Washing Line Picture Test, the 3-D Analysis Test and the Naming Test. These additions meant that new references to these appendices needed to be added to the content of the GSNSB, and the re-numbering/ordering of appendices needed to conform accordingly. As a final inclusion to these new appendices, additional scoring instructions to accompany some of the new tests (that is, the Naming Test and Washing Line Picture Test) were added.

Finally, other formatting changes included correction of the arrows pointing to the diagrams and scoring boxes, which were originally misaligned, or had become misaligned when new material was included. Further work was also carried out on the headings in the GSNSB, which appear in various colours — it was discovered that there were inconsistencies in the colour coding of the headings (and certain of the diagrams), which were adjusted to achieve uniformity throughout. All in all this process, as with the proof-reading, took approximately five months to complete and was complicated by the fact that the many of these formatting changes were different for each of the three language versions of the GSNSB, due largely to the different phrase length of Afrikaans and isiXhosa relative to English and to each other. Therefore, each of the three language versions was formatted separately.

CHAPTER FOUR: CREATION OF NEUROCOGNITIVE TESTS FOR THE SOUTH AFRICAN CONTEXT

In the academic discipline, South African students are largely trained in psychological ideas developed in North America and Europe. As a result, South African psychology fails to reflect South Africa's unique culture. It does not address the realities of the circumstances, mental functioning and problems of the majority of South Africans. Similarly, in the development of psychological tests, focus has largely been on the white population.

(Edwards & Louw, 1997, p. 27)

There is a dire need for culturally fair, diagnostically accurate neurocognitive tests for the South African context. "The field of neuropsychology is unprepared for the growth in racial, ethnic, and cultural diversity among those gaining access to cognitive assessment services and those participating in research studies. Very few neuropsychological measures have been properly validated for use among individuals who are not Caucasian, do not speak English, or lack a high school degree" (Manly, 2005, p. 278). This problem facing the discipline of neuropsychology in general is even more acute in the South African context, where socio-economic hardship is widespread and the disadvantaged in terms of appropriate psychological services constitute the majority, not the minority, of the population. "Culture is a broad and overarching concept, a complex entity that can have ethnic, geographic, generational, linguistic, and social determinants" (Kotik-Friedgut, 2006, p. 43). A neuropsychologist needs to account for such factors if meaningful diagnoses are to be made in a cross-cultural setting — deciding which aspects of the patient's presentation are part of universal cognitive function and which are specific to the patient's culture (Fletcher-Jensen, Strickland & Reynolds, 2000).

Nell (2002) further cautions against the habits of mainstream neuropsychology in assuming that neurocognitive functions are universal across all humans and not influenced by culture or education. The existence of certain universal physiological phenomena such as hearing or vibration sense, which exist within the medical realm, does not mean that cognition is also uninfluenced by culture or education, just because both fall within the ambit of medicine where the universality of physical diagnoses is taken for granted (Nell, 2000). In the South African context, Foxcroft and Roodt (2005) urge for the publication of research into the cultural appropriateness of tests and for the development of more broadly applicable and culturally sensitive tests, using appropriate cross-cultural methodologies. As Uzzell (2007, p. 2) reiterates, “Whereas questions contained in a neuropsychological instrument may be understood and suitable in the culture in which it was developed, it may have a different meaning or no meaning in another culture. Principles of neuropsychology by necessity must include cross-cultural methods that define and acknowledge diversity and universals between two or more different cultures or cultural areas”.

Anderson (2001) highlights the urgent need in South Africa for well-planned and coordinated studies to address the problems caused in neurocognitive assessment by the extent of cultural diversity in the country. Such problems manifest in the form of the high number of false positives found during neurocognitive assessments. These problems were clearly demonstrated in a study examining the performances of 20 health control patients from a variety of cultural and educational backgrounds, using the Controlled Oral Word Association Test (COWAT), the Grooved Pegboard, the Digit Symbol Modalities Test (SDMT), the Auditory Verbal Learning Test (AVLT) and the Paced Auditory Serial Addition Task. Here,

the results showed an unacceptably large number of controls (25 percent) being identified as impaired (Anderson, 2001).

With specific reference to psychometric testing, Shuttleworth-Jordan (1996) argues that in the South African context, given the dynamic nature of socio-cultural influences, a careful discrimination should be made between those pre-existing tests that still have diagnostic relevance and those that are culturally inappropriate, being careful not to assume that all western-derived tests are automatically problematic. She argues that the emphasis on developing new tests should be aimed at neuropsychology being practised in settings with illiterate or semi-illiterate individuals, especially those from rural areas — in other words, with individuals who have not in any way benefitted from acculturation with westernised populations (Shuttleworth-Jordan, 1996). Such individuals represent a large portion of Groote Schuur Hospital's patient population.

To the author's knowledge, no adaptation of any individual 'bedside' neurocognitive tests has ever taken place in the South African context. When cultural issues have in fact been addressed, as has occurred with psychometric assessment batteries (especially intelligence tests, such as the Wechsler Adult Intelligence Test (WAIS)), typically it is not the content of the tests, but rather the norms by which the patients are measured, that is changed and the measures simply translated (Nell, 2000; Van Eeden & De Beer, 2005). One example of a study that attempted to address cross-cultural issues in neuropsychology in South African was conducted by Tollman and Msengana in 1990. Here an attempt was made to translate and adapt Luria's Neuropsychological Investigation into Zulu (a South African indigenous language). This study encountered problems because Zulu speakers' culture, language and education confounded the results by making interpretation of performance extremely

difficult, and they struggled on many of the test items, despite the fact that these had been translated and certain content replaced with South African terms.

Culture: Its Influence on Neurocognitive Tests

The concept of culture is extremely difficult to define for many reasons, not the least of which is the fact that culture coexists with language, education and socio-economic status. In South Africa this complexity is taken to the extreme, where socio-economic status varies greatly, 11 official languages are spoken, many of the population are poorly educated or illiterate, and people live in both rural and urban settings. Swartz (1998, p. 7) defines culture as being about: “the process of being and becoming a social being, about the rules of a society and the ways in which these are enacted, experienced, and transmitted”. Accordingly, culture is viewed as being in constant flux. Others have incorporated the following terms into their definition of culture: knowledge, beliefs, learned traditions, ways of thinking, attitudes, values, roles, instruments, a collection of ideas, a way of life, habits, skills, motivations, and as a design for living (Ardila, 2007; Betancourt & Lopez, 1993; Harris, 1983; Kennepohl, 1999; Swartz, 1998).

Rosselli and Ardila (2003) highlight how clinical neuropsychology has achieved little in attempting to rectify cultural problems in testing. Nell (2000, p. 3) identifies the crux of the matter in emphasising how: “[p]sychological tests are conceived and standardized within the matrix of Western culture”. Neuropsychology has largely been practised by individuals educated and trained in Europe, North America and the former Soviet Union, with the Eurocentric view that universal principles could be applied to the understanding of mental processes (Kotik-Friedgut, 2006). For example, “To copy nonsense figures (e.g., Rey-Osterrieth Complex Figure) can be suspicious for many people. It may be [a] relevant item

for an American school child, but absurd for someone living in a nonpsychometrically oriented society” (Ardila, 2007, p. 26).

In the last decade, terms such as *cross-cultural neuropsychology* and *cultural neuropsychology* have begun to emerge in the literature (Ardila, 1995; Fletch-Jensen et al., 2000; Kennepohl, 1999; Uzzell et al., 2006). The field is slowly beginning to acknowledge that the influences of culture require the neuropsychologist to learn to differentiate between those concepts that are universal and those that apply to unique situations, while bearing in mind that brain functioning and culture form a complex interaction (Fletcher-Jensen et al., 2000; Kotik-Friedgut, 2006; Rosselli & Ardila, 2003). “It is well established that culture has a considerable influence on the development of the brain and its functions. The actual mechanisms, scope, and consequences for neuropsychological diagnostics, however, still require clarification” (Kotik-Friedgut, 2006, p. 44).

Attempts to produce ‘culture-free’ tests have been made internationally, yet many regard this task as impossible (Rosselli & Ardila, 2003). Many take the view that the use of non-verbal tests will nullify the influence of culture. However, due to cultural differences such as the perception of representative drawings and pictures, differences across cultures in the acquisition of cognitive strategies with which to approach problems, and varying levels of test wiseness, success has never been achieved (Rosselli & Ardila, 2003).

In South Africa, virtually no attempts have been made to create culturally fair neurocognitive tests, and the majority of psychometric tests in use have not been normed on South Africans (Anderson, 2001; Foxcroft & Roodt, 2005, Griessel, 2005). Consequently, all the neurocognitive tests in use are European or American in origin. This has resulted in high

false-positive errors in clinical practice, with normal individuals being identified as impaired on account of the inherent bias of the tests, and patients' performances being compared to inappropriate norms (Anderson, 2001).

The influence of culture on specific tests is outlined in later sections of this chapter.

Education: Its Influence on Neurocognitive Tests

The effects of education on neurocognitive test performance are well documented, with many authors highlighting that individuals with higher levels of education tend to perform better on neuropsychological testing, especially those tests involving verbal responses (Ardila, Rosselli & Rosas, 1989; Matarazzo, 1972; Nell, 2000; Rosselli, Ardila, & Rosas, 1990). A review of the literature reveals that the effect of education on nonverbal tests is less well documented (Rosselli & Ardila, 2003). A key moderator of test performance in heterogeneous societies is 'test-wiseness', referring to an individual's prior exposure to westernised testing procedures, mainly through formal education, which acquaints him/her with the concepts and skills required to complete paper and pencil tests, to understand the concept of timed-tasks, to have acquired motivation to complete tests, and to be able to sit still and complete tasks in this context (Nell, 2000). Without such skills, an individual will find neurocognitive testing extremely challenging.

Perez-Arce and Puente (1996) highlight how often in multicultural settings education level and experience are often confounded with cultural assumptions regarding test performance, where poor performance can actually simply be explained by differences in education level rather than cultural influence. This problem is further exacerbated by the fact that, "[t]est taking skills are so taken for granted in Western society that it is difficult to grasp the extent

to which they are absorbed rather than explicitly taught” (Nell, 2000, p. 3). In South Africa, the differences in the quality of education received by people from different cultural backgrounds means that a patient’s level of education must be distinguished from the quality of education he/she has received (Foxcroft & Roodt, 2005; Nell, 2000). In other words, two individuals who both have 10 years of education will not necessarily be on an equal footing in terms of the functional capacity associated with that particular level of education from a western point of view.

The influence of education on specific neurocognitive tests is outlined in the sections of this chapter to follow.

The problems alluded to in Chapter One regarding the cultural inappropriateness/bias of some of the tests chosen for inclusion in the Groote Schuur Neurocognitive Screening Battery (GSNSB) Prototype are elaborated on in detail in the following sections of this chapter. A careful examination of each of the relevant tests initially revealed that this cultural bias (mainly Anglo-American in nature) served to undermine the clinical utility and effectiveness of these tests in the South African context. This jeopardised the tests’ ability to provide culturally fair and appropriate testing, with meaningful diagnostic outcomes. The examiner can be easily misled into thinking that a patient’s poor test performance is due to a lesion when it may actually be the consequence of the cultural inappropriateness of the test items. It was through our day-to-day clinical use of the tests in the South African context over the past five years that their cultural inappropriateness and bias had been clearly demonstrated and documented. The author was already able to identify some of the most problematic items even before a systematic analysis of the test items was undertaken.

The GSNSB is divided into five primary domains of neurocognitive assessment: Orientation; Memory Function, Language Function, Spatial Cognition, and Executive Function (see Appendix A and/or Appendix B). As previously mentioned, a selection of international neurocognitive tests, deemed to be best suited for examining the domains of functions to be assessed, were chosen from the array of neuropsychological tests available for inclusion into the GSNSB. The tests were well-known 'bedside' tests in daily use in the clinical practice of the Groote Schuur Hospital neuropsychologists and were long established as being diagnostically effective in the clinical setting. The overriding problem of cultural inappropriateness/bias identified with some of these tests in the GSNSB is systematically outlined below. The problem of 'cultural bias' is further compounded by the widely varying levels of education of our patient population, by the variety of languages spoken, and by urban versus rural influences and processes of acculturation.

Memory Function

Memory function encompasses three primary neurocognitive processes: encoding of memory, retrieval of memory, and working/short-term memory (the process of holding information in conscious thought). Three distinct types of amnesia are seen clinically: mesial temporal amnesia, diencephalic amnesia and 'frontal amnesia'. Mesial temporal amnesia refers to a problem with encoding, characterised by the inability to lay down new memories, a deficit of episodic memory, with anterograde amnesia and retrograde amnesia for recent events (Butters & Miliotis, 1979; Kolb & Whishaw, 2003; Solms & Turnbull, 2002; Walsh & Darby, 1999). The patient's remote memory remains intact, along the temporal gradient, and he/she is aware of his/her memory problems (Solms & Turnbull, 2002). The hippocampi are primarily responsible for the encoding of new memories, and hence lesions resulting from

Alzheimer's disease, hypoxia, epilepsy and posterior-communicating (P-Comm) aneurysms, typically result in an amnesic picture of this quality.

Diencephalic amnesia is typically characterised by confabulation and a problem with the retrieval of memory. The patient is unaware that he/she is amnesic and his/her recall is fragmented and disorganised, with chronological confusion (Solms & Turnbull, 2002). The diencephalic structures — the hypothalamus and thalamus, along with basal forebrain nuclei and adjacent frontal cortex — all play a key role in these memory processes (Fitzgerald & Folan-Curran, 2002). This type of amnesia is typically seen with Korsakoff's Psychosis, which preferentially affects the diencephalic structures that also have connections with the adjacent frontal cortex (Damasio, 1979). It is also seen with vascular lesions (Butters & Miliotis, 1979; Joynt & Shoulson, 1979; Walsh & Darby, 1999).

Frontal 'amnesia' is characterised by dysfunction with the retrieval process of memory, and is typically seen in the context of generalised executive impairment. It is referred to as an 'amnesia', although more accurately it is the product of poor control and supervision of memory processes secondary to executive impairment. In other words, the primary memory processes are themselves intact, but it is the faulty retrieval of memory due to executive impairment that presents as apparent amnesia (Kolb & Whishaw, 2003; Luria, 1973; Solms & Turnbull, 2002; Walsh & Darby, 1999). This clinical picture is seen with pathologies resulting in lesions of the frontal lobes, such as frontotemporal dementia, tumours, injury resulting from motor vehicle accidents, and other head trauma, which can take the form of either closed or open head injury (for example, gun shots or blunt force injuries).

Testing Memory Function: The Tests and Their Weaknesses

The tests of Memory Function included in the GSNSB Prototype were the Rey Complex Figure (visual memory), the Babcock Story/Wechsler Memory Scale (WMS) story (verbal memory), the 4 Hidden Objects Test, and the Digit Span Test.

Rey Complex Figure

The Rey Complex Figure — also known as the Rey-Osterrieth Complex Figure, the Complex Figure Test (CFT) and the Rey Figure — was designed by Andre Rey in 1941 primarily to test visuospatial constructional ability and visual memory (Lezak, 2004; Mitrushina et al., 2005; Spreen & Strauss, 1991). Qualitative observation of the patient's approach to the test can also be used to test for neglect in the case of a patient with a right hemisphere lesion (such a patient will leave the left proportion of the figure relatively incomplete). In addition, this test can be used to test for the executive signs of poor planning and strategy to the task, in the case of frontal lesions (the patient might turn the figure on its side and/or begin with the detail rather than first drawing the overall gestalt) (Kaplan, 1988). Osterrieth developed the test further by standardising it and creating the 18-item scoring system, comprising 36 points (Lezak et al., 2004).

The test requires the patient to copy the figure, which is then removed, followed by the patient having to re-draw it from memory (see Appendix A). This task involves incidental learning, as the patient is not told that he/she is required to remember the figure at any point preceding the task of re-drawing it from memory. A delayed recall is also requested, usually after 30 minutes (Spreen & Strauss, 1991). The Rey Complex Figure, which has been shown to be a reliable test displaying good internal consistency, is widely used, especially since it

has a range of clinical applications (Berry, Allen & Schmitt, 1991; Mitrushina et al., 2005; Pimental & Ross, 2003; Rapport, Charter, Dutra, Farchione, & Kingsley, 1997). The Rey Complex figure did not require adaptation.

Babcock Story

The Babcock Story Recall was designed by Babcock in 1930 and is a constituent of the Wechsler Memory Scale (WMS) Logical Memory section, along with the Anna Thompson Story (Mitrushina et al., 2005). The story was designed as a test of verbal memory and was chosen for the GSNSB on account of its effective utilisation in the widely used, well-known Wechsler's Memory Scale (Horner, Teichner, Kortte & Harvey, 2002; Lezak et al., 2004; Mitrushina et al., 2005). The story contains 21 scored units of information, and is administered to the patient in a number of different ways (Lezak et al., 2004). The patient is initially read the story then asked to recall it immediately. Thereafter, the story can be read for a second time, with a second recall, or the patient can be asked to recall the story after a 10- (sometimes 30- minute) delay without an additional reading (see Appendix A).

The Wechsler Memory stories (the Babcock and Anna Thompson stories) required adaptation due to their content and language, as they existed only in English. Afrikaans and isiXhosa versions of these stories were needed, as it was difficult to interpret the performance of a person whose first language is not English. Additionally, the content of the stories — the descriptions, the references to places overseas and the Eurocentric activities of the characters — proved especially challenging for many South Africans who are often unfamiliar with content of this nature. A typical example of this bias in the Wechsler Memory stories is: *“Anna Thompson of South London, employed as a cook in a school canteen, reported at the police station that she had been held up on the High Street the night before and robbed of*

fifty-six pounds”. Another pertinent example from the Babcock Story is: “... *a river overflowed in a small town ten miles from Albany*”. From clinical experience, the average South African patient is confused by unfamiliar terms such as *pounds, South London, miles, canteen, High Street* and *Albany*, which are not found in the South African context.

Limited literature is available on the influences of culture and education on the Babcock Story when used in the clinical setting, largely because the test forms part of the Wechsler Memory Scale rather than serving as an individual test (Horner et al., 2002). Nell (2000), however, cites the data from seven studies across America and South African samples from the World Health Organisation Neurobehavioural Core Test Battery (WHO-NCTB), examining Logical Memory across levels of education. The results showed how the well-educated South Africans performed equally or better than Americans, while the poorly educated in the samples performed a full standard deviation below the well educated (Nell, 2000).

In general, the shortcomings of the Western-designed Wechsler Memory Scale (WMS), in terms of its cross-cultural appropriateness as a psychometric battery, implicitly apply to the Babcock Story too.

4 Hidden Objects Test

The 4 Hidden Objects Test is a memory test that was incorporated in the GSNSB Prototype because it offers a quick and simple assessment of both immediate and long-term memory. Due to its relative level of simplicity, the test is ideal for situations where the patient is deemed too amnesic for more complex testing, or when the patient is extremely poorly educated or lacks test wiseness. The test requires any four objects to be hidden from the

patient, first in one location and in later trials in various locations (see Appendix A). This form of memory assessment is widely used and has appeared in the Terman and Merrill (1973) Stanford-Binet tests and in mental status examinations, such as Strub and Black's MSE (Lezak et al., 2004).

The 4 Hidden Objects Test required adaptation as the objects frequently chosen for use, such as, for example, a *cellular phone*, proved culturally inappropriate and often confusing to the patients. Such items left the patient guessing, most often incorrectly, as to what the object was. Semantic errors were therefore often encountered, such as a *cellular phone* being mistaken for a *calculator*. Naturally, such added complexity detracts from the patients' ability to recall the items correctly. Logie (1995) stresses that, if items are unfamiliar to the patient, recall is likely to be impaired. Additionally, as the items being used varied each time, it was necessary to develop a standard set. No literature is available on the effects of culture or education on this test, yet the general problem of item familiarity applies to this test, as it does to other neurocognitive tests reviewed in this study.

Digit Span Test

The Digit Span Test is a test of working memory, found in the Wechsler Memory Scale (WMS). The test comprises a series of sequences of random numbers, with each trial increasing in length from three digits through to eight, which the patient is required to repeat back when administered one line at a time (see Appendix B).

The Digit Span Test required investigation because clinical experience had found that many patients, especially those whose first language was not English, were struggling with the task

when, on the basis of their lesions, they were not entitled to have a working memory problem. A number of studies have examined the effects of language on the Digit Span Test. For example, Puente and Ardila (2000) highlight that the Digit Span Test may draw on other cognitive processes when administered in languages other than English, due to variations in the number of syllables occurring across languages. With isiXhosa and Afrikaans both being spoken at Groote Schuur Hospital, there was a strong possibility that this issue was affecting patients.

A review of the number of syllables in numbers one to nine in each of the three respective languages revealed that, in isiXhosa, the words total 29 syllables altogether, while in Afrikaans, the words are very similar in length to the English words, totalling only 12 syllables. Several studies examining bilingual subjects have reported a negative linear relationship between speech rate increase and memory span (Baddeley, Thomson & Buchanan, 1975; Chincotta & Underwood, 1997; da Costa Pinot, 1991; Hitch, Halliday & Littler, 1989). These findings have important implications for those settings in which individuals speak more than one language, as they suggest that digit span might be better if administered in the language that has the fastest pronunciation rate.

The effects of education on the Digit Span Test have also been documented. A number of studies on different cultural groups in various settings, including Europe, America and South America, have shown individuals with low levels of education to perform worse than better-educated individuals on the test (Ardila & Rosselli, 1989; Garcia-Morales, Gich-Fulla, Guardia-Olmos & Pena-Casanova, 1998; Karakas, Yahn, Irak & Erzenin, 2002; Kaufman, McLean & Reynolds, 1988; Pacaud, 1989). The WHO-NCTB data cited by Nell (2000) (see 'Babcock Story' section above) produced the same findings on the Digit Span Test as they

did with the Babcock Story: the well-educated participants fared better on the test than those with a poor level of education. The effect of the language in which one is taught has also been identified as an important factor in Digit Span Test performance. Here, it is believed that individuals fair better on the Digit Span Test if it is administered in the same language in which they were taught mathematical abilities at school (Chincotta & Underwood, 1997).

In the clinical setting, the test is usually administered using up to seven digits, which is considered to be a normal working memory span (Joynt & Shoulson, 1979; Solms & Turnbull, 2002). Therefore, the author decided it was necessary to conduct an investigation as to whether test performance improved if the test was administered in Afrikaans or isiXhosa. Additionally, a standardised set of digits needed to be created for inclusion in the GSNSB.

Language Function

Principally a left hemisphere function, language can be broadly divided into production of speech, audioverbal comprehension, naming ability (word-finding difficulty), ability to repeat, and the ability to read and write. Among the more common aphasia seen clinically are Wernicke's aphasia, Broca's aphasia, and Global aphasia, although the spectrum of impairment is broad and many classifications of aphasia exist, with mixed and varying clinical presentations commonly found. Wernicke's aphasia is characterised primarily by poor comprehension, poor naming ability and poor repetition, with fluent verbal output. Writing is usually impaired, with reading ability intact, despite the inability to comprehend what has been read. Wernicke's aphasia is characteristic of left hemisphere lesions to the posterior part of the superior temporal gyrus and the posterior perisylvian region (Benson, 1979; Kertesz, 1989; McKenna, 2004; Solms & Turnbull, 2002; Walsh & Darby, 1999).

Broca's aphasia primarily involves a deficit of speech production, and typically presents with non-fluent speech output, severe difficulty with repetition, naming difficulty and, on occasion, comprehension may not be fully intact. Reading and writing are usually also affected. Broca's aphasia results from lesions of the left hemisphere, localised to the posterior part of the inferior frontal convolution (gyrus), adjacent to the face motor cortex (Benson, 1979; Kertesz, 1989).

Global aphasia, often seen in very acute stroke, before the deficits resolve into another variant of aphasia, involves the impairment of both language comprehension and production, as well as virtually no ability to read, write or repeat sentences verbally (Benson, 1979). Global aphasia usually results from large lesions encompassing both Broca's and Wernicke's areas and large parts of the remaining perisylvian convexity of the left hemisphere (Benson, 1979; FitzGerald & Folan-Curran, 2002).

Testing Language Function: The Tests and Their Weaknesses

Testing of Language Function in the GSNSB Prototype included the Cookie Theft Picture/Cookie Jar Picture test to examine production of speech, and the Boston Naming Test (BNT) to assess naming ability.

Cookie Theft Test

The Cookie Theft Test comprises a kitchen scene from a 1970s middle-class home. The test forms part of the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan & Barresi, 2001). The depicted scene is designed to elicit a verbal description as to what is transpiring — the patient's verbal fluency, naming in spontaneous speech, and the presence or absence of paraphrasia (semantic or literal), are of primary interest to the examiner (Lezak, 2004;

Temple, 2002). Time is an important factor in this task, as the patient is given 60 seconds to produce a story about the picture, during which the quantity of speech output (number of words) is recorded (see Appendix A).

The Cookie Theft Picture required adaptation as our clinical experiences had found the scene, unfamiliar to many patients, to undermine patients' performances — many paused and hesitated when trying to name and describe unfamiliar objects within the kitchen scene such as the *stool* (commonly called a *chair*), the *sink*, and the general unfamiliar depiction of the cooking area. The importance of such mishaps cannot be overestimated, as it is crucial when assessing language deficits to ensure that possible difficulties found with the test material are not contributing to the quality of the patient's performance — such problems serve to confuse and possibly mislead the neuropsychologist, who is then not sure whether to attribute the poor performance to deficit of speech production or cultural bias. The influence of education on the test has never been adequately determined, yet individuals with high levels of education are generally at an advantage on tasks involving language (Mackenzie, Brady, Norrie & Poedjianto, 2007).

Boston Naming Test

The Boston Naming Test (BNT) is used primarily to assess the patient's naming (confrontation naming) ability through depicted line drawings. The items are sometimes also used to assess for visual agnosia. "Because of the high incidence of naming problems in aphasia as well as in other neuropathological conditions, virtually all aphasia examinations contain a naming task" (Sprenen & Strauss, 1991, p. 213). The test was designed by Kaplan, Goodglass and Weintraub in 1983. Sixty pictures are shown one at a time to the patient, who is asked to name each in turn (Lezak et al., 2004). The test is known to be highly sensitive to

the detection of naming difficulties and is one of the most widely used naming tests in psychology (Spreeen & Strauss, 1991). A large number of studies have found the test to have good diagnostic properties and to be able to discriminate between clinical and normal groups (Cahn, Salmon, Butters, Wiederholt, Corey-Bloom, Edelstein, 1995; Jacobs, Sano, Dooneief, Marder, Bell, Stern, 1995). The Boston Naming Test is also available in short, 30-item form, divided into either odd or even items, with every second item being administered (Lezak, 1995). Studies using the 30-item version have also found it to be sound in discriminating between control and patient groups (Fisher, Tierney, Snow & Szalai, 1999).

The Boston Naming Test required adaptation as frequent clinical experiences had revealed a number of items culturally inappropriate for the South African context, which served to undermine the patients' performances and confuse the examiner. The difficulty is that the patients (including those without naming deficits) are unable to name (or recognise) many of the culturally unfair items, and therefore either guess (usually a semantically similar item) or simply say they do not know. The problem this poses for the clinical assessment is obvious, as the purpose of the test is to detect aphasic deficits such as semantic paraphasia in brain-damaged individuals — such patients would not be able to name items due to a lesion, but rather only be able to identify the semantic category of the item and hence name a semantically related item. For example, the patient might see a picture of a park bench and say “chair” (which misleadingly suggests paraphasia), or see a picture of a pretzel and say “snake” (which misleadingly suggests agnosia). Without being able to discern between the two, the clinician is left with the predicament of not knowing the source of the patient's error — lesion or cultural bias. This problem manifests in many of the Boston Naming Test items and also occurs in the Cookie Jar Test (mentioned above), which is also designed to look for paraphasic errors. Ardila (2007) adds credence to these clinical observations by emphasising

how many of the test's items are America-specific and therefore culturally problematic when used in cross-cultural settings.

Other authors have also highlighted the cultural limitations of this test. Authors using the test in Korea have found many of the items to be culturally inappropriate (Kim & Na, 1999).

Studies on Maoris in New Zealand, on Australians, Hispanics and African-Americans have also found scores in these populations to be significantly lower when compared to Europeans and North Americans due to the problematic American-derived items (Barker-Collo, 2001; Brauer Boonea, Victor, Wen, Razani, & Ponton, 2007; Worrall, Yiu, Hickson, & Barnett, 1995). As further evidence of this problem, some studies have attempted to correct the cultural inappropriateness by adapting versions of the test for specific contexts. For example, Cruice, Worrall and Hickson (2000) modified some of the test's items to make it more suited to the Australian population — the *beaver* and *pretzel* items were removed and replaced with *pizza* and *platypus* items, which are more recognisable to Australians.

Numerous studies have also found education to influence test performance on the Boston Naming Test, frequently finding that individuals with lower levels of education perform lower on the test and with more variability (Delouche, Hannequin, Dordain & Perrier, 1996; Hawkins & Bender, 2002; Heaton, Avitable, Grant & Matthews, 1999; Mitrushina et al., 2005; Saxton, Ratcliff, Munro, Coffey, Becker & Fried, 2000). Finally, various sources have found age to be a factor in Boston Naming Test performance. This occurs typically in older populations from the age of 60, where more variability is seen, standard deviations increase and performance decreases. This trend is more significant from the age of 70 (Ardila, 2007; Lansing et al., 1999; Ross, Lichtenberg & Christensen, 1995; Spreen & Strauss, 1998; Tsang & Lee, 2003; Welch, Doineau, Johnson & King, 1996).

Spatial Cognition

Impairment of spatial cognition can result in constructional apraxia, spatial acalculia, dressing apraxia, difficulty with the perception of two and three dimensional space, topographical disorientation, anosognosia, emotional indifference, and unilateral spatial neglect (Devinsky, 1992; Heilman, Watson & Valenstein, 1979; Solms & Turnbull, 2002).

The combination of anosognosia, neglect and deficits of spatial cognition and perception is known as the Right Hemisphere Syndrome, seen with lesions of the right hemisphere, but typically involving the inferior parietal lobe (the angular gyrus) along with the supramarginal gyrus and part of the adjacent superior temporal gyrus (Devinsky, 1992; Walsh, 1999).

Testing Spatial Cognition: The Tests and Their Weaknesses

In the GSNSB Prototype, the tests for the Spatial Cognition section can be broken down into those used for visuospatial analysis, those testing for neglect and those testing for anosognosia. The Rey Complex Figure is used for assessing for constructional apraxia, while the Cube Analysis Test is used for assessing perception. The tests in the GSNSB Prototype used for testing neglect included the Rey Complex Figure (see above) and the Scene Drawing Test.

Cube Analysis Test

The Cube Analysis Test, also known as the Block Counting task or Cube Counting, forms part of the Stanford-Binet Scale (Lezak et al., 2004). The test, which is used in the GSNSB for the testing of spatial perception, comprises 14 items in the form of two-dimensional drawings of three-dimensional piles of blocks, with some of the blocks hidden from view

(Lezak et al., 2004). The patient's task is to count how many blocks constitute each item of each of the 14 items in turn (See Appendix A).

The Cube Analysis Test initially required adaptation primarily for copyright purposes, although the neuropsychologists' clinical experiences with patients had shown that many found the test to be very difficult, with them struggling to understand the concept of three-dimensions. These clinical observations are consistent with the literature, which highlights that neurocognitive tests of visuospatial function are typically culturally biased (Ardila & Keating, 2007; Rosselli & Ardila, 2003; Sugarman, 2007). Jahonda (1981, as cited in Nell, 2000), concluded that the perception of three-dimensional space is not universal and varies across cultures, with the methods used in testing largely dictating the success of the performance on formal testing. By way of example, some African cultures are found to perform poorly on visually specialised tasks, suggestive of the fact that in their daily activities these cultures prioritise other sensory modalities over vision in their interaction with the external world (Berry, 1965; McFie, 1961; Wober, 1966).

In terms of the effect of education on the Cube Analysis Test, Lezak (1995) highlights that it is highly susceptible to level of education and is widely considered to be one of the more difficult visuospatial tests.

Scene Drawing Test

The Scene Drawing Test, which comprises a westernised depiction of a double-storey house with picket-fence, surrounded on each side by two different types of tree, is used to elicit neglect (see Appendix A). This test required adaptation to a more typical South African theme on the grounds that patients often struggle to recognise the depiction. A more

culturally appropriate scene would facilitate South African patients' ability to focus on the task at hand (copying the picture) rather than puzzling over the perhaps unfamiliar scene, especially as drawing is often difficult for illiterate patients who are unfamiliar with pen and paper tasks. It must be constantly remembered that many of the patients at Groote Schuur Hospital are very poorly educated.

Executive Function

Executive function can be simplified for academic purposes by division into four functional domains: mesial, orbital/basal, dorsolateral and deep white matter (subcortex). The dorsolateral convexity can be further subdivided into pre-motor and pre-frontal cortex.

Executive function is seen as presiding over all cognitive function in a hierarchical way, and comprises many vital functions (Damasio, 1979; Kolb & Whishaw, 2003; Luria, 1973; Walsh & Darby, 1999). The mesial cortex is primarily responsible for the selective application of voluntary arousal. Damage leads to obtunded (clouded) consciousness, confabulation, and ideational perseveration (Solms & Turnbull, 2002). The orbital/basal cortex is primarily responsible for inhibition and response suppression, with damage leading to disinhibition, social inappropriateness, distractibility and impulsiveness (Damasio, 1979; Miller, Darby, Benson, Cummings & Miller, 2001; Walsh & Darby, 1999; Solms & Turnbull, 2002).

The dorsolateral cortex governs the subordination of goal-directed behaviour to verbally regulated programmes, for motor tasks in the case of the pre-motor cortex, and for abstract thought and problem-solving in the case of pre-frontal cortex. Damage to the dorsolateral convexity produces concrete thought, loss of problem-solving ability, inability to shift sets, disorganised thought processes and lack of self-critical awareness. The deep white matter is implicated, through the areas of cortex which it connects, in the governing of spontaneous

initiative and curiosity, with damage resulting in adynamia, aspontaneity and impersistence, and even akinetic mutism in severe cases (Solms & Turnbull, 2002; Walsh & Darby, 1999).

As alluded to above in relation to frontal ‘amnesia’, lesions of the frontal lobes are often diffuse, given the nature of the pathologies that typically result in such lesions. Therefore, more often than not, a complex picture of executive function emerges, which is not localisable to one specific area of cortex.

Testing Executive Function: The Tests and Their Weaknesses

The testing of Executive Function in the GSNSB Prototype was broken down, broadly speaking, into those tests that assess frontal cortical function (mesial, orbital/basal, and dorsolateral) and those that assess frontal subcortical (deep white matter) function. Deep white matter function is assessed using is the Controlled Oral Word Association Test (COWAT or FAS Test). Mesial cortex function is assessed using the Babcock Story (also used to assess Memory Function). The testing of the orbital/basal region again involves the FAS Test. Dorsolateral function is assessed by various tests, none of which were initially identified as problematic by the neuropsychologists.

The FAS Test

The FAS Test is a language-oriented test that requires the patient to generate as many words as he/she can in one minute beginning with each of three specified letters in turn (see Appendix A). The test was developed by Benton in 1968, through examining letter frequencies occurring in the English language (Lezak, 2004; Ross, Calhoun, Cox, Wenner, Kono & Pleasant, 2007; Spreen & Strauss, 1991; Sumerall, Timmons, James, Ewing & Oehlert, 1997; Walsh & Darby, 1999). The letters ‘F’, ‘A’ and ‘S’ are most often used — so

much so that the test is often termed the FAS Test. These letters represent a descending order of frequency, thereby making the task slightly harder with each letter given (Lezak et al., 2004). The test's instructions specify that no repetition of words is allowed and that the patient may not use proper nouns (the names of people, places or products) (see Appendix A). The patient is given one minute to respond for each particular letter (Morris, 2004; Spreen & Strauss, 1991). The patient is also not allowed to use words that end differently, but begin with the same stem, that is, if the patient said *run*, they could then not also say *running* (Walsh & Darby, 1999).

Clinical experience with the FAS Test in South Africa has shown that it is potentially misleading to ask patients whose first language is not English to generate a sufficient number of words from the test as it currently exists. This is because the letters 'F', 'A' and 'S' do not appear with the same frequency in isiXhosa and Afrikaans as they do in English. What was most urgently required were new English, Afrikaans and isiXhosa versions, containing culturally appropriate letters representing an equivalent distribution of words common to the respective languages. Without such adaptations, the clinician is left to guess whether the patient's difficulties are due to a lesion or due to the difficulty of having to operate with an unfamiliar language. Nell (2000) calls for alternative forms of the test to be made in cross-cultural settings, using dictionaries to source equivalent word frequencies.

The available literature highlights that level of education is seen to influence COWAT performance, with individuals with less education tending to perform far more variably (Lezak, 2004; Loonstra, Tarlow & Sellers, 2001; Spreen & Strauss, 1991; Tombaugh, Kozak & Rees, 1999). Mitrushina et al. (2005) highlight how a concern which needs addressing is that in recent normative studies, norms for the test in individuals with little education are

lacking. The influence of culture on this test has also been reviewed, with cross-cultural findings indicating that, for example, Zulu South Africans with the same level of education fared far worse than Americans.

Rationale

In summary, having reviewed in the South African context the neurocognitive tests contained in the GSNSB Prototype originating from North America and Europe, it is important to emphasise once more the impact of the inherent cultural inappropriateness/bias in many of these tests. Clinical judgement plays a key role in assessing neurocognitive function in patients, especially when these patients come from a culturally diverse society. It is crucial to eliminate as many of the confounding variables as possible involved in the clinical judgement of patients' test performances, one such variable being the cloud of confusion created by culturally biased tests. Attempting to overcome this serious problem was one of the major goals of this study.

Five overall objectives governed this stage of the research. The first of these was to confirm that the problematic tests/test items had been correctly identified, with the help of cultural, language and clinical neuropsychology experts. This objective also involved investigating and understanding the inherent bias and specific problems associated with each test. This was achieved by examining the neurocognitive tests contained in each of the four key sections of the GSNSB (that is, Memory, Language, Spatial Cognition, and Executive Function), and confirming that these were indeed the tests requiring adaptation. The second objective was to create new, culturally and clinically appropriate South African neurocognitive tests to replace the problematic ones. The third objective was to translate the newly adapted tests into Afrikaans and isiXhosa (see Chapter Three). Fourthly, it was necessary to test the efficacy of

the adapted tests on neurocognitively intact control participants using a pilot study. Fifthly, the results of the pilot study made specific recommendations regarding further changes to be made to the tests prior to establishing the validity and reliability of the GSNSB (see Chapter Five for further changes made to certain tests). The outline that follows describes the complex processes that were followed in creating the new neurocognitive tests.

Methodology

Sample

The sample chosen for this pilot study comprised 30 neurologically normal, neurocognitively intact participants, who were screened to exclude any pathology or disease that might have neurocognitive consequences (See Appendix H for the Screening Sheet). This was done to ensure that the participants' results were not confounded by neurocognitive problems. The sample was randomly selected from the general patient population of Groote Schuur Hospital, and also from family members of patients, in order to achieve a sample that best represented the hospital's patient population in terms of culture, education, age and socioeconomic circumstances. The age range of the sample was 16 to 68 years ($M = 32.5$ years and $SD = 13.03$), while the number of years of education ranged from 1 to 16 ($M = 10.4$ years and $SD = 3.22$). The sample of 30 participants was further subdivided into three groups: first-language English speakers ($n = 10$), whose ages ranged from 17 to 39 years ($M = 28.7$ years and $SD = 8.2$), and who had an average of 12.9 years of education; first-language Xhosa speakers ($n = 10$), whose ages ranged from 16 to 42 years ($M = 30.1$ years and $SD = 8.6$), and who had an average of 9.7 years of education; and thirdly, Afrikaans first-language speakers ($n = 10$), who ranged in age from 18 to 68 years ($M = 39$ and $SD = 17.51$), and had an average of 8.73 years of education. The sample was subdivided into these three language

groups in order to be representative of the overall patient population of Groote Schuur Hospital.

In addition to the 30 neurocognitively intact participants, it was attempted to include some neurocognitively impaired patients in the sample to provide preliminary observations as to how patients, as opposed to controls, fared on the new tests in relation to the originals. At this point in the research, this was not imperative, as the main validation of the GSNSB was to follow. However, given time constraints and both the numbers and type of patient available at the time, only two right middle cerebral artery (MCA) stroke patients and three aphasics (two Broca's aphasics and one Wernicke's aphasic) were seen. All three aphasics were Afrikaans speaking, and comprised two females (aged 70 and 39) and a male (aged 40), who had eight, nine and 11 years of education respectively. All three aphasics had suffered demonstrable left middle cerebral artery (MCA) strokes on CT scan.

Materials

The materials used for this stage of the research included both the English version of the GSNSB Prototype (see Appendix A) and its two newly translated Afrikaans and isiXhosa versions, used to administer the tests' instructions. Further materials used included the nine tests earmarked for adaption, along with their newly translated versions. The material also included the newly created tests designed to replace the originals (as denoted in brackets): the Forward Digit Span (Auditory Span Test); the 4 Hidden Objects Test; the Babcock Recall Story (Township Fire Story); the Boston Naming Test (Naming Test); the Cookie Theft Test (Washing Line Picture Test); the Anna Thompson Story (Mary Selo Story); the Cube Analysis Test (3-D Analysis Test); the Scene Drawing Test (the Hut Drawing Test); and the Controlled Oral Word Association Test (FAS/BHP/NPS Test). Eight of the nine tests were

re-named, the exception being the 4 Hidden Objects Test (see Appendix A for the original tests and Appendix B for the relevant new tests).

Further materials used included the Patient Information Sheet (see Appendix J) and Consent Form (see Appendix I), and the Screening Sheet (see Appendix H) designed to exclude the presence of certain pathologies in the control participants. Finally, a specifically designed Scoring Sheet (see Appendix K) was used to ensure that all participants' test scores and their qualitative observations about the tests were captured. The Scoring Sheet ensured that all data were captured in a neat, concise way, and that no tests were accidentally overlooked.

Design

The specific design of this aspect of the research was intended to satisfy two primary goals: firstly, to use an approach adopting converging lines of evidence to guide first the implementation and then the evaluation of the test adaptations made; and secondly, to conduct a thorough pilot study that would draw on both qualitative and quantitative measures to evaluate of the efficacy of the test adaptations. To achieve the first objective, a panel of experts was assembled to offer advice on, critique, evaluate and provide feedback on the newly proposed test items. This panel comprised two experts in clinical neuropsychology from the University of Cape Town's Neuropsychology division, and three experts in both language and culture from the University of Cape Town's Southern African Languages Department, who were skilled linguists as well as being knowledgeable on the urban and rural cultures underlying their respective languages. Two of these three experts were from the Xhosa Department, while the third was from the Afrikaans Department.

Along with the experts' feedback, the final line of evidence drawn upon was the qualitative feedback (verbal accounts) gathered from the participants themselves as to how they had experienced the tests (both new and old), which versions they preferred and why, and what recommendations they offered to further improve the tests/items. These three primary sources of feedback provided a comprehensive and thorough approach to adapting the tests. This design — adopting a meticulous approach to the difficult problem of addressing 'cultural issues' within psychology — was intended to provide a study that was as exhaustive as possible, tackling this problem in a unique way never before attempted in the South African context.

To achieve the second objective of conducting the pilot study itself, the participants were tested with both the old and new versions of the selected tests, administered in a random and varying order, in one testing session. This allowed direct comparisons to be made between the participants' performances on the old and new versions in order to see which they performed better on and why. This pilot adopted a single-blind approach to the data collection phase, as the assessors administering the tests were unaware whether each participant they tested was one of the 30 neurocognitively intact individuals or a patient from one of the two patient groups. The assessors were only informed once all assessments were complete, when data analysis began, into which group each participant fell. Out of this process, both qualitative and quantitative data were obtained.

After the completion of the data collection for the pilot, the converging lines of evidence, including the qualitative observations of the participants, provided a final forum for evaluating the success of the newly adapted tests. Among the many advantages of such a design was that it allowed the test performances to be evaluated around two clear criteria —

firstly, whether the participants had performed better on the newly adapted tests than the original ones; and secondly, whether they were able to perform ‘normally’. Here, ‘normally’ can be taken to mean that the control participants scored close to full marks on the test, as neurocognitive bedside tests of this nature are not intended to challenge neurocognitively intact individuals. It must be remembered that these tests are not psychometric in nature and are therefore not scored so as to compare the individual to a standardised population norm — they serve merely to demonstrate whether the person has a deficit in a function that should be fully intact in a healthy individual.

Data Analysis

The data analyses for the pilot study adopted both qualitative and quantitative approaches. The types of analysis used were geared towards conducting two primary evaluations — whether the participants performed better on the new tests than on the original versions (and was the difference between the original versus the new test statistically significant); and, whether the participants were able to perform ‘normally’ on the new tests. Here, it must be remembered that these tests are not psychometrically based and are not intended to overly tax the cognitive functions of neurocognitively intact individuals.

For the qualitative aspects, careful observations as to how the patients fared on the original versus the new test were recorded using the specially designed Scoring Sheet. When a participant struggled with specific items/tests, these difficulties were recorded along with his/her verbal feedback regarding the problem. Additionally, the participants’/patients’ opinions of which test version they preferred, and why, were also examined and analysed. These observations were then viewed in conjunction with the descriptive statistics.

For the quantitative aspects of the data analysis, descriptive statistics were recorded to evaluate the numbers of participants performing successfully on each test version, as well as the breakdown of how many participants scored better on the original versus the adapted tests. Following from these descriptive statistics, various statistical tests were performed on the data. In order to evaluate the 4 Hidden Objects Test, the scores from immediate recall, and first and second delayed recall, were added together to form a total score for each of the two sets of objects. In investigating the impact of the use of these two different sets, the totals were subjected to a Friedman's rank test for k correlated samples (Friedman analysis of variance [ANOVA]) (heterogeneous variance).

The analysis of the Digit Span/Auditory Span Test involved the use of the raw scores as the dependent variable. A Friedman ANOVA was then applied to the isiXhosa and Afrikaans groups to see whether or not performing this test in their first language, as opposed to English, resulted in a significant difference in their score. An independent t-test by groups was used to see whether any significant between-group differences between the isiXhosa and Afrikaans participants existed when the tests were administered in English. Next, a one-way ANOVA was conducted to see whether any differences existed between the three language groups when the test was performed in the groups' first language.

Finally, an analysis was conducted using logistic regression to see whether the participants' level of education predicted their success or failure on the Digit Span/Auditory Span Test. This involved using education as the independent variable, defined continuously according to the number of years of education the participant had. The participants' Digit Span/Auditory Span Test scores denoted the dependent variable. Given that the English participants were only assessed once, the two scores obtained for each participant in the isiXhosa and

Afrikaans groups were averaged to achieve one score for each. Based on the GSNSB's scoring procedure, participants who scored less than six were coded as FAIL, while six or more was coded as PASS, which allowed a dichotomous variable to be yielded as the final dependent variable.

For the statistical analyses conducted on the remaining tests, t-tests and the chi-squared goodness-of-fit test were used to ascertain whether significant differences between the groups' score on the respective tests had occurred. Overall, these statistical findings, in conjunction with the descriptive statistics, and the qualitative observations and feedback, provided a thorough and comprehensive appraisal of the pilot study data.

Procedure

All tests were administered in the participants' first language. An exception to this was in the case of the Digit Span/Auditory Span Test, where one version was administered in English and one in the patient's first language (if not English). This was done to examine whether the language in which the test numbers were administered influenced the patients' ability to hold information in their short-term memory (STM). The participants were all tested in the same room at Groote Schuur Hospital, with the testers and their interpreters (when required) being 'blind' as to whether the participant was a control or a patient. Each testing session lasted approximately one hour. The order in which the tests were administered was randomised to ensure that the order of test administration did not influence the final results. Therefore, in some instances the old version of a test would be administered ahead of the new version and vice versa. For all the tests administered, the testing procedure followed the testing instructions as given in the GSNSB Prototype (see Appendix A for the exact instructions). The Scoring Sheet was specially designed to ensure that all data were thoroughly captured,

with space provided for the qualitative observations of the investigators as well as for the qualitative feedback/comments from the participants.

Written consent was received from all participants, who were informed of the voluntary nature of their participation and the fact that all results were anonymous and confidential. The data were securely stored and made available only to members of the research team and the translators. Ethical permission was granted by the Groote Schuur Research Ethics Committee.

When possible, assessments were tape recorded. The author gave all the assessors and interpreters prior training in proper test administration and scoring procedures, as well as the purposes of the GSNSB. The assessors were proficient in English but required the interpreters when assessing the Afrikaans and isiXhosa participants. All interpreters had worked previously at Groote Schuur Hospital (thereby making them familiar with working with medical clinicians and patients in this complex setting), and they had also translated the GSNSB, thereby adding to their overall experience.

An account of the precise procedure followed in constructing and piloting each of the nine tests requiring adaptation is now provided.

Memory Tests

Adaptation of the Digit Span: Auditory Span Test

In order to produce the Auditory Span Test, a standardised set of numbers was created by entering the numbers from one to nine into the STATISTICA statistical analysis program (see Appendix B). Simple random sampling of these numbers was then conducted increasing the

total number of percentage sampled each time in order to generate more and more numbers for inclusion in each subsequent number sequence. The resulting numbers were then ordered manually, to prevent numbers appearing in consecutive order — in ascending or descending order.

For each assessment, another set of numbers was generated on the spot, in the administration room, as is done with clinical neuropsychological assessments conducted at the bedside. This was done to avoid test-retest attenuation as a result of using the same set of numbers for both administrations. Again, care was taken to prevent numbers appearing contiguously, and/or in ascending or descending order. As previously mentioned, each participant was given one of the test versions in his/her home language, and the other version using numbers presented in English. The order in which the two sets of numbers were given was randomised, as was the order in which the language of the tests was conducted. In other words, sometimes the standardised set was presented first, with the on-the-spot generated set presented second, and vice-versa; sometimes the assessment was conducted using English numbers first, and numbers in the participant's home language second, and vice versa. Only the Forward Digit Span sequences were administered, and one trial was given at each sequence length for two through to seven digits. Therefore, the aim was to achieve a cross-randomisation of digit sets and language.

Adaptation of the 4 Hidden Objects Test

For the adaption process of this test, two sets of objects were used. Following the same approach as with the Digit Span, one standardised set of four objects was developed to compare to one randomly selected, changing set of objects, which would be chosen from any items present in the room at the time of each testing session — for example, a cellular

telephone, a bank note (money), a pen, a watch, a pad of paper, a cuff link or glasses, to name a few of the objects used. The use of ready-at-hand objects is the same as the way that patients are assessed at the bedside during clinical neuropsychological assessments with the 4 Hidden Objects Test.

The standard set of four items was created after much consultation with the isiXhosa cultural experts, as well as with the clinical neuropsychologists, drawing on their extensive clinical and assessment experience as to which objects had previously been problematic and why. The following four items were eventually settled on: a pipe, a bangle, a flower and a key. While it is obvious that no items selected can be completely ‘culture free’, it was agreed that the above-mentioned set constituted a more culturally fair set of items, as they were chosen as non-westernised items, equally likely to be found in an urban or rural setting.

Adaptation of the Babcock Story: Township Fire Story

Given the cultural bias of the original Babcock Story, the first priority for the adaptation was to change the theme and setting to an appropriate South Africa one. To this end, a brainstorming session was held to try to come up with possible scenarios to portray. As was the case with the creation of the Mary Selo Story, it was important that the story be set in a well-known South African location, describing activities familiar to all people. The more obscure and unfamiliar the content of the story, the more of a disadvantage it would be to the patient in trying to recall the information.

The initial attempts to meet the above-mentioned criteria included stories about a sporting event, a crime being committed, a trip to the shops, a motor-vehicle accident, and a holiday to a nature reserve/game park. The challenge was to provide sufficient detail, or the correct

balance of detail, to produce a meaningful story, while at the same time including content that would be familiar to most South Africans. Consequently, it was decided that it was best to stick to the ‘disaster theme’ of the original Babcock Story — this made sense, especially as the original 21-unit semantic scoring structure of the story was also retained and all the other tests adapted in this study had adhered to the original conceptualisations of the tests from which they were adapted.

After some careful thought about what possible disaster to portray in the new story, the theme of a fire in a township was chosen because it was considered to be an event that would be in some way familiar to most South Africans, as it is a regular occurrence in both rural and urban settings. Additionally, there was media coverage of a big township fire around this time, adding credence to the fact that most people would be able to relate to the new story’s content. This concept was approved by the cultural experts, who readily agreed that it would be suitable for the local peoples. A final consideration was that a ‘township fire’ disaster theme would make it easier to compare the new versus the original version of the tests, as both portrayed a disaster taking place.

The new test (see Appendix B) was then shown to the experts on the neuropsychology panel, who advised on how to best to divide the story into the most appropriate 21 semantic scoring units. These had been difficult to allocate, especially since in the original story the allocation of the units appears to have been largely arbitrary. Once all the experts were satisfied with the changes made to the test, it was officially named the Township Fire Story. Again, a mini pilot was first conducted on the test; this revealed no initial difficulties with the participants’ ability to internalise and recall the content.

Language Tests

Adaptation of the Boston Naming Test: Naming Test

It was decided to reduce the Boston Naming Test, which contains 60 items, to 30 items — a far more suitable length for use within the quick and easy-to-administer GSNSB. Williams, Mack and Henderson (1989) were in fact the first to develop this 30-item shortened form of the Boston Naming Test. This shortened version, which has proved to be empirically sound, utilises either the 30 even number items from the original Boston Naming Test, or the 30 odd numbered items (Lansing, Ivnik, Cullum & Randolph, 1999). This study chose to shorten the Boston Naming Test by retaining all the odd-numbered pictures. After a number of years of clinical experience, with many clinical neurocognitive assessments having been completed in the South African context, the Neuropsychology division were able to identify the most problematic items in the original Boston Naming Test that they had encountered in their clinical work. Among the items initially identified were the *pretzel* and *beaver* (see Figures 4.1 and 4.2), along with many other items including the *Sphinx*, the *trellis*, the *asparagus*, the *unicorn*, the *hammock* and the *pelican*, to name but a few. All of these items consistently proved difficult for the patient population at the hospital due to their highly westernised orientation. Such problems were consistent with findings in the literature, where tests had been modified for other non-American contexts. A study conducted in New Zealand, for example, found that Maori people had struggled to name the *pretzel* and *beaver* (60 percent more errors than American samples) as well as the *globe* and *asparagus* (20 percent more errors) (Barker-Collo, 2001).

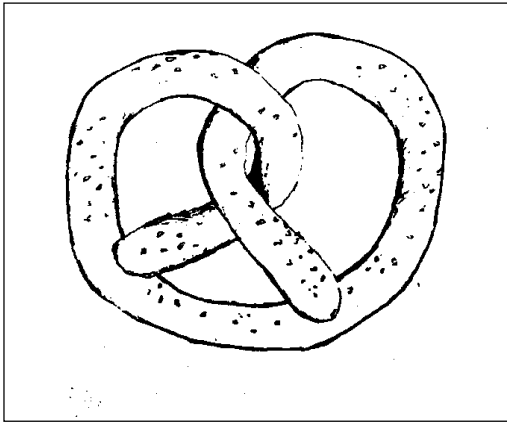


Figure 4.1. Pretzel – problematic item from BNT

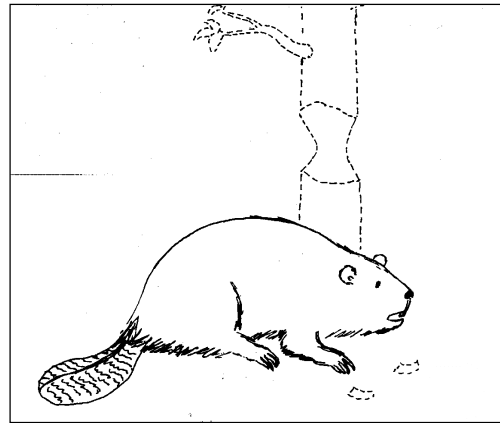


Figure 4.2. Beaver - problematic item from BNT

In total, 21 of the 30 Boston Naming Test items were identified as problematic. This determination was based not only on past clinical experience but also on the views of the panel of cultural experts at the Southern African Languages Department at the University of Cape Town. In addition, after further consultation with the panel of neuropsychologists, it was decided to replace all 30 original items as a result of copyright. The new replacement items were designed and constructed using a process that involved ongoing collaboration with the panel of cultural and language experts, as well as taking into consideration the opinions of members of the general public regarding what they deemed appropriate/inappropriate and why.

The primary aim was to develop items similar to the original Boston Naming Test, but specific to the general South African context (its culture and environments). The plan of first adopting general consensus as to the effectiveness of the test changes was the initial step in the adaptation process. Items from the original Boston Naming Test were replaced with new ‘South African’ items of equivalent complexity — for example, the *pelican* was replaced by an *ostrich*. Here, the pelican is a *type* of bird, so the goal was to replace it with an equivalent item, that is, a type of bird that would be familiar to South Africans. In this instance, an

ostrich was deemed to be a worthy equivalent — the items changed in this manner are noted in Table 4.1.

Table 4.1

List of adapted items

Old Item	New Item
Latch	Padlock
Pelican	Ostrich
Hammock	Goal net
Tennis racket	Cricket bat
Dominoes	Dice
Igloo	Thatched hut
Asparagus	Carrot
Scroll	Newspaper
Rhinoceros	Hippo
Beaver	Snake
Snail	Locust
Dart	Spear

The task of adapting the nine remaining unproblematic items proved fairly straightforward; again, equivalent items were used to replace the originals — for example, the *pencil* was replaced with a *pen*, the *saw* was replaced with an *axe*, the *bed* was replaced with a *chair*, etc.

The above-mentioned approach to adapting the items could not be used for a number of the original Boston Naming Test pictures, as South African equivalents could not be found —

such items included the *pyramid*, *Sphinx*, *escalator*, *globe*, *unicorn* and the *volcano*. When the panel of cultural experts were consulted, this problem was confirmed, due to the fact that trying to identify/create items that are fair to more than one ‘culture’ in a single context is extremely challenging. For example, there was no possible ‘supernatural creature’ to replace the *unicorn* that could be both drawn as a picture and at the same time be fair to isiXhosa, as well as Afrikaans and English people. After further research and deliberation over this problem, it was decided that the best solution was to replace the *unicorn* with an item of similar difficulty, but not a supernatural creature. The underlying principle was that the item should be an animal similar to a horse but not as common — consequently, a *zebra* was chosen, as it met this criterion.

The same approach was used for the other items identified as not having South African equivalents — the *volcano* became an *open fire*, while the pyramid was changed to *isivivane*, an indigenous word for a pile of stones made by travellers at the side of the road. It was, however, decided to replace this item, as it was realised that only isiXhosa people would be familiar with the term — there were no Afrikaans or English equivalents, rendering the item biased even within the South African context. The suggestion that a picture of a *pestle and mortar* be used was turned down because the cultural experts pointed out that this item was depicted differently in rural isiXhosa and urban English settings and was therefore problematic.

The above-mentioned description eloquently demonstrates the real-world challenges that the test-adaptation process involved and, at the same time, shows the clear advantage of a ‘converging lines of evidence’/inter-disciplinary approach to the problem. Another challenge for this aspect of the research was that the patient population of Groote Schuur Hospital does

not have the same socio-economic status or access to education as populations in western contexts. This relative deprivation required that test items in some instances be made 'simpler' than those used in other contexts. Having the cultural experts to provide insights into the familiarity of items to Afrikaans and isiXhosa peoples, as well as people from rural and urban backgrounds, proved invaluable. As part of the panel assembled for this study, the neuropsychologists also provided invaluable input. Their contribution in judging the adequacy of the newly developed test items and whether they were clinically useful (for example, whether they would be able to discern semantic paraphasia and other deficits), was of key importance. They also helped by assisting in the ordering of the Naming Test items, to ensure that a hierarchy of difficulty similar to the original Boston Naming Test was maintained.

Before the main pilot study began, all the newly created Naming Test items (see Appendix B) were tested through a 'mini pilot' study. The items were shown to a sample comprising a few people with similar demographic characteristics to the participants in the main pilot study, with none of the items being demonstrated as overtly problematic. As a result, the 30 new items were finalised, and then drawn in black ink in the style of the original Boston Naming Test items.

Adaptation of the Cookie Theft Test: Washing Line Picture Test

Through the Neuropsychology division's prior clinical experience with this test, it had already been demonstrated that, as with many of the other tests in use, both the patient's level of education, and whether he/she was from a rural or urban background, were both important factors in determining his/her test performance. It is of vital importance that the scene depicted in a test assessing narrative speech production (verbal fluency) and paraphasia is

familiar and fair to all patients. Bearing these factors in mind, and after confirming the idea with the panel of experts, it was decided that a rural scene was needed, where typical everyday South African activities were taking place. Consequently, a washing line was chosen as the focal point of the newly depicted scene, with a woman hanging out the washing to dry on the line. Because this test requires the patient to tell a story about the picture, two children playing were then added to the scene to help the patient to connect different aspects of the scene — one child is shown splashing in the mud, which soils the washing on the line, while the other is shown playing with a sleeping dog and about to pull the dog's tail. The result is a scene where all the characters are active in some way, which facilitates the telling of a story. This allows the examiner to assess the patient's verbal fluency, use of verbs and function words, and to look for paraphasias — all key aspects of assessing speech production in aphasia.

Once the Washing Line Test had been created, the cultural experts were consulted in order to determine whether the scene as a whole was fair and familiar to all three cultural groups. After reviewing various versions created by an artist, it was agreed that the picture (see Appendix B) was culturally fair and suitable for piloting in the South African context. The neuropsychology experts were then also asked to comment on the scene, primarily to help evaluate whether the Washing Line Test was of equivalent complexity to the original Cookie Theft Test and to ensure that other possible clinical considerations had been taken into account. Following from this process, the neuropsychology panel suggested that the picture should include additional background elements to elicit more commentary from the patient. As a result of this suggestion, the artist added chickens feeding on the ground and a bird of prey flying in the sky. As was the case with the new Naming Test, a 'mini pilot' study was

then conducted with the same participants mentioned above. This revealed that none of the participants had problems creating a story from the depicted scene.

Adaptation of the Anna Thompson Reading Test: Mary Selo Story

Although designed as a verbal memory test in the Wechsler Memory Scale (WMS), the Anna Thompson Story was included in the GSNSB primarily as a reading test. Given its European origin and contents, an alternative was needed to meet the requirement of being suitable for use in the South African context. Again, the need for this adaptation had already been established by the clinical experiences of the neuropsychologists using this test in the hospital setting. Consequently, the Mary Selo Story was created, set in a South African location (the city of Port Elizabeth), with culturally neutral activities being portrayed (for example, going to the beach and running to a bus stop for shelter from the rain).

The key consideration in designing the Mary Selo Story was to follow the same structure of the verbal memory tests (even though this was not its intended use in the GSNSB) so that it might be potentially used in this way in the future. Therefore, it was ensured that the new story retained the same scoring structure as the original — 21 units, or separate pieces of information. This task proved problematic (see the ‘*Translation of Adapted Tests*’ section below).

Deciding the name of the central character in the story was also important as a specifically South African surname was required. This also applied to the first name. To investigate the potential familiarity of the first name and, after discussions with the cultural experts, participants were consulted as part of the mini pilot. Through this, it was discovered that the surname *Selo* and the first name *Mary* were both suitable and there was general consensus on

their familiarity. In addition, all of the ‘mini pilot’ participants were able to read the passage, commenting that it was more than adequate for the South African context. Finally, the neuropsychology panel confirmed that the new story (see Appendix B) was clinically acceptable, and that the 21 units of information it contained had been appropriately chosen and subdivided (the breakdown of verbal memory stories into 21 units appears, upon examination, to have been largely arbitrary).

Spatial Cognition Tests

Adaptation of the Scene Drawing Test: Hut drawing Test

Discussion with the cultural experts regarding a replacement test for the Scene Drawing Test began with an appraisal of the weaknesses of the original. Once again, starting from the neuropsychologists’ clinical experiences, it was agreed that the test did not represent an identifiably South African scene. For example, the experts highlighted the fact that typical South African dwellings, especially township houses and shacks, do not have chimneys, and have roofs that are substantially different from the one depicted in the original test. This observation also applied to the fence in the original picture, which is not the kind of fence that you would see in a township or urban setting. It was further noted that the tree is also not typical in appearance to any South African species, being far more European in appearance. All in all, it was decided that the scene was far too westernised to be left unchanged.

Initial attempts at changing the scene centred on the drawing of a local street scene, including cars and buses. After numerous attempts, however, it was decided to reject this concept as it was too difficult trying to create the scene while at the same time avoiding making it look too much like an urban/city depiction to the detriment of rural people. In addition to this, it was difficult to create a scene of this sort that would be sufficiently detailed, yet at the same time

easy enough for a patient to copy; bearing in mind that the purpose of the test is to elicit potential visual neglect, a scene with similarly equal left-right proportions is therefore very important. The initial objects chosen — the car, street setting and bus — were too difficult to satisfy these key objectives.

Consequently, an alternative conceptualisation was sought. It was then decided to retain the original idea of a house, but to adapt it so that the new picture would resemble a typical South African township or rural dwelling. Therefore, a rendition was required that depicted trees of an African variety and a dwelling without a chimney, with a traditional roof (made of thatch), and an alternative, a more simple fence (made of sticks). All the while, it was important to bear in mind the intended clinical purpose of the test. During the drawing process, it was discovered that trying to draw two different-looking trees, which still looked indigenous to Africa, was problematic. After further consultation with the cultural experts, it was decided to replace a possible tree on the left of the picture with a traditional cooking pot (known locally as a *potjie*).

The new scene (see Appendix B), complete with the newly added traditional pot, was accepted by the cultural experts. When consulting the neuropsychology panel, they advised that more detail needed to be added to the picture to afford more opportunity to detect possible visual neglect. To this end, a door latch (crucially on the left of the hut's door) and a cloud in the sky, also to the left of the picture, were added. These details improved the overall symmetry of the picture, while at the same time providing more details for the patient to potentially miss 'on the left'. This example again illustrates the importance of combining cultural considerations with clinical understandings of neurocognitive disorders when

designing new tests. The ‘mini pilot’ study confirmed that the new scene was good in this initial form.

Adaptation of the Cube Analysis Test: 3-D Analysis Test

The adaption of the original Cube Analysis Test proved far from easy. The first challenge was to creatively draw different blocks, which were sufficiently different from the originals. On the first attempt, the blocks had a clear perspective problem, due mainly to the lack of uniformity and texture of the lines, but also due to distortion of the 3-D perspective. This problem resulted from the fact that sufficient accuracy could not be achieved using free-hand drawing. Therefore, a second attempt was made to draw the cubes accurately using a computer drawing program. This, however, still did not rectify the problem as the 3-D perspective of the blocks was distorted and the lines irregular. A third attempt, adjusting the accuracy, produced a better perspective but the cubes still required rearranging so that the order of complexity of the 14 items in the test was graded. When, after this third attempt, the 3-D perspective was still unsatisfactory, a professional architect was consulted to solve the perspective problems being encountered, which resulted in greater accuracy.

Executive Tests

Adaptation of the Controlled Oral Word Association Test (FAS): BHP and NPS Tests

The adaptation of the COWAT/FAS Test began with the consultation of both Afrikaans and isiXhosa dictionaries in order to ascertain word frequencies. This approach, recommended by Nell (2000), was vitally important in ensuring that the adapted tests were equivalent to the original. Because this test is designed to test a patient’s ability to generate words, it is essential that there are sufficient words in the relevant language beginning with each letter

given to the patient, so as not to disadvantage him/her. This is vital in order to accurately elicit neurocognitive deficit.

The adaptation of the Controlled Oral Word Association Test was to prove a difficult and complex task, especially given that the study involved two new and diverse languages: Afrikaans and isiXhosa. The basic principle followed was that the first letter chosen should occur the most frequently of the three, with the second less frequently, and the third the least frequent. Following this principle, substitutes for the original letters F, A and S were sought using isiXhosa and Afrikaans dictionaries. For the Afrikaans version, the Pharos/NB corpus database was used. This database, the publication of a number of South African publishers, is the most comprehensive lexical database of the Afrikaans language, consisting of 30 million words in fiction and non-fiction. After many consultations, the letters *B*, *H* and *P* were finally selected from the 402 351 unique word tokens. These were then reviewed and approved by the cultural experts.

When applying the above procedure to isiXhosa word frequencies, problems arose. It soon became apparent that there was no readily available lexical data for isiXhosa. Despite searching libraries and consulting online sources, sufficient data could not be found. At this point, the advice of the cultural experts was sought; they advised that *The Greater Xhosa Dictionary*, which is published by the University of Fort Hare, be consulted. The problem remained that this dictionary had only ever been published in the third volume, while most of the other dictionaries are either Xhosa-Afrikaans or Xhosa-English dictionaries. Finally, an isiXhosa dictionary published by Pharos was used (Pharos, 1998). The selection process was further complicated by the fact that, due to the abundance of prefixes (for example, *um*, *aba*, *isi*, *ulu*) used in isiXhosa, certain letters have to be immediately excluded as prefixes could

just be added to any word, thereby nullifying the purpose of the test. Once the word-spreads had been calculated, the letters *N*, *P* and *S* were selected and were subsequently approved by the isiXhosa experts. Given that no clinical or cultural problems had been found when using this test with English patients and that there were no copyright infringements, the original letters of the COWAT — *F*, *A* and *S* — were retained for the English version in the GSNSB.

Administration of the tests

The procedure used in administering the tests to the research participants was as follows. A designated room was found at Groote Schuur Hospital so that all assessments could take place in one setting — all participants were seen in this room. On the day of his/her assessment, each participant was shown to this room and briefed about the anonymity and confidentiality of his/her participation; the fact that he/she could withdraw from the study at any point; and the intended purpose of study. This was done with the aid of the Patient Information Sheet. The participants were then asked to sign the Consent Form.

In the case of Afrikaans- or isiXhosa-speaking participants, the entire testing session, including the briefing, was conducted in their first language. This was done using the translated version of the GSNSB Prototype (see Chapter Three). The Afrikaans and isiXhosa interpreters used had been members of the panel of cultural experts assigned to assist with the test-adaptation process. This was highly advantageous, as it ensured that they had the requisite experience with working with the test material. In addition to this, the interpreters were also formally trained by the author to administer the GSNSB. The interpreters were also chosen due to their experience in working in the Groote Schuur Hospital context, assisting and teaching medical students, and helping medical staff with bedside interpreting. This was

of great benefit to the study, as they were all already familiar with dealing and conversing with hospital patients.

The order of administering the new and the original tests was randomised so as to avoid practice and ordering effects. Each testing session lasted approximately one-and-a-half hours, given that both the old and new versions had to be administered and all qualitative feedback required recording. Where possible, the relevant test responses were tape recorded, to ensure accurate scoring following the completion of the testing session. Given financial and time constraints, and the difficulty of getting participants to return to the hospital a second time, it was not possible to conduct the testing over two sessions. All the participants' test responses and their qualitative observations were recorded on the specifically designed Scoring Sheet.

As previously mentioned, all assessments were conducted in the participants' first language. However, for the FAS Test, the Afrikaans participants were required to perform both the letters *F*, *A* and *S* in English, and the new letters *B*, *H* and *P* in Afrikaans. The isiXhosa participants were also required to perform the letters *F*, *A* and *S* in English, while the letters *N*, *P* and *S* were assessed in isiXhosa. Once the testing was complete, the participants were asked which version of the test they preferred and which of the two they had found more familiar. A normal performance was deemed to be a score of 35 or more over the three letters combined.

The second set of tests that was an exception to the rule of assessing the participant exclusively in his/her first language was the Digit Span/Auditory Span Test, which was also administered in more than one language. As with the other tests earmarked for adaptation, two versions of this test were also administered. The first of these was a series of random

numbers thought up on the spot, varying from participant to participant. The second was the standardised set of number sequences, generated for the pilot study. The random, varying set of numbers was included to prevent test-retest attenuation as a result of using the same set of numbers in administering both the 'old' and 'new' tests. As with the other tests, the order in which these two versions were administered was randomised, along with the language used. In other words, on some occasions the randomised version was assessed in English and the new version in either Afrikaans or isiXhosa, depending on the first language of the participant; on other occasions, the new version was administered in the English and the randomised version in the participant's first language. For the English participants, both versions were administered in English.

For the Babcock and Township Fire stories, it was considered as self-evident that the first-language Afrikaans and isiXhosa participants would fair relatively poorly in a memory task that was administered in English. This assumption was backed by the neuropsychologists' clinical experiences. Therefore, it seemed only logical to administer these tests in the first language of each participant.

The final stage of the procedure followed during the pilot study was a follow-up meeting with the interpreters, which occurred once the data collection had been completed. This meeting was held to check that no scoring errors had been made and to confirm that all data had been collected. In order to eliminate possible misunderstandings that could have resulted from the interpretation process, the interpreters were asked to comment on any problems they had noted with aspects of the testing process or on any identified recurring mistakes on the part of the participants. The interpreters were then asked to share their overall experiences of the pilot study and to give feedback on any possible weaknesses they might have observed.

Results

The 4 Hidden Objects Test

The total number of participants who were administered this test was 29 ($N = 29$) not 30, as one of the participants reported being too fatigued to complete this particular test. A Friedman ANOVA was performed on the data, the outcome of which was that there was no significant difference (the chi-squared Friedman ANOVA value was exactly zero) on the participants' performance between the two differing sets of objects, $\chi^2_{\text{F}}(1, N = 29) = 0.000$, $p = 1.000$. Descriptive statistics for the 4 Hidden Objects Test are provided in Table 4.2.

Table 4.2

Descriptive statistics for the 4 Hidden Objects Test

Group	Mean		Standard Dev	
	Old Set	New Set	Old Set	New Set
isiXhosa	16.00	16.00	0.00	0.00
Afrikaans ^a	15.90	15.90	0.32	0.32
English	16.00	16.00	0.00	0.00

Note: $N = 10$ for each group, unless otherwise specified.

Maximum score = 16.

^a $N = 9$.

In other words, these results revealed that the participants' performance on the 4 Hidden Objects Test did not differ between the original (culturally biased) objects and the newly chosen (culturally fair) ones.

Digit Span/Auditory Span Test

The Friedman ANOVA procedure produced a non-significant finding for the comparison between these two tests, $\chi^2_{\text{F}}(1, N = 20) = 1.667, p = 0.197$. This demonstrated that the test performance was not influenced by the language the test was administered in. The Kendall's co-efficient of concordance was 0.83, suggesting a high degree of confidence in this finding. A *t*-test performed on the data also revealed no significant between-group differences; that is, when the Digit Span was administered in English numbers, the isiXhosa group ($M = 5.80, SD = 0.63$) did not score significantly higher than the Afrikaans group ($M = 5.70, SD = 1.34$), $t(18) = 0.21, p = 0.830$. A one-way ANOVA performed on the data also showed no significant difference between groups, $F(2, 27) = 1.32, p = 0.280$. This showed that when either test version was administered in the participants' first language, no significant differences existed between the isiXhosa, Afrikaans and English individuals (see Table 4.3 for descriptive statistics).

Table 4.3

Descriptive statistics for Auditory Span Test

Group	Mean		Standard Dev	
	ENG	HL	ENG	HL
isiXhosa	5.80	6.10	0.63	1.10
Afrikaans	5.70	6.10	1.34	0.99
English		6.80		1.23

Note: $N = 10$ for each group. ENG = English, HL = Home Language.

A final question to answer was whether the participants' level of education influenced their performance on the Digit Span/Auditory Span Test. In order to complete a logistic regression analysis, three participants were re-included in the sample ($N = 32$) — they had been previously excluded from the other statistical procedures because they were not alike on the ethnic group–home language match, that is, they were not from the same ethnic group as the majority of the participants representing their language group (for example, an English speaker who was not white). Of these re-included participants, one was withdrawn from this particular analysis as a result of having only one year of education, which was deemed to be an outlier that would skew the results.

Table 4.4 gives the descriptive statistics for the participants' levels of education. Of the 32 participants, 14 were coded as FAIL and 18 as PASS. The Quasi-Newton estimation method used for this analysis revealed that level of education was a significant predictor of success or failure, Wald's $\chi^2(1, N = 32) = 5.693, p = 0.017$. This demonstrated that a higher level of

education was significantly related to a good performance on the Digit Span/Auditory Span Test.

Table 4.4

Descriptive statistics for Education

Group	Mean	Standard Dev	Minimum	Maximum
isiXhosa ^a	9.70	2.31	5.00	12.00
Afrikaans ^b	9.00	2.00	6.00	12.00
English ^c	13.39	2.26	10.00	16.00
All Groups	11.00	2.94	5.00	16.00

Note: Four of the participants in the English group were non-white.

^a*N* = 10. ^b*N* = 9. ^c*N* = 13.

The classification of cases for the Digit Span/Auditory Span tests revealed an odds ratio of 4.667, with 68.75 percent of the cases having been correctly classified. The odds ratio of greater than one showed that the classification was better than would be expected with pure chance. Additionally, the results displayed in Table 4.5 show how the participant's level of education predicted success on the test roughly 20 percent better than it predicted a failure.

Table 4.5

Classification of cases

Observed	Predicted		Percent Correct
	Failure	Success	
Failure	8	6	57.14
Success	4	14	77.78

Babcock Story/Township Fire Story

The qualitative results for these two tests revealed how many of the participants found specific elements of the original test problematic (despite all participants passing, according to the requirements of the tests as outlined in the GSNSB Prototype). For example, the town name *Albany* was either left out of their recalls altogether or substituted with *Albertinia* or *Alberton*, which are places in South Africa. A similar example was the word *mile*, which was again either left out of the recall or recalled as *kilometres*. A similar result from the new test was that many participants changed the word *house* to *shack* when recalling the story. Other qualitative results were the presence of contamination between the stories when the second story was being recalled, especially when dates or numbers of people killed were mentioned — this was due to the fact both stories were administered during the same testing session. Also of note was the fact that qualitatively the majority of the participants reported finding the content of the new test more familiar when asked for their feedback.

When examining the descriptive statistics, the results showed that both the English ($M = 35.5$; $SD = 9.1$) and Afrikaans ($M = 29.9$; $SD = 9.3$) participants fared better on the new test.

On the other hand, the results showed that the isiXhosa ($M = 28.8$; $SD = 9.6$) participants did better on the original test (see Table 4.6). Statistical analyses of the differences in the groups' test performances between the old and new versions of the tests were done using t-tests. The results found no significant differences between the tests for any of the three participant language groups ($p < 0.05$). The raw test scores for the three groups' performances on the Babcock Story and the Township Fire Story are displayed in Table 4.7.

Table 4.6

Descriptive statistics for the Babcock and Township Fire stories

	Mean	Minimum	Maximum	Std Dev
Babcock Recall Story for English Group	31.3	17	42	8.4
Township Fire Recall Story For English Group	35.5	23	53	9.1
Babcock Recall Story for isiXhosa	31.6	12	52	12
Township Fire Recall Story for isiXhosa	28.8	14	43	9.6
Babcock Recall Story for Afrikaans Group	27.1	11	49	12.45
Township Fire Recall Story for Afrikaans Group	29.2	17	43	9.31

Table 4.7

Raw scores for the Babcock and Township Fire stories for all three groups

ENGLISH											
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Total
Babcock											
1st Recall	2	10	11	2	9	7	10	5	8	12	76
2 nd Recall	8	14	13	9	12	9	15	15	11	15	121
delayed	7	14	12	9	10	10	16	12	11	15	116
	17	38	36	20	31	26	41	32	30	42	313
Township											
fire 1st recall	4	15	7	11	7	6	7	11	11	12	91
2nd recall	10	18	13	13	14	9	10	15	16	16	134
delayed	9	20	11	12	12	15	8	13	14	16	130
	23	53	31	36	33	30	25	39	41	44	355
isiXhosa											
	Case 11	Case 12	Case 13	Case 14	Case 15	Case 16	Case 17	Case 18	Case 19	Case 20	Total
Babcock 1st recall	5	6	8	11	3	8	12	16	11	4	84
2nd recall	7	11	15	12	10	13	14	17	14	8	121
delayed		7	13	15	8	13	12	19	16	8	111
	12	24	36	38	21	34	38	52	41	20	316
Township											
fire 1st recall	7	4	6	9	9	6	4	12	10	9	76
2nd recall	10	6	13	13	9	10	11	17	15	10	114
delayed		4	7	15	9	10	11	14	17	11	98
	17	14	26	37	27	26	26	43	42	30	288
AFRIKAANS											
	Case 21	Case 22	Case 23	Case 24	Case 25	Case 26	Case 27	Case 28	Case 29	Case 30	Total
Babcock 1st recall	12	5	4	10	13	3	6	6	11	5	75
2nd recall	13	3	5	9	18	8	10	10	17	7	100
delayed	11	3	4	10	18	8	9	10	15	8	96
	36	11	13	29	49	19	25	26	43	20	271
Township											
fire 1st recall	6	8	8	8	11	6	9	7	11	7	81
2nd recall	7	5	13	10	16	7	12	10	16	12	108
delayed	7	4	13	11	15	5	14	10	16	8	103
	20	17	34	29	42	18	35	27	43	27	292

Boston Naming Test/Naming Test

The results showed that the majority of the participants, 25 out of the 30, reported preferring the new Naming Test to the Boston Naming Test, saying that it was more familiar to them. For the original Boston Naming Test, the participants' average score was 17.3 out of 30, their average score being 27.27 out of 30 for the Naming Test. Of the three respective language groups, the isiXhosa-speakers scored the lowest on both versions — an average of 13.1 for the original version and 26.8 for the adapted test. The average scores for the English and Afrikaans groups were similar, the Afrikaans average being 18.4 on the old version and 27.4 on the new one, while the English average was 20.4 on the original version and 27.9 for the new test (see Table 4.8).

Table 4.8

Controls' performances on Boston Naming Test/Naming Test

<i>N</i> = 30	isiXhosa	English	Afrikaans	Total
Boston Naming Test				
<i>M</i>	13.1	20.4	18.4	17.3
<i>SD</i>	5.47	6.98	3.84	6.23
Naming Test				
<i>M</i>	26.8	27.9	27.4	27.37
<i>SD</i>	1.99	1.91	1.65	1.85

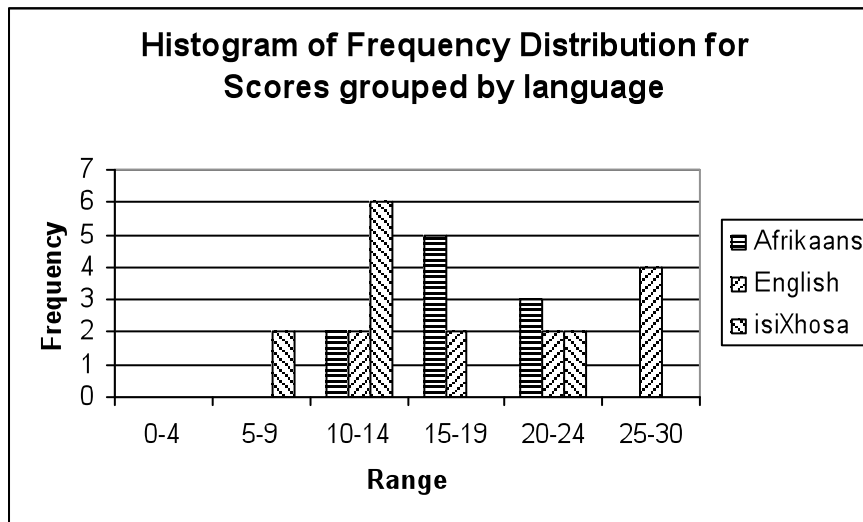


Figure 4.3. Language group scores on the Boston Naming Test

Two interesting findings emerged regarding the three language groups' scores on the Boston Naming Test (see Figure 4.3). The first was that the isiXhosa participants scored worse on the Boston Naming Test than both the Afrikaans and English participants. The second finding here was that there was a large disparity within the isiXhosa group, as they tended to score either very well or very poorly on the original test.

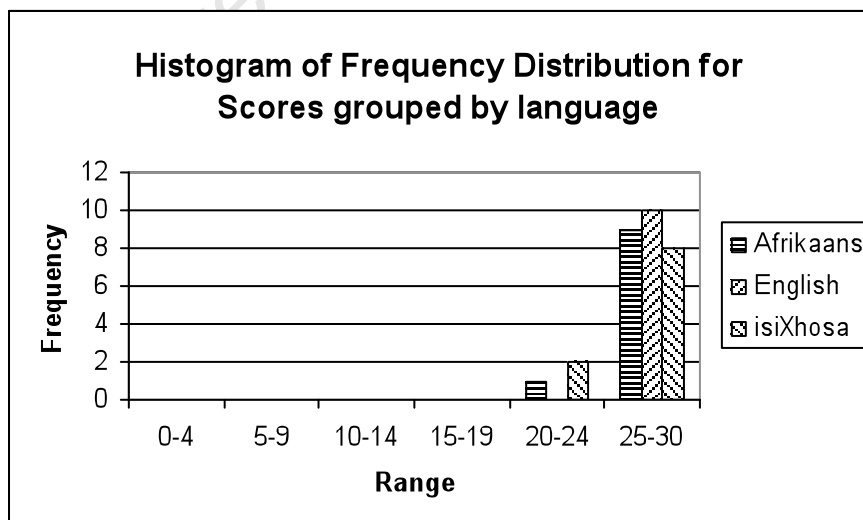


Figure 4.4. Language groups scores on the Naming Test

The findings regarding the three language groups' performance in the new Naming Test were convincing (see Figure 4.4). All three language groups' performances improved dramatically, demonstrating that cultural bias had significantly diminished in the new test.

To further corroborate the findings of the descriptive statistics, statistical t-test analyses were performed to examine the differences between the participants' performances on the original versus the new Naming Test. The results revealed a highly significant performance difference ($p < 0.05$) between the old and new tests — the Afrikaans and isiXhosa groups showed a significance level less than 0.001, while the English group showed a significance level of 0.004 (see Table 4.9).

Table 4.9

t-tests for Boston Naming Test/Naming Test

($p < 0.05$)

Group Tested	T-stat
t-test for all groups	0.000001
Afrikaans	0.000010
isiXhosa	0.000005
English	0.003816

Further findings identified a large number of the Boston Naming Test items as problematic in the South African context. A number of the 30 items examined frequently proved problematic; of these, the *trellis* (25 incorrect answers), *asparagus* (23 incorrect answers), *pretzel* (23 incorrect answers), *pyramid* (22 incorrect answers), *hammock* (22 incorrect answers), *Sphinx* (22 incorrect answers), *unicorn* (20 incorrect answers), *pelican* (18 incorrect

answers), and *beaver* (18 incorrect answers), were recorded as the most culturally inappropriate.

In addition to the number of errors made on specific items, a number of qualitative observations emerged regarding the ‘themes’ of the errors made on the original test. These observations further highlighted the urgent need to change the items in the light of cultural bias and educational deprivation. Twenty-three of the participants incorrectly named the *pretzel*; of these, nine named it a *snake*, and four a *worm*. Eighteen of the participants incorrectly named the *beaver*, six naming it a *rat*, while four thought it was a *mouse*, and one a *hamster*. In addition, all the participants who misnamed the *pelican* called it various incorrect types of *bird*. For the *hammock*, they frequently named it a *net*, while the *unicorn* was frequently named a *horse*. These results highlighted the large degree of similarity between the answers of the participants who misnamed these particular items.

When examining the new Naming Test (see Figure 4.7 below), the qualitative results revealed that only a handful of items were problematic. From the new test, the *hippopotamus* (see Figure 4.5), *mug* (see Figure 4.6) and *dragonfly* (see Appendix B) items were frequently misnamed — nine participants misnamed the *mug*, with eight of the nine (six being isiXhosa speakers) calling it a *cup*. More Afrikaans than English participants named the *mug* correctly. The *hippopotamus* was misnamed by 11 participants, five of whom called it a *pig*, four a *rhinoceros*, while two did not know its name. Interestingly, only one English participant named the *hippopotamus* incorrectly; all those who named it a *pig* were isiXhosa speakers and all but one of those who named it a *rhinoceros* were Afrikaans.

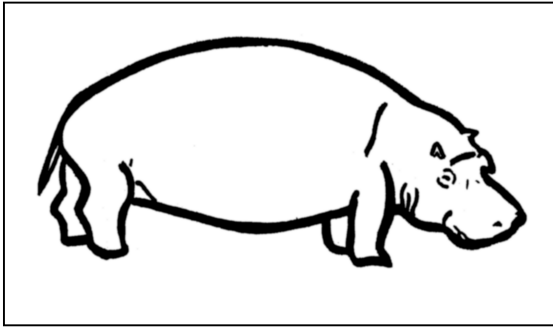


Figure 4.5: Hippopotamus

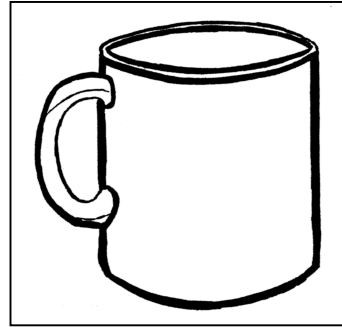


Figure 4.6: Mug

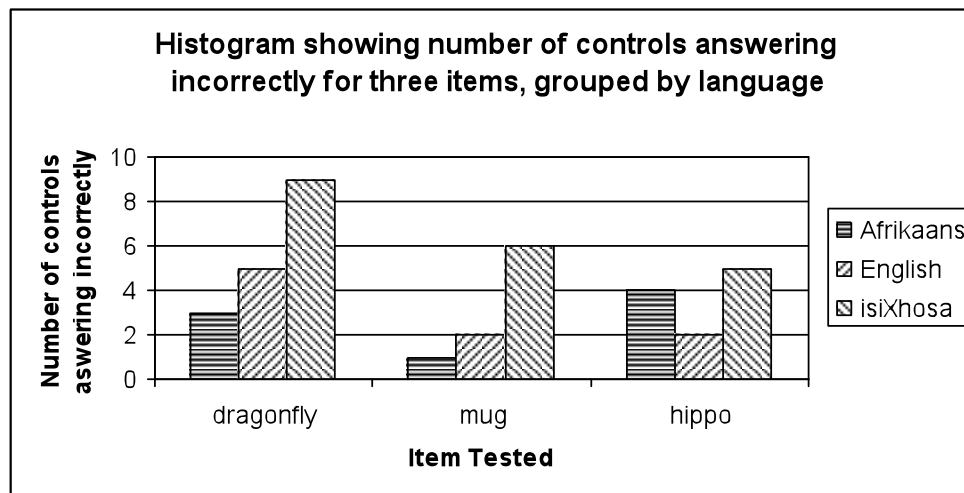


Figure 4.7. Problematic Naming Test items

Cookie Theft Test/Washing Line Picture Test

The qualitative observations revealed that the majority of the participants, 18 of the 30, preferred the new Washing Line Picture Test over the Cookie Theft Test, finding it more familiar (see Table 4.10). Six of the participants preferred the original and six had no preference. Broken down further, only one isiXhosa participant preferred the original test, and one had no preference. Of the English participants, five preferred the new test, three had no preference and two said the original was more familiar to them. Of the Afrikaans participants, five preferred the new test, saying it was more familiar to them, two had no preference, and three thought that the original test was more familiar.

Table 4.10

Controls' familiarity for the Cookie Theft and Washing Line Picture tests

<i>N</i> = 30	Afrikaans	English	isiXhosa	Total
Found Cookie Theft more familiar	3	2	1	6
Found Washing Line more familiar	5	5	8	18
No preference	2	3	1	6

For the quantitative analysis of the significance of the Cookie Theft/Washing Line Picture Test data, a chi-squared goodness-of-fit test was used — this analysis was not used on the individual language groups due to too small a sample size (see Table 4.11). Results showed that there was a highly significant difference between responses ($p = 0.008$), with the Washing Line story being found the more familiar of the two.

Table 4.11

Chi-squared test: significant difference between responses of familiarity - 0.008 significance level ($p < 0.05$)

Preference	No. of people (observed)	Expected Frequencies
Cookie Theft Test	6	10
Washing Line Picture Test	18	10
No preference	6	10

Anna Thompson Story/Mary Selo Story

As previously stated, these two test versions were used as a reading test in the GSNSB, not as memory tests. The results revealed that the participants all performed equally well on both the Anna Thompson and the Mary Selo stories, with no discernable differences observed between the two.

Aphasic Patients

Boston Naming Test/Naming Test

The results from the tentative investigation of these two tests using three aphasic patients were as follows. The first patient, a Wernicke's aphasic, misnamed the first five items of the new Naming Test, and was also unable to say whether he was able to recognise the pictures at all. The test was thus discontinued. With the original Boston Naming Test, the same problem occurred, resulting in the test again being discontinued.

The second patient, a Broca's aphasic, scored 11 out of 30 on the original, making many of the same errors that the control participants had made, for example naming the *pretzel* a *snake* and the *unicorn* a *horse*. When given the new Naming Test, this patient scored 22 out of 30. Among the items misnamed, she called the *locust (grasshopper)* a *bat*, the *newspaper* was named *letters*, and the *ostrich* a *chicken*. Interestingly, her answers were all semantically related to the actual items, which had been correctly named by the Afrikaans control participants who were from the same language/cultural group as she was.

Finally, the second Broca's aphasic perseverated on both the original and the new test, frequently using neologisms when attempting to name the items. With the new Naming Test,

she named the *spear* a *pole* and the *cricket bat* a *ball*, again revealing semantic errors on items that the control participants had found unproblematic.

Cookie Theft/Washing Line Story

Both versions of this test were able to elicit repetitions, problems with sentence structure and naming problems. Here follows an example from the Wernicke's aphasic Washing Line Picture Test response:

Die man ... hy hang die wasgoed. Hy hang die wasgoed ... hy hang die wasgoed. Die (indiscernible word). En daar is 'n man (pointing at the woman putting up the washing) en daar is 'n man, man, man (pointing at each of the children).

(Translation: The man, he hangs the washing. He hangs the washing ... he hangs the washing. The (indiscernible word) and there is a man (pointing at the woman putting up the washing) and there is a man, man, man (pointing at each of the children).

The first Broca's aphasic had problems naming objects in the pictures and with sentence structure. For example, while describing the Cookie Theft Test, he said:

Die 'laitie' staan op die dinges. Onder hom. Die water stroom ... daar. Die vrou vat 'oekie' note nie.

(Translation: The little guy stands on the thing. Under him. The water streams... there. The woman also does not take note.)

Finally, the second Broca's aphasic's performance when describing both tests revealed her problems with elocution, and her general slowness and difficulty with speech — for example, she had problems when it came to saying the words '*koekie*' (cookie), '*daar*' (there) and

'grond' (ground). The tests also revealed her incorrect use of prepositions and verbs — she told the story of the boy in the Cookie Theft Test falling *onto* the stool, while with the Washing Line Picture Test she said that the little boy was *running* in the water (rather than standing, splashing in the water as he is actually doing).

Anna Thompson Story/Mary Selo Story

When given these two tests to read, the first Broca's aphasic reported she could not read, and the Wernicke's aphasic refused to do these two tests. The second Broca's aphasic was fatigued and was only able to repeat the first sentence of the Mary Selo story, so it was discontinued.

Cube Analysis/3-D Analysis Test

A summary of the results from these two tests can be seen in Table 4.12 . In total, 15 of the 30 participants scored higher on the new test, with 13 performing equally well on both versions. Only two participants performed better on the original test. An additional qualitative observation was that, when the times taken to complete each were compared, 23 of the participants were able to complete the new test significantly quicker than the original version.

Of the 30 control participants, 16 (five isiXhosa, five Afrikaans and six English) preferred the original test, while the remaining 14 favoured the adapted version. Interestingly, some of the participants (two isiXhosa, two Afrikaans and three English) preferred the original test, yet scored better on the new version. The participants who preferred the original test described the new test as lacking in proper perspective, with 'too many hidden blocks', making them more difficult to count. Some of these participants admitted, however, that the blocks in the new test (being bigger) were easier to see. For example, an Afrikaans participant who

performed equally well on both tests, but who preferred the original, observed: “The new one is funny (*snaaks*), it’s something wrong with it”. However, an isiXhosa participant who did equally well on both tests preferred the new version, saying: “The new one is easier and clearer to see”. Predominantly, the participants who preferred the new version described it as clear to see and understand because the blocks were larger.

Table 4.12

*Cube Analysis/3-D Analysis Test performances
and participants’ preferences*

Controls’ scores	Old	New	preference
1 isiXhosa	2/2	0/2	Old
2 isiXhosa	0/2	0/2	Old
3 isiXhosa	0/2	0/2	Old
4 isiXhosa	0/2	1/2	Old
5 isiXhosa	0/2	1/2	Old
6 isiXhosa	0/2	1/2	New
7 isiXhosa	0/2	1/2	New
8 isiXhosa	0/2	1/2	New
9 isiXhosa	0/2	1/2	New
10 isiXhosa	2/2	2/2	New
11 Afrikaans	1/2	1/2	Old
12 Afrikaans	0/2	0/2	Old
13 Afrikaans	1/2	0/2	Old
14 Afrikaans	0/2	0/2	Old
15 Afrikaans	0/2	1/2	Old
16 Afrikaans	2/2	2/2	New
17 Afrikaans	0/2	1/2	New
18 Afrikaans	1/2	2/2	New

19	Afrikaans	0/2	1/2	New
20	Afrikaans	0/2	1/2	New
21	English	1/2	1/2	Old
22	English	2/2	2/2	Old
23	English	1/2	0/2	Old
24	English	1/2	1/2	Old
25	English	2/2	2/2	Old
26	English	0/2	1/2	Old
27	English	2/2	2/2	New
28	English	1/2	1/2	New
29	English	1/2	2/2	New
30	English	0/2	1/2	New

Patients' scores		Old	New	preference
1	isiXhosa	0/2	0/2	Old
2	isiXhosa	0/2	0/2	New

Scene Drawing/Hut Drawing Test

The results for these two tests showed that 28 participants scored equally well on both versions of the tests, the two others fared better on the new test (see Table 4.13). Only five (three isiXhosa, one Afrikaans and one English) of the 30 participants preferred the original, all of whom had fared equally well on both versions. The remaining 25 participants all preferred the new test, saying that it related better to them in terms of their daily lives, and that it appeared more familiar. One isiXhosa participant who did equally well on both versions, stated that: "It's more traditional and the tree reminded me of the Eastern Cape, you can rest under them".

Table 4.13

*Scene Drawing Test/Hut Drawing Test performances
and participants' preferences*

Controls scores	Old	New	preference
1 isiXhosa	1/1	1/1	Old
2 isiXhosa	1/1	1/1	Old
3 isiXhosa	1/1	1/1	Old
4 isiXhosa	1/1	1/1	New
5 isiXhosa	1/1	1/1	New
6 isiXhosa	1/1	1/1	New
7 isiXhosa	1/1	1/1	New
8 isiXhosa	1/1	1/1	New
9 isiXhosa	0/1	1/1	New
10 isiXhosa	0/1	1/1	New
11 Afrikaans	1/1	1/1	Old
12 Afrikaans	1/1	1/1	New
13 Afrikaans	1/1	1/1	New
14 Afrikaans	1/1	1/1	New
15 Afrikaans	1/1	1/1	New
16 Afrikaans	1/1	1/1	New
17 Afrikaans	1/1	1/1	New
18 Afrikaans	1/1	1/1	New
19 Afrikaans	1/1	1/1	New
20 Afrikaans	1/1	1/1	New
21 English	1/1	1/1	Old
22 English	1/1	1/1	New
23 English	1/1	1/1	New
24 English	1/1	1/1	New

25	English	1/1	1/1	New
26	English	1/1	1/1	New
27	English	1/1	1/1	New
28	English	1/1	1/1	New
29	English	1/1	1/1	New
30	English	1/1	1/1	New

	Patients' scores	Old	New	preference
1	isiXhosa	0/1	0/2	Old
2	isiXhosa	0/2	0/2	New

Right Middle Cerebral Artery (MCA) Patients

Cube Analysis/3-D Analysis Test

When given to the two right middle cerebral artery (MCA) stroke patients, both fared equally poorly on both the original and the new versions of the tests. When asked which they preferred, both stated they preferred the new test, as it was 'bigger and better'.

Scene Drawing/Hut Drawing Test

The two right middle cerebral artery stroke patients both fared better on the new Hut Drawing Test than on the original — although visual neglect was clearly evident in all four of their drawings (that is, old and new). When asked which test they preferred, both commented that the new version was more familiar to them, but that the original was slightly easier to draw.

Controlled Oral Word Association Test (FAS/BHP/NPS)

The qualitative results for these two tests showed that all the control participants did relatively well on both the original and the new test. Some of the participants began to repeat

certain words they had already given towards then end of the final trials, which was noted and judged to be due to fatigue. It was also observed that a few of the isiXhosa and Afrikaans participants switched to providing English words during the tasks, even though they were instructed that they could answer in their first language.

Descriptive statistics (see Table 4.14) revealed that the English participants fared the best of the language groups on the test. The isiXhosa ($M = 26.2$; $SD = 10.5$) and Afrikaans ($M = 23.3$; $SD = 9.3$) participants scored better overall on their new test (that is, either the NPS or BHP Test) than on the original FAS, but still not as well as the English participants.

Table 4.14

Descriptive statistics for the FAS/BHP/NPS Test

	Mean	Standard Dev	Minimum	Maximum
English group FAS	35.4	14.1	20	62
Afrikaans Group FAS	18.1	10.2	3	34
Afrikaans Group BHP	23.3	9.32	10	38
isiXhosa Group FAS	20.8	8.2	7	33
isiXhosa Group NPS	26.2	10.5	9	41

Further statistical results came from the use of Group Dependent t-tests (using the STATISTICA software package) to examine statistically the difference in performance on the original versus the new test — this was done for both the isiXhosa and Afrikaans groups. Here, the results showed a statistically significant difference ($p < 0.05$) for the Afrikaans participants for their performances on the original versus the new test. For the isiXhosa

participants, no significant difference ($p < 0.05$) was demonstrated — this result might be explained by the presence of an outlier in the data, a participant (case 17) who scored a full 17 words poorer on the new test compared to the original administered in English. Once this outlier was removed from the data, a significant difference between the isiXhosa participants' performances on the original versus the new test was demonstrated ($p < 0.05$). The results also revealed that only five of the English participants achieved a perfect score of 2-out-of-2 for the new test according to the GSNSB's allocated cut-off score of 'more than 35-words' for the cumulative total over the three letters given (see Table 4.15). Also, eight isiXhosa and seven Afrikaans participants failed the new test according to these same cut-off scores.

Table 4.15

FAS/NPS/BHP Test scores for all three groups

ENGLISH											
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9	Case 10	Total
F	12	10	25	6	7	10	14	17	17	11	129
A	11	6	20	7	6	6	11	16	8	16	107
S	8	9	17	7	8	8	17	21	13	11	119
	31	25	62	20	21	24	42	54	38	38	355
isiXHOSA											
	Case 11	Case 12	Case 13	Case 14	Case 15	Case 16	Case 17	Case 18	Case 19	Case 20	Total
F	5	3	7	7	4	6	12	11	6	10	71
A	3	0	4	9	0	6	4	2	4	9	41
S	3	0	6	8	0	4	12	13	8	15	69
	11	3	17	24	4	16	28	26	18	34	181
N	4	3	7	6	4	9	3	4	13	14	67
P	9	7	11	14	2	9	4	10	11	12	89
S	8	5	7	9	4	11	4	8	9	12	77
	21	15	25	29	10	29	11	22	33	38	233
AFRIKAANS											
	Case 21	Case 22	Case 23	Case 24	Case 25	Case 26	Case 27	Case 28	Case 29	Case 30	Total
F	3	7	8	12	11	4	7	8	5	10	75
A	1	3	5	7	6	2	8	4	2	8	46
S	7	9	7	14	11	1	12	9	8	9	87
	11	19	20	33	28	7	27	21	15	27	208
B	6	4	6	18	10	3	19	11	9	12	98
H	12	7	5	14	6	5	7	9	8	14	87
P	4	9	6	10	6	1	14	6	11	11	78
	22	20	17	42	22	9	40	26	28	37	263

Discussion

The aim of the pilot study was to provide a thorough, multifaceted approach to the process of neurocognitive test adaptation in South Africa, using converging lines of evidence to implement, and then evaluate, the test adaptations. This task proved extremely challenging for a variety of reasons, not least of which was the complex matrix of the influences education, culture (and acculturation) and language on the clinical setting. This undertaking was unique, as no previous work of this kind had previously been undertaken in the South African context. The results of the pilot study served to confirm the pressing need for culturally fair neurocognitive tests in South Africa, emphasising and reiterating the day-to-day problems encountered by the neuropsychologists working in this challenging setting. This need was highlighted most poignantly by the poor performances of controls on the original tests and the homogeneity of their errors. It is also important to remember that the test piloting process was intended not only to provide effective test adaptations, but also to create a standard set of test items that could eventually be consistently used individually in the clinical setting — this applies for all the tests under review. Given the reliance on the qualitative approach, and the importance of clinical judgement in this setting, the ability to evaluate a patient's performance using a test that has been repeatedly used and investigated (including the possibility of inter-subject comparisons), is of utmost importance.

When evaluating the results from the memory tests, the comparison between the old and new versions of the 4 Hidden Objects Test produced no statistically significant difference.

Additionally, ceiling effects occurred, with the majority of control participants performing near to perfect, irrespective of language group. Given that the neuropsychologists deemed this to be the most elementary form of memory assessment they use (not cognitively taxing), this finding was both pleasing and unremarkable. Walsh (1987) emphasises that bedside

neurocognitive tests should not be taxing to normal individuals, which applies to all the tests under review in this study.

The finding that the old items did not appear culturally biased, despite the neuropsychologists having experienced problems with similar items in the past, might be explained by the fact that the items chosen were actually familiar to the participants at hand and hence easy for them to remember. Alternatively, the possibility arises that the items selected for the original test were not as culturally biased as others previously used in the clinical setting, given that over the years many items have been found to be problematic clinically. Language plays a key role in the 4 Hidden Objects Test, and is also closely related to culture (Ardila, 2007; Nell, 2000; Swartz, 1998; Uzzell, 2007). Therefore, given that the participants in this sample may have had a relatively high level of acculturation, a standard set of new, more culturally fair items is better than the uncertainty and potential bias of a random, westernised set of items.

The findings from the Digit Span/Auditory Span tests were similar to those of previous research in that the Afrikaans and English participants performed similarly irrespective of language. This was most likely due to the fact that syllable length and articulation rate in these two languages are very similar (Chincotta & Underwood, 1997). All three of the language groups performed similarly when the test was administered in English, which was a pleasing result — here, the mean score of over ‘6’ was consistent with what the neuropsychologists deem to be a normal performance, especially in a relatively poorly educated population (Kaplan, Fein, Morris & Delis, 1991; Lezak, 2004). It is likely that this commonality might be explained by the fact that English has short syllable length and articulation rate, as this observation is consistent with previous studies (Baddeley, Thomson

& Buchanan, 1975; Chincotta & Underwood, 1997; da Costa Pinot, 1991; Hitch, Halliday & Littler, 1989).

The results for the isiXhosa group, when the test was administered in isiXhosa, were most interesting, given that the articulation rate and syllable length are far longer in the isiXhosa language. On closer qualitative investigation, many of the isiXhosa participants reported that they were actually internally converting the numbers into English when they were read to them in isiXhosa. This makes their overall performance that much more interesting. When questioned as to why they did this, the majority of participants replied that they always worked with numbers in English — this finding seemed to apply regardless of their proficiency in English. There are two possible reasons for these fascinating findings. Firstly, very few South Africans are totally monolingual and most are able to speak at least one other language. Secondly, mathematics and basic numeracy are almost always taught in English, even in isiXhosa settings (often pupils are not taught at school in their home language).

Previous studies investigating the Digit Span Test have also found that performance increases when the task is administered in the same language that mathematics is taught to the individual at school (Chincotta & Underwood, 1997). Griessel (2005) supports this notion, explaining that, “many South Africans have been or are being educated in their second or third language, either by choice or as a result of Apartheid educational policies of the past. Particularly, in instances where a measure taps previously learned knowledge ... it may be fairer to assess test-takers in the language medium in which they were or are being educated, instead of in their first-language”.

This observation leads to another key point concerning the findings from the Digit Span Test. The finding that the participants’ level of education correlated highly with their performance

is also consistent with prior research outcomes (Ardila & Rosselli, 1989; Garcia-Morales, Gich-Fulla, Guardia-Olmos & Pena-Casanova, 1998; Karakas, Yahn, Irak & Erzengin, 2002; Kaufman, McLean & Reynolds, 1988; Pacaud, 1989). There are a number of possible explanations why this might be the case, including test-wiseness; the motivation of people to do well at tests and their awareness of the importance of doing so; and the ability to use rehearsal strategies (especially when using numbers), which may have been developed at school from the advantages of numeracy and numerical familiarity.

The results highlighted the fact that the participants' level of education significantly predicted success or failure on the test versions; it was, however, a better predictor of success than failure. Interestingly, some participants expected to pass on the basis of their level of education actually failed. This observation suggests that the quality of the education an individual receives is more important than the quantity and, given that the majority of the participants were non-whites who had been previously disadvantaged, this finding serves to highlight once again the diversity and challenges abundant in this clinical setting. As Nell (2000, p. 96) reiterates: “[y]ears of schooling is therefore a crude indicator of educational attainment because it says nothing about those aspects of school quality that are taken for granted in Western settings”. Needless to say, this crucial observation applies to all the neurocognitive test performances evaluated in this thesis.

For the newly developed Township Fire Story, the significant finding from the pilot study was that two of the three participant groups — namely the English speakers and the Afrikaans speakers — performed better on it than on the original Babcock Story. However, overall the isiXhosa speakers fared marginally worse on this new test in comparison to their

performances on the Babcock Story, despite the qualitative finding that the majority of them reported finding it more familiar. This result required further investigation.

A possible explanation lay with the presence of contamination (intrusions) between the two stories, as noted by the qualitative observations of participants' performances. The possible problem of contaminations occurring between two stories administered together, and thereby affecting the quality of recall, has been noted by Lezak (2004) in relation to the two Logical Memory stories of the Wechsler Memory Scale (WMS). Because the two stories share the same overall structure (that is, a disaster occurring, with a given date and place, etc.), and because they were both administered in the same testing session, a number of instances were recorded from the recalls where the date or place from the one story was recalled when recounting the other story. This did, on the whole, lower the score of whichever test was randomly administered second in any of the given sessions.

In addition to this problem, the practice effects resulting from the participant being familiar with the testing procedure and test format by the time the second story was administered could well have resulted in their fairing better with the second story relative to the first. This possibility was supported by the finding that 58 percent (17 out of 30) of the participants who were administered the Township Fire Story first fared better on the Babcock Story when it was administered second. However, despite the fact that both practice effects and contamination of story content were found to be present in this pilot data, what was still not certain was what weighting might be attributed to these factors and, more importantly, why the isiXhosa participants were the only group to fair worse on the new version. The participants' level of education could not explain this result, as the isiXhosa participants' average education was in fact one year more than the Afrikaans group's.

The results of the adapted language tests were also most pleasing. There was a significant improvement in performance in the new Naming Test over the original, as was evident from both the statistically significant results and the feedback from participants, the vast majority of whom reported a preference for the new version due to the perceived familiarity of the pictures. Close examination of the participants' performances on the original Boston Naming Test revealed the complex and subtle influence of culture, education and acculturation, and served to confirm the urgent need for the test adaptations. When the data were categorised by language group, it became clear that the disparity between the participants' old versus their new test performances was smaller for the English relative to the Afrikaans and isiXhosa groups. In other words, the English group had a relatively small difference between their old and new test scores, while there was a large difference for the isiXhosa group. Just how severe the cultural bias was on the original test is emphasised by the fact that the isiXhosa group's average score of 13.1 is far lower than Alzheimer's patients in the United States, who averaged 19.1 on the same shortened version (Lansing et al., 1999).

These findings highlight the clear presence of cultural bias when these tests are used in the South African context, as the largely 'westernised' European English speakers found the original test *relatively* unproblematic. This finding is consistent with the findings of cultural bias in the test as demonstrated in other studies where the test was administered to Australian, Maori, Hispanic and Asian populations (Barker-Collo, 2001; Brauer Boonea, Victor, Wen, Razani & Ponton, 2007; Worrall, Yiu, Hickson & Barnett, 1995). It is also noteworthy that, with the exception of a handful of the English participants who had a tertiary education, all the items incorrectly named were misnamed by participants with varying levels of education. This demonstrated that the tests' inherent problem was the cultural and language bias, but

that level of education also impacts on test performance. Again, this finding echoes those of previous studies, where it was concluded that level of education significantly affects the task (Hawkins & Bender, 2002; Heaton, Avitable, Grant & Matthews, 1999; Mitrushina et al., 2005; Saxton, Ratcliff, Munro, Coffey, Becker & Fried, 2000; Spreen & Strauss, 1998).

Most significant was the fact that in the original Boston Naming Test the majority of items resulting in errors were the same items previously identified as problematic by the neuropsychologists in the clinical setting. A clear example of this is the *pretzel* item, where 23 of the 30 participants got it wrong, with not one isiXhosa control getting it right.

Similarly, the *hammock*, *unicorn*, *Sphinx* and *beaver*, to name but a few (in total, there were 10 items which 18 out of the 30 participants or more failed), were also equally problematic to the isiXhosa participants as they were to the English ones — across the board, these items were sufficiently problematic to all the participants to warrant a loss of clinical confidence in them. This was further substantiated by the finding that many of the participants' incorrect answers mimicked paraphasias, adding credence of the problem of distinguishing between a genuine error on pathological grounds against one made as a consequence of cultural bias. This has also been experienced in other studies, such as the Cruice et al. (2000) findings in Australia, where the *pretzel* and *beaver* items were subsequently replaced.

The substantial improvement seen with the adapted Naming Test was evident not only by the participants' subjective preference for it, but also by the fact that the standard deviation across all participants was just 1.85, regardless of language spoken or level of education. The overall mean score of 27.37 out of 30 for the new test was also far more acceptable, as well as being far closer to a norm of 25.4 out of 30 for the original test derived in the United States (Lansing et al., 1999). This overall mean can be compared to the overall mean of 17.3 out of

30 for the original test in this study. It is also highly significant that not one of the isiXhosa participants preferred the original test, the majority saying that the new items were clearer and were encountered everyday unlike the original test items, many of which they reported never having seen before. Three out of the four participants with tertiary education preferred the original, saying that the new one was boring and lacked variety, which further serves to emphasise the cultural and educational disparities between a westernised background and the current clinical context.

The convincing findings from the new Naming Test demonstrated the significant step forward made by this study in providing culturally fair assessment for South African patients. In the light of the continued quest for clinical fairness and accuracy, further areas of possible improvement to this test were identified. Firstly, it was evident from the results obtained that a number of the newly created and piloted items were problematic — the *hippopotamus*, the *mug* and the *dragonfly* were all frequently misnamed. Consultation with the cultural experts over these results produced some interesting observations. For example, upon investigation it was pointed out that the *hippopotamus* is not found in the rivers of the Eastern Cape region (from whence the majority of isiXhosa people originate). Therefore, for them, the animal bearing the closest resemblance to this strange creature was a *pig*, being the response the majority of participants gave.

For the *mug* item, a large number of the participants misnamed it as a *cup*, revealing possible ambiguity with the item and the potential for paraphasic errors to be missed. Consultation with the cultural experts revealed that there are two separate words for these two items (*ikommityi* and *imug*) in the isiXhosa language, so the errors could not be explained by the possibility of one word existing for both. Therefore, the participants' confusion with this item

probably stemmed from the level of acculturation in the sample, and the possible fact that isiXhosa individuals from rural backgrounds do not regularly encounter mugs.

Finally, the third problematic item identified from the new Naming Test was the *dragonfly*, which 17 of the 30 participants named incorrectly. However, on closer scrutiny of the qualitative feedback results, and after consultation with the cultural experts, it was seen that although the majority of the participants did not know the name, they acknowledged that they were familiar with the item and had seen it before. In support of this qualitative feedback, many of them were able to identify it as a type of insect. The cultural experts confirmed that both Afrikaans and isiXhosa words do exist for *dragonfly*, hence ruling out the lack of a name for this item in these languages as a potential source of their errors.

In addition to these three problematic items, another area identified for improvement in the new Naming Test was the average score of 27.37 out of 30, which was relatively higher than international averages for the original Boston Naming Test. Therefore, along with the need to replace the above-mentioned items, it was also necessary to include one or two more difficult (yet still culturally fair) items to replace some of the simpler ones.

For the final adapted language test, the Washing Line Picture Test, the results revealed that no statistically significant difference could be discerned between the original and the newly created test; however, the vast majority of the pilot participants preferred the new version on the grounds that it was more familiar to them. Therefore, the new version was deemed fit for inclusion into the GSNSB ahead of its validation.

When the language tests were tentatively administered to the three aphasic patients, the results were positive. For example, the new test's ability to detect paraphasias was evident from the performance of one of the Broca's aphasics where the *ostrich* was named a *hen*, the *newspaper* named *letters*, and the *locust* a *bat*. The fact that the pilot study results had already shown that normal controls were able to consistently name these items correctly adds credence to the assumption that these errors were a result of the patient's pathology rather than the work of cultural bias. The Wernicke's aphasic performed poorly on the new test and it was discontinued after five consecutive errors. However, the results still showed that the new test was able to pick up perseverations, neologisms and semantic errors. Overall, the new test can be deemed a success, as the controls performed well on it in relation to the original, while the patients struggled, but not seemingly on cultural or educational grounds.

When the Washing Line Picture Test was given to the aphasic patients, the results were most promising, especially when analysed in relation to the controls' performances. Here, the new test was successful in eliciting a number of the patients' aphasic deficits. For example, with one of the Broca's patients, the new test managed to demonstrate her lack of verbal fluency (slow speech rate), problems with pronunciation, and poor sentence structure, noted by the misuse of prepositions and verbs. Although some of these problems were also revealed by the original Cookie Theft Test, what is most salient here was that the controls did not have similar problems on the adapted test. All in all, the new test is sufficiently capable of detecting aphasics' deficits, in a way that South African participants find familiar. This tentative finding boded well for the validation process to follow.

The findings from the new Mary Selo Test and the original Anna Thompson Story revealed that both tests proved equally effective as reading tests for the control participants. However,

given the overall confidence of the cultural experts in the Mary Selo Test's suitability as being culturally fair to both rural and urban individuals, regardless of language group, the test was deemed appropriate for inclusion into the GSNSB. An additional advantage of the successful creation of this test is that it has laid the groundwork for the possibility of the test being used as a memory test, should it be required, and provided that it is first validated appropriately.

With regard to the tests in the Spatial Cognition section, the new 3-D Analysis Test proved the most difficult of the nine newly developed tests both to create and to perfect. Part of this complexity lay in the fact that the original Cube Analysis Test is considered to be one of the most difficult forms of assessment of spatial cognition and is also widely acknowledged as being especially sensitive to both education and cultural factors (Ardila & Keating, 2007; Lezak, 1995; Sugarman, 2007). Given this background, developing a more appropriate version for the South African context was never going to be straightforward. When the possibility arises that certain cultures are weaker on spatial tasks than others due to their specific environmental exposure as part of their upbringing, the responsibility of making the items in a new test uniform and accurate in perspective becomes that much more acute and the challenge greater.

It was clear from both the qualitative results and feedback from the participants' performances on the 3-D Analysis Test that this initial attempt still required much work. On a positive note, these qualitative results did, however, reveal that the majority of participants subjectively preferred the new test, primarily on the grounds that they found the items in the original test to be too small, while the new test's items had intentionally been enlarged.

Part of the problem with designing this new test was finding a way to draw 3-dimensional figures that were both uniform and clear in their perspective. Only after the initial attempts proved unsuccessful was it discovered that the use of a computer program was required. Although it was decided early on in the test's development that the items could not be accurately drawn free hand, it was only once the use of basic computer programs was attempted and failed that an architect's advanced program was used. Overall, the results indicated that the participants performed poorly on both the original and the new test, highlighting the difficulties with this type of perceptual task, as noted by Lezak (1995).

The results of the other adapted test in the Spatial Cognition section — the Hut Drawing Test — proved far more pleasing, the vast majority of participants preferring the new test. Twenty-eight of the participants performed the same on both the original and the new version, while two performed better on the new version. Both the right middle cerebral artery stroke patients preferred the new version, which was able to effectively demonstrate their neglect. Given that this is not a cognitively taxing test for neurocognitively normal individuals, it was not surprising that the majority of the participants performed equally well with the original.

From the Executive Function section, the findings from the new NPS Test (isiXhosa), BHP Test (Afrikaans) and the FAS Test showed an overall improvement over the original FAS Test when administered in English to isiXhosa and Afrikaans first-language speakers. These results were not surprising, as one would expect people to perform better on a test administered in their first language as opposed to one administered in their second- or third-language (Nell, 2000). The evidence of a large outlier in the isiXhosa sample might well have explained this group's overall average score being slightly lower.

All in all, the English participants scored higher than the other two language groups, who performed similarly to one another. On inspection, the English participants had an average level of education of 12.9 years as opposed to the Afrikaans participants' average of 8.7 years and the isiXhosa participants' average of 9.7 years of education. Evidently, when examining the overall performance of the participants, the original cut-off scores created in the GSNSB (that is, 'more than 35 words', 'between 25 and 35 words', and 'less than 25 words') were too high for the South African populace, while the English participants' higher scores could be explained by their superior level of education. This finding that the participants' level of education impacted on their COWAT performance is again consistent with those of previous studies (Loonstra, Tarlow & Sellers, 2001; Spreen & Strauss, 1991; Tombaugh, Kozak & Rees, 1999).

The problem of the original cut-off scores being too high was rectified prior to the 'validation' phase of the GSNSB by lowering the cut-off scores by 10 points. Therefore the new cut-offs were: 'more than 25 words', 'between 15 and 25 words', and 'less than 15 words', respectively. Through this approach, it was hoped to achieve uniformity across the test performances of the three language groups. Such uniformity was highly important because one set of predetermined cut-off scores was required in order to be included in the GSNSB prior to the commencement of the validation process.

The justification for this specific change stemmed from the performances of the Afrikaans and isiXhosa participants, who scored on average 23 and 26 words respectively for the test, as opposed to the English participants' 35 words. Clearly, the level of the education of the English sample was unrepresentative of the general Groote Schuur Hospital patient

population. Therefore, the cut-offs were lowered to achieve a level where all three groups passed. The success of this modification of cut-off score would then be ascertained when the test was examined more stringently as part of the validation study (see Chapter Six).

In summary, the overall results of this pilot phase of the new neurocognitive tests' development already represented a significant contribution to patient care in South Africa. The newly adapted tests had been shown to greatly reduce the impact of language, culture and acculturation on neurocognitive test performance — clearly demonstrated by the overall improvement in the test performances of the neurocognitively intact individuals on the new tests when compared to their performances on the originals. Despite the importance and positive implications of these results, two important facts had to be borne in mind. Firstly, further adaptation of certain tests was still required (see Chapter Five). Secondly, this adaptation process was not designed to contribute towards a culture-free screening tool, as such a tool cannot exist. Rather, it was a real and meaningful attempt to nullify some of the gross cultural bias known to exist clinically in many of the neurocognitive tests used in South Africa. Such an endeavour had never been undertaken before in South Africa, and therefore the intention of this aspect of the research was to improve the clinical confidence of the given tests, thereby reducing the amount of clinical judgement and guesswork that was required.

On the basis on the above-mentioned findings and conclusions, the Auditory Span Test, the 4 Hidden Objects Test, the Washing Line Picture Test, the Mary Selo Story and the Hut Drawing Test were all suitable for immediate inclusion into the GSNSB ahead of the validation phase of the study (see Chapter Six). The Township Fire Story, the Naming Test and the 3-D Analysis Test required further investigation.

CHAPTER FIVE: RE-PILOTING THREE NEUROCOGNITIVE TESTS

This chapter builds on the work carried out in the development of the newly created South African neurocognitive tests by describing the re-piloting of three of the tests that had been shown to be problematic.

Rationale

Following the initial piloting of the nine newly developed South African neurocognitive tests (see Chapter Four), further modifications to certain of these tests were required, based on the findings and recommendations of the initial pilot study. Specifically, the Naming Test, the 3-D Analysis Test and the Township Fire Story all required additional investigation and further modification before they could be regarded as suitable for introduction into the Groote Schuur Neurocognitive Screening Battery (GSNSB) ahead of its validation. Consequently, the following re-piloting study was conducted to investigate and then rectify the specific concerns with each particular test. These test-specific concerns are outlined in the ‘Procedure’ section below.

The aim of re-piloting the Township Fire Story was to attempt to rectify the disparity found between the isiXhosa participants’ performance on the test versus the Afrikaans and English speakers, as the isiXhosa speakers were the only group who had performed better (marginally so) on the Babcock Story than the new version. The aim of re-piloting the Naming Test was to rectify the slight ceiling effect that had been noted by introducing new, more difficult (yet still culturally fair) items, while at the same time replacing the *mug* and *hippopotamus* items, which had been found to be inappropriate. The central goal in re-piloting the 3-D Analysis

Test was to see if the performance of neurocognitively intact individuals on the test could be improved by adjusting the perspective of the depicted items.

Methodology

Sample

Thirty neurocognitively intact participants were selected from the general Groote Schuur Hospital patient population in order to constitute a demographically representative sample. This was an entirely new sample — that is, none of these participants had been involved in the initial pilot study. Of these 30 participants, 12 were female and 18 were male, with the average age of the sample being 42 years. The youngest participant was 16, while the oldest was aged 74. Each participant was first screened using the same Screening Sheet used in the initial pilot study (see Appendix H), to exclude pathologies that might have neurocognitive consequences, thereby ensuring that they were neurocognitively normal. As was the case with the initial pilot study, this sample comprised ten isiXhosa, ten English and ten Afrikaans participants. The average age of the isiXhosa group was 35.6 ($SD = 12.31$) and their average level of education was 8.4 ($SD = 2.91$) years. The average age of the English group was 49.7 ($SD = 12.75$) and their average level of education was 10.2 ($SD = 1.93$) years. The average age of the Afrikaans group was 44.22 ($SD = 17.21$) and their average level of education was 7.8 ($SD = 3.35$) years.

Materials

As previously mentioned, the materials used for the re-piloting included the original 3-D Analysis Test, Township Fire Story and the Naming Test, as initially piloted (see Chapter Four). In addition, given that further modifications had been made to these tests ahead of their re-piloting, the materials used also included these three tests in further modified form.

For this re-piloting the English, Afrikaans and isiXhosa GSNSB Prototypes were used to administer the tests' instructions. The remaining materials used for this aspect of the research comprised the Consent Form (see Appendix I) and Patient Information Sheet (see Appendix J) and the Screening Sheet (see Appendix H) used to exclude potential pathologies in the control participants. Finally, the specially designed Scoring Sheet (see Appendix K) was again used to ensure that all participants' test scores, as well as the qualitative observations about the tests made by both the participants and the assessors, were accurately recorded.

Design

The design chosen for this re-piloting study was envisaged to mimic that of the original pilot study given the paramount goal of adopting converging lines of evidence in order to accurately evaluate the test modifications. Therefore, once again both qualitative and quantitative measures were adopted. The qualitative data drawn upon incorporated the feedback and advice of the neuropsychologists, the cultural and language experts, and the patients themselves, while the quantitative data took the form of the control participants' test scores on the three tests. The participants' feedback again involved asking which tests items they preferred and why, and which items they found problematic, and why.

The central purpose of this design was to provide a comprehensive appraisal of the tests to ensure their suitability for inclusion in the GSNSB ahead of its validation. Given that six of the nine newly developed neurocognitive tests had already met the required standard, it was of utmost importance to ensure that the remaining three tests also met these criteria, the ultimate aim being to ensure that the best possible clinical service could be provided to future patients.

Data Analysis

The data gathered from the three tests was analysed from both a qualitative and a quantitative point of view. While the qualitative protocol was the same for all three tests, the exact quantitative measures varied depending on the specific investigation required for each individual test.

For the Township Fire Story, the data analysis involved the use of descriptive statistics, along with the one-way ANOVA procedure to investigate possible differences in test performance between the three language groups. The factorial ANOVA procedure was also used to investigate whether the participants' language or level of education affected their performances on this test.

For the re-piloting of the four new Naming Test items, chi-squared tests and t-tests were used to compare the three language groups, in order to determine whether the participants from the three groups performed similarly. The qualitative analysis was used to investigate whether the participants found the items recognisable. In addition to this, once the old items from the original Naming Test (that is, the *mug* and *hippopotamus*) had been replaced with new improved items, factorial ANOVA was used to determine the success of the new changes by examining the influence of age, education and language on the participants' performances.

The data analysis used for the 3-D Analysis Test required a comparison to be made between the participants' performance on the original versus the newly modified version. For this analysis, t-tests were used. Additionally, factorial ANOVA was used to investigate the influence of language, age and education on 3-D Analysis Test performance.

Procedure

General Procedure

The general procedure followed during this re-piloting process largely mimicked that of the initial pilot study. All assessments were conducted at Groote Schuur Hospital and each participant was first given the Consent Form and Patient Information Sheet. Both isiXhosa and Afrikaans interpreters were employed to ensure that each participant was assessed in his/her first language, using the relevant test instructions from the newly translated isiXhosa and Afrikaans GSNSB Prototypes. These interpreters were the same individuals who had served as translators of the GSNSB Prototypes and were members of the panel of cultural and language experts, as well as having participated in the initial pilot study. All participants' answers were recorded using the Scoring Sheet.

Each assessment lasted from 30 to 45 minutes. For the Naming Test and the 3-D Analysis Test, both the original and the modified versions were administered in order to compare the participants' performances on both tests. For the Township Fire Story, only the modified version was used, and was administered to all 30 participants.

Procedure Followed in Modifying the Tests

The procedure followed in further adapting each of the three tests was specific to each test.

Modification of the Township Fire Story

With regard to the Township Fire Story, the major finding of the pilot study had been a discrepancy between the performance of the isiXhosa group relative to the Afrikaans and English groups. While the English (averaging 31.3 on the original test and 35.5 on the new version) and Afrikaans participants (averaging 27.1 on the original test and 29.2 on the new

version) fared better on the new story than on the original Babcock Story, the isiXhosa participants had performed better on the original test (averaging 31.6 on the original test and 28.8 on the new version).

It was suspected that there were three possible reasons for this disparity. The first of these was the presence of an outlier in the isiXhosa data, which served to skew the groups' overall results. This participant had scored just 12 and 17 for the original and new stories respectively. The second possible reason was that many contaminations had occurred between the content of the Babcock Story and the Township Fire Story, especially in the isiXhosa sample. This was largely a consequence of the logistical limitations of the previous pilot study where both tests had to be administered in the same session. Furthermore, there was evidence of a practice effect, as 58 percent of the participants who were administered the new story first had subsequently performed better on the Babcock Story when it was administered second. A third minor reason was that the qualitative results had shown that, on the whole, the participants had struggled with the word *week* in the story, thinking that the story had actually occurred recently, despite being told that it was not true. It was also observed that the words *the flames* appeared twice in the new story, which affected the scoring as two points could not be given for the same item. As a result of these observations, when re-piloting the Township Fire Story, it was administered without the Babcock Story to a new sample of controls in order to eliminate the effects of contamination.

After consulting the cultural and neuropsychological experts again and examining the qualitative findings from the initial study, some modifications were made to the story. Firstly, the word *week* was replaced with the word *year* in order to make the story sound more likely to be fictitious. Secondly, the final sentences were changed from '*... while trying to put out*

the flames and rescue their possessions from the fire. It took eight hours before the fire was extinguished’ to ‘...while trying to save their possessions. In rescuing a child who was trapped in a shack, a woman broke her arm’. This was done so that ‘*the flames*’ did not appear twice. Once these changes had been made (see Appendix B), they were reviewed by the panel of experts, who subsequently approved them.

Modification of the Naming Test

The Naming Test required further modification as the *mug*, *hippopotamus* and *dragonfly* items had all been identified as problematic during the pilot study. It had also been noted that a slight ceiling effect was present — normal controls had performed about two points higher (with an average of 27.37) on the new Naming test than the United States average of 25.4 for the Boston Naming Test (Lansing, et al., 1999). To rectify this, it was decided to include some more difficult items.

The cultural and language experts were consulted to discuss why the *mug*, *hippopotamus* and *dragonfly* had been found problematic, especially by the isiXhosa speakers. Following investigation of the qualitative feedback of the participants together with the experts’ observations, it was discovered that, in actual fact, the *dragonfly* was a culturally fair item, despite the high failure rate. This was because almost all the participants reported that they recognised the picture, even though the name escaped them. It was therefore decided to retain the *dragonfly* item in the final test. Investigation of the *hippopotamus* item revealed that a large number of the participants, especially among the isiXhosa group, had struggled to name it because this animal is not found in the Eastern or Western Cape, where many of them had originated. Therefore, a replacement item was sought.

Finally, the *mug* item was found to have been problematic because isiXhosa people from rural settings were considered not to have been familiar with a *mug* due to their lack of acculturation, with the result that many of the participants incorrectly named it as a *cup*. Consequently, a replacement item for the *mug* was also sought. This culture-based error, which could easily be mistaken for a semantic paraphasia, underscores the importance of providing culturally fair items.

The procedure followed in identifying and creating replacement items was therefore based on the need to find items which were not only culturally fair but which were also more difficult than the items they were replacing. Although this task proved far from easy, four potential items were identified: a *spur* (from a chicken's leg) (see Figure 5.1), a *husk* (as part of a corn on the cob) (see Figure 5.2), a *pylon* (see Figure 5.3) and an *aloe* (a plant common to the Western and Eastern Cape regions) (see Figure 5.4).

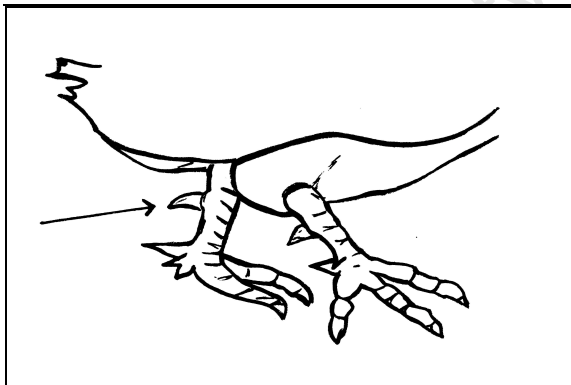


Figure 5.1: Spur



Figure 5.2: Husk

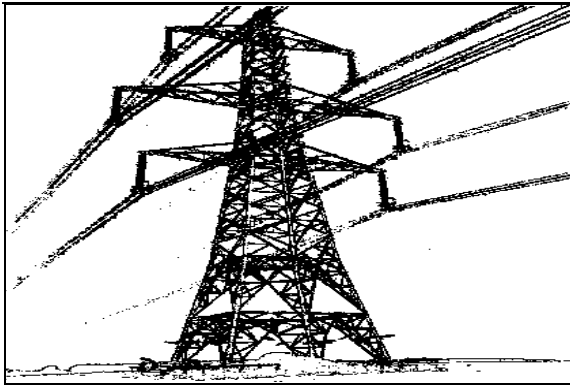


Figure 5.3: Pylon

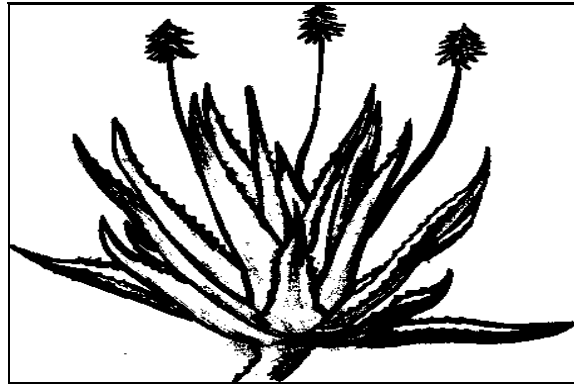


Figure 5.4: Aloe

These four items were all approved by the cultural and neuropsychology experts as suitable for re-piloting. They agreed that the items should be familiar to both rural and urban participants from all three language groups. It was decided that, although only two items from the original Naming Test needed to be replaced, all four new items would be re-piloted and the two on which the participants performed best would be included in the final test.

Modification of the 3-D Analysis Test

The 3-D Analysis Test required further modification following the initial finding that controls had performed poorly on both this test and the Cube Analysis Test. The new test was found to be problematic for a number of reasons. Firstly, the depicted items were too small for all participants to see properly and, secondly, some of the blocks making up the 3-dimensional figures were distorted in perspective (see Figure 5.5 for this problem). This had been noted both by the neuropsychologists and from the participants' qualitative feedback. The result was that some of the items were difficult to interpret visually. A third problem was that, given the sensitivity of visuospatial tests to the influence of education and cultural factors, and the accepted difficulty of this particular visuospatial task highlighted in the available literature,

some of the items in the test were too complex for use in the South African context (Lezak, 1995; Rosselli & Ardila, 2003).

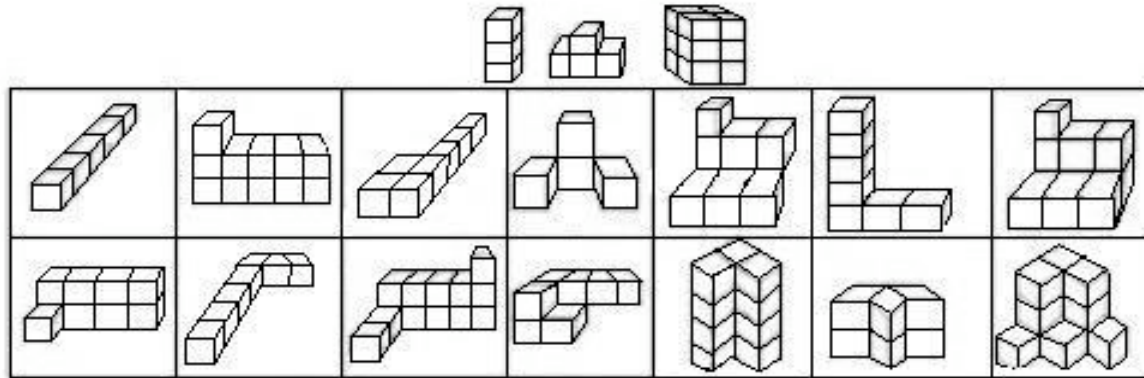


Figure 5.5. Distortion problems in the originally created 3-D Analysis Test

To address these problems, a number of modifications were made. Firstly, an analysis was conducted to determine which items the majority of participants in the initial pilot study had failed. Not surprisingly, the more complex items (that is, those composed of the most 3-dimensional blocks) presented with the highest error rates. In addition, the items where the perspective was most distorted also produced high error rates. These newly identified items were then further simplified by removing some of the blocks and, in certain cases, also altering the perspective of the figures. The most problematic of the items were replaced with entirely new, more simple designs. A computer program called Blender was used to accurately adjust and rectify the 3-dimensional perspectives of the figures.

Secondly, all of the test's 14 items were enlarged in order to make them easier to see — this was also done using the Blender computer program. Once these changes had been made, the neuropsychological and cultural experts were consulted to judge the efficacy of the modifications and to suggest further improvements to the 14 items prior to re-piloting.

Results

Township Fire Story

Qualitatively, none of the participants expressed any concerns or dislikes regarding the content of the story. The means and standard deviations derived from the quantitative analysis are provided in Table 5.1.

Table 5.1.

Means and standard deviations of participants' performances on the Township Fire Story

	Language Group		
	isiXhosa	Afrikaans	English
<i>M</i>	35.6	41.4	49.7
<i>SD</i>	12.31	17.21	12.76

Figure 5.6 represents the performance of the participants from the three language groups on the re-piloted Township Fire Story. Here, the isiXhosa participants performed the best on the test, with the Afrikaans and English participants fairing similarly.

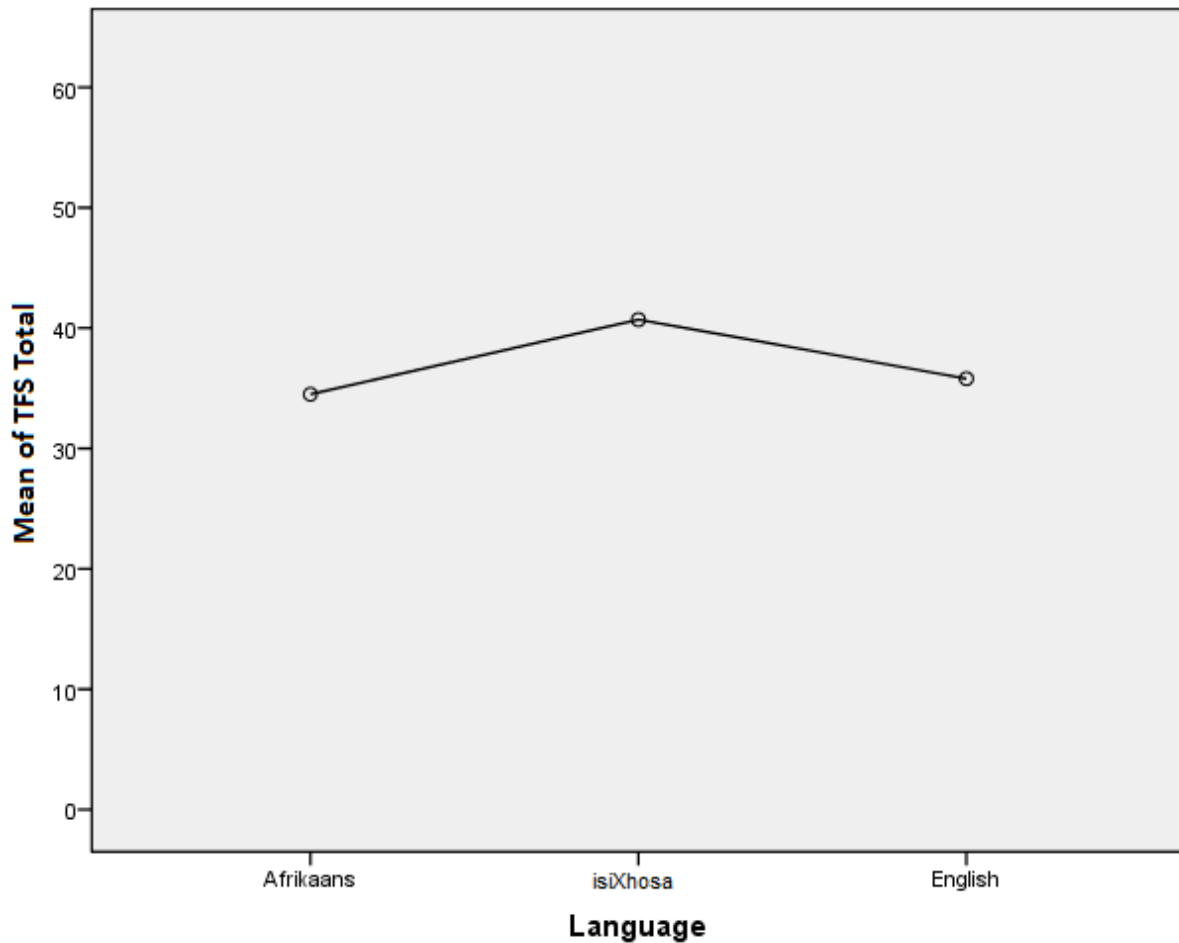


Figure 5.6. Cell mean plot showing performance by language group for the Township Fire Story

One-way ANOVA was used to calculate the probability of the three languages' performances resulting from the same population. ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated. However, given that ANOVA is a robust test and the sample sizes were equal, the analyses were continued. No significant effects were noted, $F(2, 27) = .415, p = 0.664$, demonstrating that the Afrikaans, isiXhosa and English participants all performed similarly on this updated test.

Factorial ANOVA was used to investigate whether the participants' level of education affected their test performance. Table 5.2 shows the descriptive statistics (means and standard deviations) of the analysis. The question whether the participants' level of education, in

conjunction with the language they spoke, influenced their test performance was also investigated. Given that the data was derived from a small sample, the level of education of the participants was divided into two groups: those with less than nine years of education and those with nine or more years, which also allowed for two groups of as equal size as possible. No significant main effect was found for language, $F(2, 30) = 1.24, p = 0.307$, but a significant effect was found for education, $F(1, 30) = 7.5, p = 0.012$. There was no significant interaction effect between education level and language spoken, $F(2, 30) = 0.574, p = 0.571$. These results indicated that level of education significantly affected performance on the updated Township Fire Story (0.012).

Table 5.2.

Descriptive statistics for the Township Fire Story

Factor A: Education		Factor B: Language			A marginal means
		English	Afrikaans	isiXhosa	
9 years or more	<i>M</i>	39.88	46.75	45.25	42.94
	<i>SD</i>	14.015	19.97	10.468	
	<i>N</i>	8	4	4	
less than 9 years	<i>M</i>	19.5	26.33	37.67	30.21
	<i>SD</i>	17.678	15.253	12.42	
	<i>N</i>	2	6	6	
B marginal means		35.80	34.50	40.70	

Naming Test

The participants' qualitative feedback revealed that all reported being able to recognise both the *aloe* and the *pylon* items, despite their high error-rate in naming both items correctly. It was also revealed that most participants were unfamiliar with the *spur* and the *husk* items — many of them called the *spur* a *claw* and the *husk* was called *leaves* (or *blare* in Afrikaans).

The quantitative results showed that the four newly introduced items all produced similar performances across the 30 control participants, with 15 failing on the *husk*, 22 failing on the *pylon*, 14 failing on the *aloe*, and 21 failing on the *spur*. To confirm whether or not these items were influenced by the participants' language group, chi-squared tests were used using the $\alpha = 0.05$ significance level. The results showed that language did not influence the results on the four items: for the *pylon* $\chi^2(2, N = 30) = 4.42, p = 0.050$, *spur* $\chi^2(2, N = 30) = 2.85$, *husk* $\chi^2(2, N = 30) = 2.400$ and *aloe* $\chi^2(2, N = 30) = 1.070$ items.

Given these findings, it was decided to replace the problematic *hippopotamus* and *mug* items in the Naming Test with the *aloe* and *pylon* and thereafter run additional analyses in order to determine the success of this modified test. With its two newly replaced items, the participants' average score for the test was calculated to be 25.97 out of 30, with a standard deviation of 3.13 (see Table 5.3 for the language breakdown of descriptive statistics).

Table 5.3

Means and standard deviations of Naming Test scores for isiXhosa, English and Afrikaans groups

	Language Group			
	isiXhosa	English	Afrikaans	Total
	Naming Test			
<i>M</i>	25.9	26.2	25.8	25.97
<i>SD</i>	3.21	3.46	3.05	3.13

The analysis also involved using a cell mean plot to portray the performances of the three language groups, which were all similar (see Figure 5.7). One-way ANOVA was used to calculate the probability of the three languages' performances resulting from the same population — no significant effect was found, $F(2, 27) = 0.04, p = 0.960$, demonstrating, through this high probability, that the Afrikaans, isiXhosa and English participants all performed similarly on this new test.

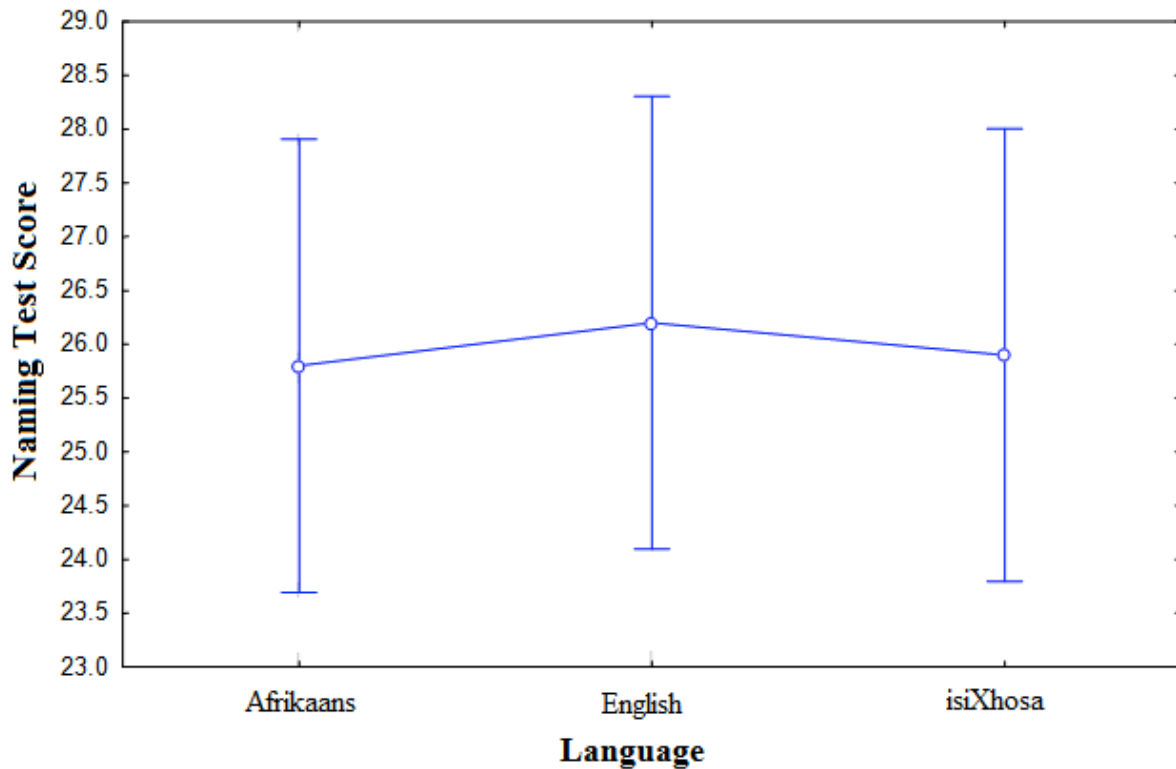


Figure 5.7. Cell mean plot showing Naming Test performances by language group

Factorial ANOVA was used to investigate whether the participants' level of education affected their test performance. The question whether the participants' level of education, in conjunction with the language they spoke, influenced their test performance was also investigated. Given that the data were derived from a small sample, the level of education of the participants (with a mean of 8.8 years) was divided into two groups: those with eight years of education or less and those with more than eight years, which is conveniently the difference between primary school and high school, while also allowing for two groups of as equal size as possible. The results revealed that assumptions of normality and homogeneity of variance were upheld. Table 5.4 shows the descriptive statistics (means and standard deviations) of the analysis. No significant main effect was found for language, $F(1, 39) = 0.62, p = 0.546$, but a significant effect was demonstrated for education,

$F(2, 39) = 11.57, p = 0.002$. There was no significant interaction effect between education level and language spoken, $F(2, 39) = 0.43, p = 0.655$. These results indicated that level of education significantly affected performance on the modified Naming Test ($p = 0.002$).

Table 5.4

Mean Naming Test score

Factor A: Education	Factor B: Language			A marginal means	
	English	Afrikaans	isiXhosa		
8 years or more	<i>M</i>	27.25	28.00	27.20	27.41
	<i>SD</i>	2.55	0.82	2.05	
	<i>N</i>	8	4	5	
less than 8 years	<i>M</i>	22.00	24.33	24.60	24.08
	<i>SD</i>	4.24	3.14	3.85	
	<i>N</i>	2	6	5	
B marginal means		26.20	25.80	25.90	

The participants were divided into two age groups in order to establish the effect of age on Naming Test performance, these two groups being individuals aged 40 or younger and those over 40 years old. To establish the effect of age, a t-test was used, resulting in a significant effect being found, $t(28) = 2.64, p = 0.014$. The results showed that the participants aged below 40 scored on average higher ($M = 27.54$) than those over 40 ($M = 24.76$).

3-D Analysis Test

Table 5.5 shows the descriptive statistics for how the three language groups performed on the new 3-D Analysis Test. The results of a t-test, performed in order to compare the participants' performance on the original versus the modified 3-D Analysis Test, were that no significant difference was found, $t(28) = 0.49, p = 0.620$. The participants scored an average of 8.63 out of 14 on the new test, and an average of 8.1 for the original 3-D Analysis Test.

Table 5.5

Means and standard deviations of 14 item 3-D Analysis Test scores for isiXhosa, English and Afrikaans groups

	Language Group			
	isiXhosa	English	Afrikaans	Total
	3-D Analysis Test			
<i>M</i>	8.7	10.8	6.4	8.63
<i>SD</i>	3.02	4.34	4.27	4.21

To investigate the possible role the participants' level of education and language played on the test results, analyses were performed using factorial ANOVA. Here, assumptions of normality and homogeneity of variance were upheld. Once again, the participants' level of education was broken down into two groups: more than eight years and eight years or less.

Table 5.6 provides the descriptive statistics (means and standard deviations) for this analysis.

A significant main effect was found for education, $F(1, 39) = 16.28, p < 0.001$, but not for language, $F(2, 39) = 0.88, p = 0.427$. This revealed that there was no significant difference in performance between the three language groups when compared to each other. A significant interaction effect was found, $F(2, 39) = 3.44, p = 0.048$.

Table 5.6

Mean score for new 3-D Analysis Test

Factor A: Education	Factor B: Language			A marginal means	
	English	Afrikaans	isiXhosa		
8 years or more	<i>M</i>	12.63	9.25	9.2	10.82
	<i>SD</i>	1.30	3.59	3.11	
	<i>N</i>	8	4	5	
less than 8 years	<i>M</i>	3.5	4.5	8.2	5.77
	<i>SD</i>	4.95	3.78	3.19	
	<i>N</i>	2	6	5	
B marginal means		10.8	6.4	8.7	

From the significant main effect finding for education, it was evident that the participants' level of education did influence their 3-D Analysis Test performance. Given this outcome, cell mean plots (see Figure 5.8) were used, demonstrating that the participants with more than eight years of schooling did significantly better than those with eight years or less.

The findings from the factorial ANOVA analyses also revealed a significant interaction effect between the participants' level of education and the language that they spoke. To explore this interaction effect further, a Tukey's HSD (Honestly Significant Difference) was used to ascertain where exactly the differences had occurred. It was discovered that the significant difference lay between the English group with a high education level and the Afrikaans group with a low education level ($p = 0.001$). A significant difference was also discovered within the English group between the high and low education levels ($p = 0.011$). Referring again to Figure 5.8, the cell mean plot shows that the interaction effect occurred at the isiXhosa language level. The isiXhosas' level of education did not influence their test performance, whereas the English and Afrikaans groups performed significantly worse if they had a low level of education (that is, eight years or below).

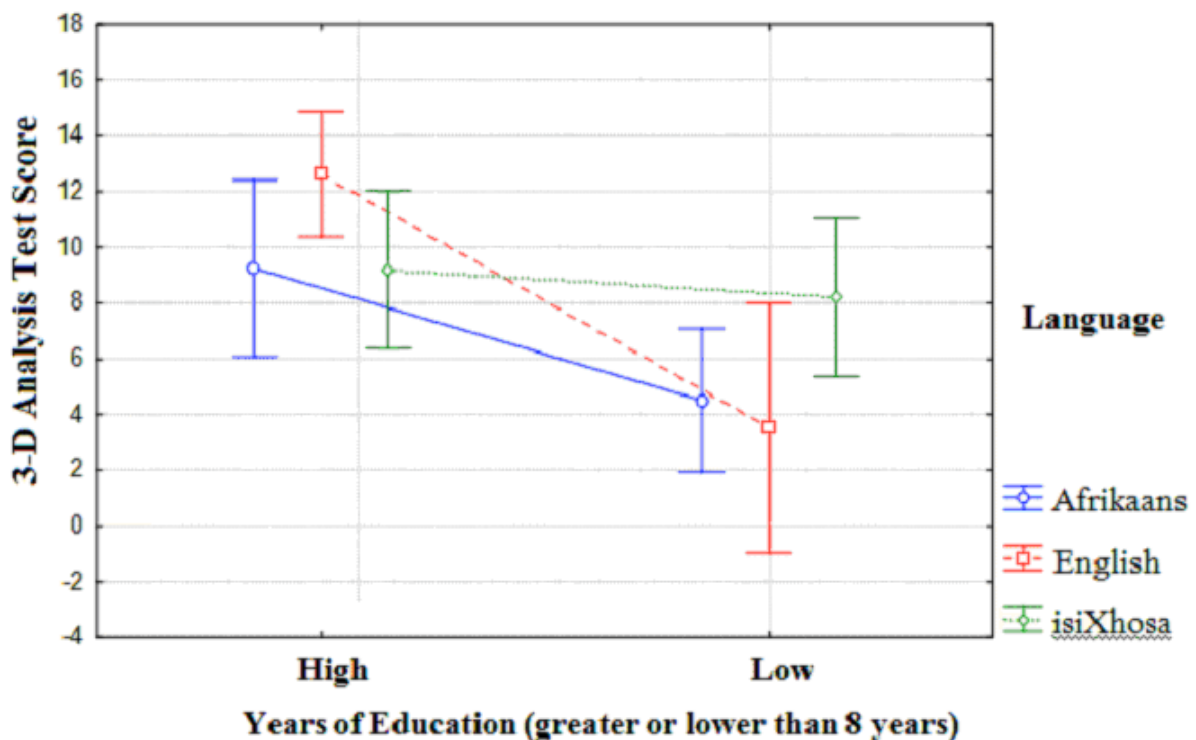


Figure 5.8. Cell mean plot showing 3-D Analysis test performances by language group and education level

In order to determine whether the participants' age influenced their test performance, an independent sample t-test was performed. For this purpose, the participants were divided into two groups according to age — those younger than 45, and those who were aged 45 years and older — which, being an even divide in terms of participant numbers, also allowed for ANOVA tests to proceed even if the assumption of homogeneity of variance was violated. The results showed that age did not influence 3-D Analysis Test performance, evident in that no significant effect was found, $t(28) = 0.85, p = 0.400$.

It was decided, based on the participants' poor performance on the new 3-D Analysis Test — despite the further modifications — to try to isolate the individual test items that were producing the highest error rates. Through this process, six of the 14 items were found to have been failed by more than a third of the controls. Subsequently, it was decided in consultation with the neuropsychology experts to remove four of the six most problematic items, in order to create a simpler, 10-item 3-D Analysis Test. Once the four items had been excluded, the average score of the participants was 7.03 out of 10, compared to the original 8.63 out of 14 (which is the equivalent of 6.16 out of 10). Descriptive statistics can be seen in Table 5.7.

Finally, in order to see what effect the participants' level of education and language spoken had on this new 10-item test, factorial ANOVA was used (see Table 5.8 for descriptive statistics (means and standard deviations)). A significant main effect was found for education, $F(1, 39) = 16.72, p < 0.001$, demonstrating the strong influence that level of education played on the participants' test performances. No significant effect was found for language, $F(2, 39) = 1.29, p = 0.294$, which clearly did not affect test performance. No significant

interaction effect, $F(2, 39) = 2.21, p = 0.131$ was demonstrated between level of education and language.

Table 5.7

Means and standard deviations of 10 item 3-D Analysis Test scores for isiXhosa, English and Afrikaans groups

	Language			Total
	isiXhosa	English	Afrikaans	
	3-D Analysis Test			
<i>M</i>	7.5	8.1	5.5	7.03
<i>SD</i>	2.64	3.11	3.1	3.07

Table 5.8

Mean score for 10-item version of new 3-D Analysis Test

Factor A: Education	Factor B: Language			A marginal means
	English	Afrikaans	isiXhosa	
8 years or more	<i>M</i> 9.38	7.75	8.20	8.65
	<i>SD</i> 0.74	1.71	2.49	
	<i>N</i> 8	4	5	
less than 8 years	<i>M</i> 3.00	4.00	6.80	4.92
	<i>SD</i> 4.24	2.97	2.86	
	<i>N</i> 2	6	5	
B marginal means	8.10	5.50	7.50	

Discussion

The results of this re-piloting study can be seen as representing a significant step towards the goal of providing culturally appropriate, diagnostically meaningful neurocognitive tests in the South African medical context. The modifications made to all three of the tests were successful, given the overall test performances of the control participants. All three neurocognitive tests were therefore deemed suitable for inclusion into the GSNSB.

The central goal of re-piloting the Township Fire Story was to rectify the problem of the isiXhosa participants having performed better on the Babcock Story, when the Afrikaans and English speakers had faired better on the Township Fire Story. The Naming Test was re-

piloted in an attempt to reduce the slight ceiling effect present, while at the same time replacing two items, the *mug* and *hippopotamus*, previously identified as culturally inappropriate, with culturally fair replacements. The 3-D Analysis Test was re-piloted in order to correct the high error rate occurring in control participants across all language groups. To achieve this, the size and perspective of the test items were adjusted, and the most problematic items were identified, removed and replaced, eventually resulting in a simplified test comprising 10 items (as opposed to the original 14 items).

The findings from the re-piloting of the Township Fire Story were most pleasing, indicating that the modifications to the test, along with the exclusion of the contamination (intrusion) effects caused by administering two stories in the same session, produced successful results. The isiXhosa group now performed the best of the three language groups and, most importantly, all the groups performed equally well. Education was found to significantly influence the participants' performances on the Township Fire Story, which is consistent with the literature (Nell, 2000).

The success of the re-piloting of the Naming Test is evident from a number of the findings. The inclusion of the more difficult, yet still culturally fair, *pylon* and *aloe* items resulted in the average score on the test dropping to 25.97 out of 30, thereby eliminating the ceiling effect found with the original test where the participants had averaged 27.37 out of 30. This new average was also similar to the average of 25.4 out of 30 found in United States populations, which is a pleasing similarity (Lansing, et al., 1999).

Of equal importance was the finding that, based on their qualitative feedback, the participants from all three language groups found the newly included items to be familiar to them, despite

the obvious difficulty the complexity of the items presented for them. On the basis of this feedback and the recommendations of the panels of experts, the *aloe* and *pylon* were included instead of the *spur* and *husk* items, because in general the participants had found these relatively more familiar. These qualitative observations were supported by the quantitative results which found that language did not have a significant effect on the participants' performance, unlike the previous study, where the isiXhosa speakers had performed worse than the other two groups on the original *mug* and *hippopotamus* items. This finding was confirmed by the standard deviations that were found with the three language group scores, which were all similar (between 3.05 and 3.46), indicating that the groups performed comparably on the modified Naming Test.

With regard to the modified Naming Test, another important finding was that education was demonstrated statistically to affect the participants' performance, the participants with eight or fewer years of education scoring significantly lower. This is consistent with, for example, the findings of Hawkins & Bender (2002) and Saxton, Ratcliff, Munro, Coffey, Becker & Fried (2000), who also found education to be a determinant of Boston Naming Test performance. As noted previously in relation to the other tests, this finding has important implications for neurocognitive testing in the South African context, given the deprivation of education that many South Africans have experienced. The extreme range of education levels seen in the patients seen at Groote Schuur Hospital reminds one that neurocognitive tests need to be robust in terms of their ability to test core neurocognitive functions (Walsh, 1985). Despite this finding, it was determined that the combined effect of the 'level of education' and 'language spoken' did not significantly influence how the participants performed.

The influence of the participants' age on their Naming Test performance was also investigated, the outcome indicating that age did influence performance, the participants aged over 40 fairs significantly worse than those aged below 40. In contrast, previous studies have, however, found age to only significantly influence Boston Naming Test performance from the age of 60 onwards (Ardila, 2007; Lansing et al., 1999; Ross, Lichtenberg & Christensen, 1995; Tsang & Lee, 2003; Welch, Doineau, Johnson & King, 1996). In this study, the participants older than 40 scored significantly lower than those aged below 40. A likely reason for this discrepancy is the fact that one cannot overlook the unique effect of education in the South African context (the other studies were conducted on western samples). Here, many of the current generation of those South Africans who are over 40 years old and who are 'non-whites' suffered under the 'bantu' education system during the Apartheid era and were denied the same quality of education received by the white population. Given this, the finding of the influence of age on this test can be accounted for by the strong relationship in South Africa between age and having experienced a low quality/level of education (Nell, 2000).

The re-piloting of the 3-D Analysis proved complicated, with the initial results revealing how the newly modified version still remained problematic due to its complexity. Despite the initial modifications — re-sizing of items and adjusting the perspective in all the visually distorted items — having been implemented, the control participants were still performing poorly on the test. In addition to this, there was still no statistically significant difference between their performance on this modified test and that on the original 3-D Analysis Test (they had averaged 8.63 out of 14 on the modified test and 8.1 on the original).

Given persistently poor control performances, statistical analyses were undertaken to investigate the possible causes of this problem. Firstly, it was discovered that the participants' level of education had a great effect on their test performance. Again, this finding is consistent with observations in the literature that this test is highly sensitive to level of education (Lezak, 1995). Secondly, with regard to the isiXhosa-speaking participants, an interaction effect was discovered with their level of education. The isiXhosa speakers' 3-D Analysis Test performances were not influenced by whether they had many years of education or just a few. The English participants seemed the most affected by their level of education, those with higher levels performing significantly better on the test than those with lower levels, although the probability value found for this finding was quite low (0.046). Therefore, further research on this finding would be required to investigate it more closely. Finally, the results of the statistical analyses revealed that age did not significantly influence performance on the 3-D Analysis Test.

Given this obvious susceptibility of the 3-D Analysis Test to the influence of education, it was decided to simplify the test further. Firstly, an analysis was performed to determine which of the test's 14 items produced the highest error rates, in order to identify which items to replace. Following from this, it was decided that the first of the additional modifications required would be to replace four of the six most problematic items with new, simpler ones. The neuropsychology experts concurred that the four identified items were indeed the best candidates for replacement. As a result, the final 3-D Analysis Test included in the GSNSB comprised 10 items.

To test the efficacy of the 3-D Analysis Test following the further modifications, additional statistical procedures were run. Here, the findings were far more pleasing: the average score

of the participants improved to 7.03 out of 10, far higher than the average score of 8.63 out of 14 in the original test. It was also found that neither the participants' age nor their first language affected their test performance on the new 10-item test, although their level of education was still a significant predictor.

On reflection, this phase of the research in the context of the goals it set out to achieve can be deemed a success. Specifically, the performances of isiXhosa speakers on the Township Fire Story greatly improved. The Naming Test no longer produced a ceiling effect, and all its items can be deemed as culturally fair as possible. The newly created 10-item 3-D Analysis Test raised the test performances of all three language groups to a level that can be considered 'normal' for individuals without neurocognitive deficits.

The overall goal of this re-piloting study was to provide culturally fair tests for inclusion in the GSNSB ahead of the central validation phase of this research. All three re-piloted tests — based on the quantitative and qualitative findings, as well as the approval of the panel of cultural, language and neuropsychology experts — were shown to be suitable for inclusion. This suitability was based largely on their demonstrated ability to be culturally fair across three language groups, across participants with widely ranging levels of education. The next task of establishing the clinical effectiveness of the new tests within the GSNSB constituted part of the overall 'validation' process. At this junction it must be remembered — as has been highlighted in the Chapter Four and by a number of authors — that the task of addressing the issues of 'culture' in relation to neuropsychological assessment, especially in a multicultural context, is a daunting and largely neglected one (Ardila, 2007; Foxcroft & Roodt, 2005; Mitrushina et al., 2005; Nell, 2000; Salazar, Garcia & Puente, 2007).

CHAPTER SIX: ESTABLISHING THE RELIABILITY AND VALIDITY OF THE GROOTE SCHUUR NEUROCOGNITIVE SCREENING BATTERY

The fourth and central problem to be addressed was the need to validate and test the reliability of the Groote Schuur Neurocognitive Screening Battery (GSNSB). These two objectives were essential for the development of the GSNSB as a clinically effective screening tool in the South African context. To achieve this goal it was necessary to adopt a comprehensive approach to validation — as is used when developing full clinical assessment batteries — even though the GSNSB is intended to be used as a screening tool and not for clinical assessment.

Validation in Relation to Neuropsychological Evaluation

Broadly speaking, the validity of a test or test battery refers to the degree to which it does what it intends — fitting the purpose for which it was designed — and whether performances achieved by the patients using the test/s are predictive of dysfunction of a specific kind due to brain pathology (Durrheim, 1999; Hebben & Milberg, 2002). Other authors such as Retzlaff and Gibertini (1994), and Anastasi and Urbina (1997), have defined validity as “the degree to which the accumulated evidence supports the specific interpretations that the test developers, or users, claim” (Lezak et al., 2004). Even if individual tests display good validity, it does not mean that they can simply be grouped into a battery and used, without first evaluating the validity of that battery. This is because, as soon as more than one test is utilised in combination with others, the overall relationship between the tests results in new possible outcomes, given that the results from the battery are more than just the sum of the tests (Russell, Russell & Hill, 2005). For example, if only one individual test is used to investigate

a particular cognitive function, it is possible that, given that a test's probability of correctly identifying brain damage is never 100 percent, chance could have produced the result.

However, if multiple tests are used to investigate the same cognitive function, the possibility of chance producing the results diminishes (Russell et al., 2005).

Conventionally, given the psychometric underpinnings of virtually all assessment and screening batteries, validity is assessed using statistical means and the generation of standardised normative data. This approach is unproblematic when a fixed battery approach is adopted because the scores are invariant and fixed, allowing for them to be compared (Russell et al., 2005). Here, an array of statistical procedures, and comparisons with established test norms for the tests in use, are used to demonstrate a battery's validity.

An alternative psychometric indication of validity that is often drawn upon — especially where screening batteries are concerned — is the measurement of the battery's sensitivity and specificity, achieved by calculating the percentage of cases the test/s accurately predict as members of a specified group (Mitrushina et al., 2005). The specificity of a test refers “to the probability of correctly identifying a normal individual or an individual from another clinical population intact with respect to the test under consideration (i.e., correct rejection of abnormality)” while the sensitivity of a test refers “to the probability of correctly detecting abnormal function in an impaired individual (i.e., the “hit rate”)” (Lezak et al., 2004, p. 149). This sensitivity/specificity approach is, however, limited, as factors such as population base-rates and variables such as level of education can influence patients' performance on tests, resulting in this measure only being meaningful when used on similar populations (Ostrosky-solis et al., 2000).

When the flexible battery approach is used, the demonstration of validity through statistical means becomes problematic, as there is no stability or consistency with test use over time (Mitrushina et al., 2005). This is the main criticism of the flexible battery approach, frequently made with reference to, for example, the Boston Process Approach. Consequently, in circumstances where a flexible battery approach is used, the demonstration of validity often relies on clinical judgement. Here, the clinician uses his/her clinical judgement to evaluate the accuracy of the tests used. This approach has been criticised as being too subjective and, consequently, no validation studies are available from use of this method (Garb, 1998).

Certain test batteries are designed to test a specific domain of cognitive function or one particular pathology (for example dementia), while others are designed for more than one domain. When this latter type of battery is used, Reitan's rule is often utilised as a way of using combinations of tests to establish a neurocognitive function (Russell et al., 2005). This rule prescribes the use of two or more tests to localise a lesion, based on the principle that one of the tests will be more sensitive to the function under investigation than the other. In other words multiple tests are used, some known to be sensitive to a particular lesion and brain condition, others known not to be, with the ratio between test scores then established to identify the location (Russell et al., 2005).

Teuber (1955, 1975) also "developed the method of double dissociation, which involves a comparison of the effects of at least two tests as applied to the two hemispheres of the brain in order to determine the relationship between tests and the hemispheres" (Russell et al., 2005, p. 789). In conjunction, these two approaches provide an effective means of evaluating the validity of a battery across domains of function. An example of the use of this double

dissociation approach is the Luria Neuropsychological Investigation (LNI), which utilises hypothesis-testing through the decision-trees designed by Luria to investigate syndromes in a four-step approach (Christensen & Caetano, 1999; Hebben & Milberg, 2002).

In summary, regardless of whether a fixed or flexible battery approach is adopted, the common denominator is that the majority of test batteries in use are fundamentally psychometric, on account of their varying use of standardised tests and their reliance (to a greater or lesser extent) on normative data. Hence, even when hypotheses are generated under the flexible approach, the approach still draws on quantified test scores to some extent, rather than qualitative observation of the patient's performance, as achieved under Luria's qualitative approach. Consequently, the most widely accepted methods adopted to validate assessment and screening batteries are imbedded within the psychometric tradition.

Against this background, the GSNSB may be said to comprise aspects of various neuropsychological approaches: psychometric, qualitative, screening and assessment.

Validation of the Groote Schuur Neurocognitive Screening Battery

There are some key issues surrounding the nature of the validity and reliability of clinical neuropsychological tests and testing. "The usual requirements that a 'good' test meet reasonable criteria for validity and reliability, and have appropriate norms are often not easy to satisfy in neuropsychological assessment" (Kaszniak, 1989, as cited in Lezak, 1995, p. 119). This statement reveals how, in reality, many good neuropsychological tests that have the ability to elicit abnormal performances in patients have been developed out of clinical experience and research — rather than having been standardised on big groups (or small ones for that matter) (Lezak, 1995). The difficulties that arise are complex. A 'good' test, sensitive

for visuospatial inattention, for example, would prove reliable and valid if, when given to normal individuals, it elicits no phenomenon at all. Yet, giving the same test to patients with documented visuospatial inattention may elicit the phenomenon in only some of the cases (Fan, Lezak, Yuan & Hu, 1988) — “and if given more than once soon after onset of the pathological condition, might prove highly unreliable as patients’ responses to this kind of test can vary from day to day” (Lezak, 1995, p. 119). Additionally, many ‘good’ tests have little purpose in neuropsychology, despite being able to satisfy the conventional statistical criteria.

The above-mentioned issues are highly relevant and are compounded by another issue: “Not all tests in neuropsychology will meet all validity criteria, and many seem to meet none beyond a very loose ability to differentiate between normal control subjects and patients with significant cognitive deficits (Mapou, 1988). Moreover, validity will vary with the use to which a test is put” (Lezak et al., 2004, p. 108). Apart from the usual validity a test must display (that is, measuring the actual deficits it claims to measure), two types of validity are relevant when examining neuropsychological tests — face validity and predictive validity. Face validity refers to “the quality of appearing to measure what the test is supposed to measure, [and] becomes important when dealing with easily confused or upset patients who are thus more likely to reject tasks that seem nonsensical to them” (Lezak et al., 2004, p. 109). Predictive validity, the ability of the test to predict future, real-life situations related to the construct, is important in neuropsychological tests, but is extremely difficult to accurately achieve (Durrheim, 1999).

As previously mentioned, a central aim of this research was to validate the GSNSB once the necessary adaptations to the tests had been completed, and once the tool had been fully

translated into both Afrikaans and isiXhosa. The tests initially chosen for inclusion into the GSNSB were selected on the grounds that they were simple, straightforward tests, which exhibited good validity and reliability, and were well established in their use as individual tests, able to effectively elicit specific cognitive deficits (qualitatively different cognitive performances) depending on their use. It was therefore important that any changes made to the tests maintained these qualities. The final validity of the changes made would ultimately be demonstrated by investigation of the validity of the GSNSB as a whole, notwithstanding the fact that the test changes made still had to exhibit good face validity, as was demonstrated in the pilot study (see Chapter Four).

Validation in relation to neuropsychological evaluation essentially meant the demonstration that the GSNSB (comprising the adapted tests) was effectively able to distinguish between the presence and absence of pathological conditions in each of its respective domains of cognitive function. Each pathological condition has a characteristic pattern of deficits, depending on the neuro-anatomical area of cognitive function affected (Luria, 1966, 1973; Walsh, 1991; Walsh & Darby, 1999). For example, the Right Hemisphere Syndrome accompanying a right temporo-parietal lesion typically includes the presence of some degree of anosognosia, unilateral neglect and constructional apraxia. 'Good' neurocognitive tests are able to elicit such patterns of deficit, while other areas when tested reveal no deficit. In the presence of a pathological condition, the nature/pattern of the presenting deficit will vary according to the region of the brain affected. Each of these patterns of cognitive deficit is qualitatively different to the neuropsychologist (Walsh, 1991).

The central task of validating the GSNSB essentially involved obtaining scores to denote the presence versus absence of a particular pathological condition. With this research, validation

involved examining and comparing patterns of deficit/or symptom (ultimately denoted by the GSNSB scores) in various lesion groups and distinguishing between the groups on these grounds. Put simply, the method adopted in validating the GSNSB involved the demonstration that patients with known anatomical lesions produced the characteristic test performances (patterns of cognitive deficits) known to accompany these lesions. This process involved taking patients with various known anatomical lesions and assessing them using the entire GSNSB, in order to demonstrate that their test performances (on the tests specifically chosen and adapted for the task) were consistent with the known patterns and qualities of deficits associated with each specific lesion-site.

It is important to remember the significance of the qualitative judgements of the patients' test performances underlying the allocation of the scores. This, in conjunction with the pattern of scores that emerges, is central to demonstrating validity:

An early mistake made by psychologists was to place reliance on composite or summarizing scores ... derived from a variety of subtests measuring a plethora of functions. The result was often to conceal quite specific and important deficits in the process. This could be likened to summarizing scores from all four limbs to derive an index of 'motor power'. A 90 percent power score does not convey the information that the patient has three limbs with normal power together with a left arm paresis. Moreover, a composite score can be made up in endless individual ways. What is needed is a description which communicates both the level of function in separate areas together with the *nature* of any dysfunctions.

(Walsh, 1991, p. 258)

By utilising the qualitative features of the patients' performances in deriving the GSNSB scores — while at the same time using a theory-driven neurocognitive screening battery that

was divided into meaningful and systematic domains of cognitive function — we were able to overcome this concern when validating the GSNSB.

Reliability in Relation to Neuropsychological Evaluation

In general, the reliability of a test or test battery refers to its ability to produce the same results over repeated trials (Durrheim, 1999; Hebben & Milberg, 2002; Mitrushina et al., 2005). Reliability in relation to measures of neurocognitive functions is not easy to achieve, as the clinical presentations and patterns of deficit under observation are in a state of flux, often resolving or declining over time (Bleiberg, Garmoe & Halpern, 1997; Hebben & Milberg, 2002; Mitrushina et al., 2005; Walsh & Darby, 1999). If patients' responses are ambiguous, or fluctuating, inter-rater reliability can be affected — thus, on occasion, it is less than perfect (Mitrushina, et al., 2005).

Test-retest reliability refers specifically to the ability of the test or test battery over time, and is investigated by administering the measure on separate occasions and then producing a correlation of the two scores (Durrheim, 1999; Hebben & Milberg, 2002). Inter-rater reliability refers to the consistency of the test or test battery when administered by two or more examiners and is of vital importance in instances where items require the judgement of an assessor for scoring (Hebben & Milberg, 2002).

Although reliability across different patient groups is a highly sought-after quality in neuropsychological assessment batteries, this is unfortunately not achieved in many batteries currently in use, the Halstead-Reitan Battery, for example, being one of the few that exhibits this quality (Hebben & Milberg, 2002). The demonstration of the reliability of the GSNSB incorporated comparisons across four patient groups and a control group, as well as three

language groups — once again highlighting the stringency of the process adopted by this study.

Rationale

The overriding goal of this research was to provide South Africa with a culturally appropriate, diagnostically meaningful neurocognitive screening tool. To achieve this task, the tool's reliability and validity needed to be formally demonstrated. This final aspect of the overall research focused on testing the reliability and validity of the GSNSB. Thus far in the study, the GSNSB had been successfully translated from English into two South African languages, and had been further enhanced and developed through the introduction of the various newly developed culturally-fair neurocognitive tests.

Based on the findings of both the initial pilot and the re-piloting studies, all nine newly developed neurocognitive tests were incorporated into the GSNSB ahead of its validation. Apart from the extensive translation work (see Chapter Three), this involved a number of adjustments to the instructions, cut-off scores and scoring instructions. For example, the FAS Test cut-off scores in the GSNSB Prototype were adjusted, and additional instructions were added to accommodate the new Naming Test, given that the Boston Naming Test had not been part of the original study (see Chapter Three for details of all the changes made). Once these changes had been proof-read by an editor, the GSNSB was ready to be validated.

The following chapter outlines the process followed in establishing the reliability and validity of the GSNSB. This process involved using patients from four different lesion groups, and a control participant group, in order to determine whether the GSNSB showed consistency over

assessments and whether it was able to differentiate between various lesion sites within the brain and between the presence versus the absence of neurocognitive deficit.

Methodology

Sample

The patients included in this sample were chosen on the basis of two criteria. The primary criterion was that they had an established, localised lesion. The second criterion was that neurocognitive impairment was present. Where possible a random sampling approach was adopted, although this was not always easy to achieve for a number of reasons, including: patients with specific lesions were sought; many neurocognitive impairments resolve quickly over time; and certain disorders, such as the Right Hemisphere Syndrome, are relatively rare in the overall patient population.

In total, 75 participants were assessed to test the GSNSB's reliability and validity. The sample comprised 15 participants for each of the four patient groups, namely: hippocampal lesions, left hemisphere lesions, right hemisphere lesions and frontal lesions. Fifteen neurocognitively intact individuals — first screened using the Screening Sheet (see Appendix H) to exclude possible pathologies involving cognitive sequelae — were also included in this sample. The hippocampal lesion group consisted of patients with either unilateral or bilateral hippocampal lesions. The 15 left hemisphere cases were made up of mainly left middle cerebral artery (MCA) strokes, and some patients with left internal capsule strokes with cortical extension, all localised to the left perisylvian convexity and sparing the frontal cortex. The right hemisphere group comprised 15 right middle cerebral artery (MCA) strokes, again sparing frontal cortex. Finally, the 15 frontal cases consisted of patients with prefrontal lesions (either unilateral or bilateral), although it was extremely difficult to find patients with

isolated prefrontal lesions. The reason for this is that the pathologies presenting with prefrontal lesions are typically traumatic brain injuries (TBIs) or dementias, both of which usually involve varying degrees of diffuse brain damage. The control participants, who were thoroughly screened to ensure the absence of any neurocognitive deficits, were derived from Groote Schuur Hospital patients in order to be demographically similar to the patient sample. To further ensure the absence of neurocognitive impairment, most of these controls were taken from the orthopaedic wards (for example, individuals with minor injuries such as broken limbs). Fifteen of the 75 participants, three from each of the five research groups, were reassessed for reliability purposes, bringing the total number of assessments performed to 90.

The overall sample ranged in age from 17 to 84 years, with an average age of 53.24 ($SD = 17.34$) years. The sample comprised 39 males and 36 females. The education level of the overall sample ranged from no years of education up to 16 years, the average level of education of the sample being 8.52 ($SD = 3.22$) years. Further descriptive statistics for the sample divided by pathology group and controls are provided in Table 6.1.

Table 6.1

Means and standard deviations for Age and Years of Education for Hippocampal-, Left Hemisphere-, Right Hemisphere- and Frontal-lesion groups

Group	Lesion				
	Hippocampal	Left	Right	Frontal	Control
	Age				
<i>M</i>	69.13	54.00	52.53	50.00	40.53
<i>SD</i>	9.25	20.47	12.28	17.06	13.66
	Years of Education				
<i>M</i>	8.67	7.27	8.07	8.80	9.80
<i>SD</i>	3.27	2.49	3.31	2.51	4.13

In the sample, of the 60 patients with lesions, 8 were isiXhosa speakers, 29 were English speakers and 23 were Afrikaans speaking (further demographic information for the lesion groups divided by language is provided in Table 6.2). The sample of 15 controls was chosen to comprise five isiXhosa-, five English- and five Afrikaans-speakers.

Table 6.2

Means and standard deviations for Age and Years of Education for isiXhosa, English and Afrikaans groups

Lesion Group	Language		
	IsiXhosa	English	Afrikaans
	Age		
<i>M</i>	39.88	60.86	56.57
<i>SD</i>	20.68	13.08	16.59
	Years of Education		
<i>M</i>	9.25	8.41	7.57
<i>SD</i>	3.20	2.86	2.86

Materials

The entire newly finalised GSNSB was used in all three of its translated forms (see Appendices B, C and D). This version of the GSNSB, including the Orientation, Memory Function, Language Function, Spatial Cognition and Executive Function sections, was administered to all participants. The Screening Sheet (see Appendix H), used to screen the controls in the pilot studies for possible pathologies involving neurocognitive impairments, was again utilised to screen the 15 neurocognitively intact controls. The Consent Form (see Appendix I) and the Patient Information Sheet (see Appendix J) were used to ensure that all

participants were properly informed about the study's intentions and purpose, and that their participation was voluntary. Finally the Scoring Sheet (see Appendix K), previously used in the pilot studies, was again utilised to accurately record all the patients' test responses, along with any of the assessors' qualitative observations concerning the patient and his/her test performance.

Design

This phase of the study was designed to allow the evaluation of both the reliability and validity of the GSNSB. A single-blind study, drawing on converging lines of evidence — both qualitative and quantitative data — was designed in order to avoid possible bias in the scoring and administration of the GSNSB to the patients/controls. Two assessors were each given responsibility for two of the GSNSB's four sections when it came to the qualitative allocation of the data (see below) — one being responsible for the Language Function and Spatial Cognition sections, and the other for the Memory Function and Executive Function sections.

The task of demonstrating the GSNSB's validity was divided into two distinct levels of analysis, drawing on the double dissociation approach, and on Reitan's rule (Russell et al., 2005; Teuber 1955, 1975). The first, primary level, involved demonstrating the GSNSB's core validity — its ability to distinguish left hemisphere lesions from right hemisphere lesions and from controls, and frontal lesions from hippocampal lesions and from controls. This combination was seen as clinically important as, in the case of amnesia for example, it is important to be able to distinguish between a primary 'encoding' (hippocampal involvement) versus a secondary 'retrieval' problem (executive dysfunction). This distinction was therefore highlighted as key to the validation process.

A broader reason for these particular divisions was to demonstrate the GSNSB's ability to identify and differentiate between syndromes —meaningful combinations of deficit that point to a focal area of brain damage — through its theory-driven design, rather than only testing whether it could detect the general presence of deficit ('organicity'), without being able to identify the nature or location of the deficit. Traditionally, screening tools have primarily been designed to satisfy this latter criterion, by looking at 'organicity' — that is, treating brain damage as a unitary phenomenon. The ability to differentiate between syndromes provides a far more detailed and accurate multidimensional appraisal of a patient's neurocognitive state than the one-dimensional screening 'organicity' provides.

The next, secondary level of analysis had two objectives. The first of these involved investigating other combinations of the GSNSB's sections, examining the GSNSB's ability to distinguish hippocampal lesions from left hemisphere, right hemisphere and frontal lesions; left hemisphere lesions from hippocampal, right hemisphere and frontal lesions; right hemisphere lesions from hippocampal, left hemisphere and frontal lesions; and finally, frontal lesions from left hemisphere, hippocampal and right hemisphere lesions. Although these secondary combinations were of clinical relevance in demonstrating the GSNSB's screening abilities, given the nature of neurocognitive syndromes and how they co-exist clinically, they were less crucial than the primary discriminations, which are more closely linked theoretically.

The other objective of this secondary level of analysis was to investigate whether the participants' performance on the 'Orientation' section of the GSNSB covaried with their performances on any of the battery's other four sections. This was necessary to determine

whether ‘Orientation’ performance was in any way a positive or negative predictor of performance on one of these sections. The ‘Orientation’ section was included as it provides a basic evaluation of the patient’s lucidity, conscious arousal and awareness, allowing the assessor to decide whether continuing with the neurocognitive assessment is worthwhile, or whether the patient is too confused or disoriented to be able to cope with being tested further.

For this validation study, the author had sole responsibility for finding and screening the suitable patients and controls within the hospital. Once these participants had been identified, they were assigned randomly to either one of the two research assistants (both neuropsychology masters students) for assessment. The assessors, who had been trained along with the team of interpreters in the administration of the GSNSB, were ‘blind’ as to the type of participant — patient or control — they had been assigned (that is, they were not aware whether the person they were assessing was a control or a patient with either a hippocampal lesion, a left or right hemisphere lesion, or a frontal lesion).

The design of the investigation of the GSNSB’s reliability involved the reassessment of 15 of the 75 participants — three randomly chosen from each of the five research groups. These reassessments were assigned to the assessor who had not conducted the first assessment in order to allow the investigation of inter-rater reliability concurrently with test-retest reliability. Inter-rater reliability was also established qualitatively when it came to the qualitative allocations to the participant groups, carried out by the two assessors. Here, the errors made in allocation by the two assessors were compared.

Following the data collection process, each of the assessors was assigned the data for which he/she was responsible (that is, two GSNSB sections each, along with the control data, which

was shared between them). They were then asked to ‘blindly’ allocate each of their 45 scoring sheets to one of their three participant groups (the data from the 15 reassessed participants was not included for this specific aspect of the design). This allowed for the qualitative investigation of the GSNSB’s ability to differentiate hippocampal from frontal lesions and from controls, and left hemisphere from right hemisphere lesions and from controls. The allocation was made purely on a qualitative basis, using observations made regarding the participants’ performances on each test and the patterns of deficit they had noted when analysing each GSNSB section — the overall score breakdown and summary of the GSNSB were not consulted. In other words, the assessors had to use the decision-trees to qualitatively evaluate the participants’ performances, without examining how they had performed quantitatively. Following from these qualitative investigations, statistical procedures were run in order to investigate the data from a quantitative perspective too.

As the use of converging lines of evidence had proved successful throughout this research, it was therefore deemed important to continue with this approach. Because the design of the GSNSB incorporated both qualitative aspects (with its decision-tree approach and information regarding clinical phenomena) and quantitative aspects (with its cut-off scores and scoring procedures), it was advantageous to design a validation study incorporating both these approaches.

Data Analysis

The collection of qualitative data was of particular importance in informing the effectiveness of the decision-trees and for guiding further improvements to the GSNSB. This is because key insights and observations can be gained from analysing how and why a patient failed on a particular test item, instead of focusing only on the fact that they had failed. Added to this, it

was vital to examine the multiple possible determinants of failure for the participants' failed items, in order to further develop effective decision-trees that take such factors into account.

For the qualitative analysis of the validation data, the allocation approach described above was adopted. Once the allocations had been completed by both assessors, the number of errors made was recorded, along with the information regarding which type of participant had been misallocated to which group. This process also involved looking at the qualitative factors that helped in the successful allocations, and at those that contributed to the errors made in each specific case.

In conjunction with the qualitative analyses, quantitative measures were also adopted. Here, analysis of the GSNSB's validity was broken down into both primary and secondary analyses. The primary, theoretically central demonstration of validity involved first using the one-way ANOVA statistical procedure to investigate whether the GSNSB's respective sections were able to differentiate patients with left hemisphere lesions from patients with right hemisphere lesions, and from controls, and differentiate patients with hippocampal lesions from patients with frontal lesions, and from controls. At the secondary level of analysis, the other combinations of sections of the GSNSB (as mentioned above) were also investigated using one-way ANOVA.

Also central to the primary analysis of validity was the importance of analysing the possible influence of the key variables that had been encountered during the development of the neurocognitive test. This involved analysing the influence of the variables of age, first language and level of education on the participants' performances on the GSNSB. The factorial ANOVA statistical procedure was used to explore the possible influence of these

crucial factors, while the Tukey's HSD (Honestly Significant Difference) procedure was employed to see where the significant differences lay. Factorial ANOVA was also used to investigate whether the sections of the GSNSB and the individual tests contained therein were able to differentiate normal controls from participants with established lesions. In order to do this, the total scores from the four sections and the individual test scores were used. Chi-squared tables were used for the test scores of the GSNSB that constituted categorical data (for example, in instances where test performance was reduced to a score of '0', '1' or '2').

For the secondary level investigation of whether performance on the 'Orientation' section of the GSNSB covaried in any way with performances on any of the GSNSB's other four core sections, a series of Pearson's product moment correlation coefficients (r) was made. For each investigation, a scatterplot was first derived in order to inspect whether there was at least a rough relationship between the two variables, thereby indicating whether or not it was worthwhile to proceed with the remaining analysis.

In order to investigate the reliability of the GSNSB, Pearson's product moment correlation coefficients (r) were calculated to compare the participants' initial assessment with their second assessment, for each of the five sections of the GSNSB, as well as for the total GSNSB score. Correlations were made between the individual scores in each section, as well as between the total scores for each of the five sections. To accurately determine the correlations, the individual scores of the participants were first normalised by converting them into percentages. These summed individual scores were then correlated with the reassessment scores.

Procedure

Before each assessment was begun, the participants were given the Patient Information Sheet and Consent Form to sign, and reminded that they were free to withdraw from the study at any point should they not wish to continue. All data collected was securely stored and only made available to the researchers. Each testing session lasted between an hour and an hour-and-a-half, during which time the entire GSNSB was administered. Each assessment was conducted in the participant's first language — that is, in the case of the isiXhosa and Afrikaans participants, the respective interpreters conducted the assessment in the presence of one of the two assessors. The interpreters, along with the assessors, were all 'blind' as to which participant — patient or control — was being tested.

Along with the participants' test responses, qualitative observations regarding the participants, their performances on the tests, and the testing session, were made throughout the assessment (for example, if there was an interruption during one of the tests, or if the participant had a headache or a paralysed limb, etc.). All the test responses, along with the qualitative observations made, were recorded on the Scoring Sheet. The 15 reassessments for the reliability study were performed within two days of the initial assessment, in order to ensure that the patients' neurocognitive deficits had not resolved and that the patients were not lost to the study once they were discharged from the hospital. Once the data had been collected, the two assessors conducted the qualitative allocations.

Results

Evaluating the Groote Schuur Neurocognitive Screening Battery's Reliability

Of the 15 participants randomly chosen for reassessment, six were male and nine female; six spoke English as a first language, six were Afrikaans first-language speakers and three were

isiXhosa first-language speakers (see Table 6.3 for further demographic characteristics of re-tested participants).

Table 6.3

Means and standard deviations of Age and Years of Education for reliability study participants (N=15, Male=6, Female=9)

	Age
<i>M</i>	55.53
<i>SD</i>	18.40
	Years of Education
<i>M</i>	7.87
<i>SD</i>	3.99

Table 6.4 displays the results from the Pearson's product moment correlation coefficients (*r*) calculated for all of the five sections of the GSNSB to investigate its reliability.

Table 6.4

Reliability Correlation Coefficient for Orientation, Memory, Language, Spatial Cognition and Executive Function sections

Section	Correlation Coefficient
<i>Total Battery (all five sections)</i>	0.99
<i>Orientation Function section</i>	0.91
<i>Memory Function section</i>	0.98
<i>Language Function section</i>	0.99
<i>Spatial Cognition section</i>	0.98
<i>Executive Function section</i>	0.96

Primary Evaluation of the Groote Schuur Neurocognitive Screening Battery's Validity

Qualitative findings

Left hemisphere versus right hemisphere lesions

For the investigation of the GSNSB's ability to distinguish left hemisphere lesions from right hemisphere lesions, as well as from neurocognitively intact controls, the qualitative findings showed that, of the 90 allocations done (45 by each of the assessors), only one participant was misallocated. Here, a right hemisphere patient was mistaken for a left hemisphere patient. This represented a 100 percent accuracy rate for the first rater, and a 97.78 percent accuracy rate for the second.

Hippocampal versus frontal lesions

In the qualitative investigation of the GSNSB's ability to distinguish frontal versus hippocampal lesions versus controls, the findings revealed that of the 90 qualitative allocations made, 11 participants were misallocated, five by the one assessor and six by the second, representing an 87.78 percent accuracy rate overall. This represented an 88.89 percent accuracy rate for the first rater, and an 86.67 percent accuracy rate for the second. Of these 11 errors, one control participant and six hippocampal lesion patients were mistaken for frontal lesion patients, and four frontal lesion patients were mistaken for hippocampal lesions.

Therefore, in total, out of the 180 allocations made over the five sections of the GSNSB, 12 errors were made, constituting a relatively good accuracy rate (93.34 percent) in terms of the GSNSB's ability to differentiate, on qualitative grounds, frontal from hippocampal lesions, left hemisphere from right hemisphere lesions, and neurocognitively intact individuals from patients with four types of lesion. Although not perfect, it must be remembered this battery is designed for screening purposes and therefore the level of stringency required is naturally less than in the case of clinical assessment, where errors can potentially have far more severe implications. Screening serves as the starting point of a clinical investigation, rather than the diagnostic conclusion.

Quantitative findings

Memory Function section

To ascertain whether the Memory Function section of the GSNSB could differentiate between frontal lesions, hippocampal lesions and neurocognitively intact controls, one-way ANOVA was used (see Table 6.5 for descriptive statistics). ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p =$

0.003). Given that ANOVA is a robust technique and the sample sizes were equal, the analysis was continued. A significant main effect, $F(2, 42) = 39.50$ ($p < 0.001$) was found. Here, effect size was evaluated by calculating *eta*-squared, which was 0.65, thereby demonstrating that the group the participants were from (frontal lesion, hippocampal lesion and control) accounted for 65 percent of the variation in Memory Function section performances.

Table 6.5

Mean scores for Memory Function section

		Group		
		Hippocampal	Frontal	Control
Score for Memory Section	<i>M</i>	5.87	13.27	17.27
	<i>SD</i>	4.32	4.35	0.80
	<i>N</i>	15	15	15

A post-hoc test (Tukey's HSD) was performed to identify where the difference occurred. When compared, a large significant difference ($p < 0.001$) was found between participants with frontal lesions and those with hippocampal lesions. Findings also showed that a large significant difference ($p < 0.001$) was present between the control group and the hippocampal lesion patient group. Finally, when examined, the comparison between the control group and the frontal lesion group showed a significant difference ($p = 0.010$). Thus, there was a significant difference in performance on the Memory Function section by the hippocampal lesion group when compared to the performances of the frontal lesion group and the controls (Figure 6.1 displays how the different groups performed on the Memory Function Section).

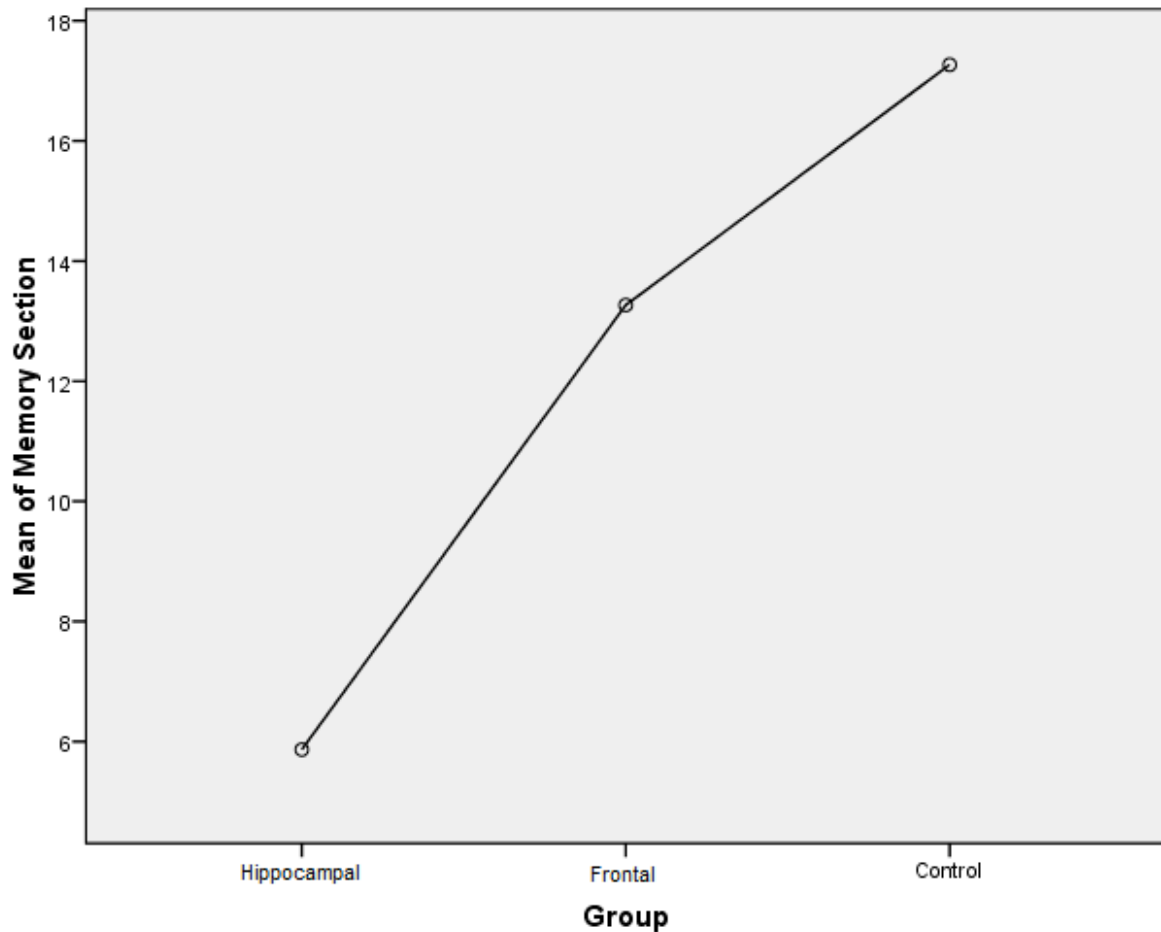


Figure 6.1. Cell mean plot showing total Memory section performances by group

In order to investigate which variables other than lesion-site might have influenced the total score on the Memory Function section, factorial ANOVA was used (see Table 6.6 for descriptive statistics). ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated. However, because ANOVA is a robust technique, and the sample sizes were equal, the analysis was continued. No significant effects for either education, $F(1, 39) = 3.873, p = 0.560$, or language, $F(2, 39) = 0.904, p = 0.413$, were found. However, a significant interaction effect between language and education, $F(2, 39) = 6.584, p = 0.003$, was established. Therefore, neither first language spoken nor level of education significantly affected the participants' performances on the Memory Function section.

Table 6.6

Mean scores for Memory Function section

Factor A: Education		Factor B: Language			A marginal means
		English	Afrikaans	isiXhosa	
9 years or more	<i>M</i>	10.45	15.00	0	8.48
	<i>SD</i>	6.170	3.266	0	
	<i>N</i>	11	7	1	
less than 9 years	<i>M</i>	13.00	9.60	16.14	12.913
	<i>SD</i>	6.423	6.096	1.345	
	<i>N</i>	9	10	7	
B marginal means		11.725	12.3	8.07	

In order to ascertain if there was a possible interaction between lesion-site and age for the total Memory Function section score, a factorial ANOVA was again performed (see Table 6.7 for descriptive statistics). ANOVA's assumption of homogeneity of variance was violated; however, because the assumption of normality was upheld, the analysis was continued because the size of the lesion groups was equal, as were the age groups. Given that the assumption of homogeneity was violated, participants were divided into sub-groups as a requirement so that the ANOVA procedure could continue. For these groups, the participants were separated into those who were aged 53 or younger and those who were older than 53. The cut-off age of 53 was chosen as this was the age that divided the group the most equally — 21 and 24 participants respectively. A significant main effect was demonstrated for pathology type, $F(2, 39) = 10.86, p < 0.001$, but not for age, $F(1, 39) = 0.147, p = 0.704$.

Additionally, no significant interaction effect, $F(2, 39) = 1.356, p = 0.270$, was seen. In summary, lesion group significantly affected performance on the Memory Function section performance. The participants' age, however, did not significantly affect performance on this particular section. Finally, no significant interaction was found between pathology group and age.

Table 6.7

Mean scores for Memory Function section

Factor A: Age		Factor B: Pathology			A marginal means
		Hippocampal	Frontal	Control	
older than 53 years	<i>M</i>	5.57	14.67	17.5	14.76
	<i>SD</i>	4.327	1.751	.577	
	<i>N</i>	14	6	4	
53 years or younger	<i>M</i>	10	12.33	17.18	9.83
	<i>SD</i>	.	5.362	.874	
	<i>N</i>	1	9	11	
B marginal means		5.87	13.21	17.27	

Township Fire Story

To see if the Township Fire Story was able to distinguish a hippocampal lesion from a frontal lesion and from a normal control, one-way ANOVA was used (see Table 6.8 for descriptive statistics). The assumptions of ANOVA were followed, but neither the assumption of normality nor the assumption of homogeneity of variance were upheld. However, the analysis proceeded as sample sizes were equal and ANOVA is a robust technique. For lesion type, a

significant effect was demonstrated, $F(2, 42) = 26.7, p < 0.001$, indicating that lesion-site was a key determinant of score on the Township Fire Story. To investigate where the differences between the two lesion groups and the controls lay, a Tukey's HSD was performed. Here, it was found that a significant difference ($p = 0.015$) existed between the hippocampal lesion group and the frontal lesion group. A significant difference ($p < 0.001$) was also found between the control group and the hippocampal lesion group. Finally, a significant difference was also demonstrated between the frontal lesion group and the controls ($p < 0.001$). A calculated effect size of 0.56 indicated that 56 percent of the variation in the participants' performance on the Township Fire Story could be accounted for by the group they were from.

Table 6.8

Mean scores for Township Fire Story

		Group		
		Hippocampal	Frontal	Control
Score for Township Fire Story	<i>M</i>	8.47	19.47	35.8
	<i>SD</i>	8.81	10.83	11.13
	<i>N</i>	15	15	15

To investigate the effect of the participants' first language and level of education on the Township Fire Story, a factorial ANOVA was performed (See Table 6.9 for descriptive statistics). ANOVA's assumption of homogeneity of variance was upheld, but the assumption of normality was violated. However, because ANOVA is a robust technique and the sample sizes were equal, the analysis continued. No significant main effect was found for education level, $F(1, 39) = 0.280, p = 0.600$, and none for language, $F(2, 39) = 0.399, p = 0.674$.

Additionally, no significant interaction effect, $F(2, 39) = 2.617, p = 0.086$, was demonstrated between these two variables. These results indicated that neither the participants' level of education nor the language they spoke significantly affected how well they fared on the Township Fire Story.

Table 6.9

Mean scores for Township Fire Story (TFS)

Factor A: Education		Factor B: Language			A marginal means
		English	Afrikaans	isiXhosa	
9 years or more	<i>M</i>	22.00	28.43	1	23.26
	<i>SD</i>	16.935	15.447	.	
	<i>N</i>	11	7	1	
less than 9 years	<i>M</i>	19.89	14.80	26.71	19.77
	<i>SD</i>	16.937	12.770	10.468	
	<i>N</i>	9	10	7	
B marginal means		21.05	20.41	23.50	

Auditory Span Test

To ascertain whether the Auditory Span Test could distinguish between frontal lesions, hippocampal lesions and controls, one-way ANOVA was used (see Table 6.10 for descriptive statistics). A significant main effect for pathology type, $F(2, 42) = 5.232, p = 0.009$, was established, indicating that pathology strongly influenced performance on the Auditory Span Test.

Table 6.10

Mean scores for Auditory Span Test

		Group		
		Hippocampal	Frontal	Control
Score for Auditory Span Test	<i>M</i>	5.13	4.53	6.00
	<i>SD</i>	1.13	1.30	1.31
	<i>N</i>	15	15	15

To investigate where the differences between the two lesion groups and the controls lay, a Tukey's HSD was performed. Here, it was found that no significant difference ($p = 0.394$) existed between the frontal lesion group and the hippocampal lesion group. No significant difference ($p = 0.151$) was found between the control group and the hippocampal lesion group. Finally, a significant difference was demonstrated between the frontal lesion group and the controls ($p = 0.007$). The cell mean plot in Figure 6.2 represents the mean differences of the performance of the three groups on the Auditory Span Test. A calculated effect size of 0.2 indicated that 20 percent of the variation seen in the participants' Auditory Span Test performances was accounted for by the group they were from.

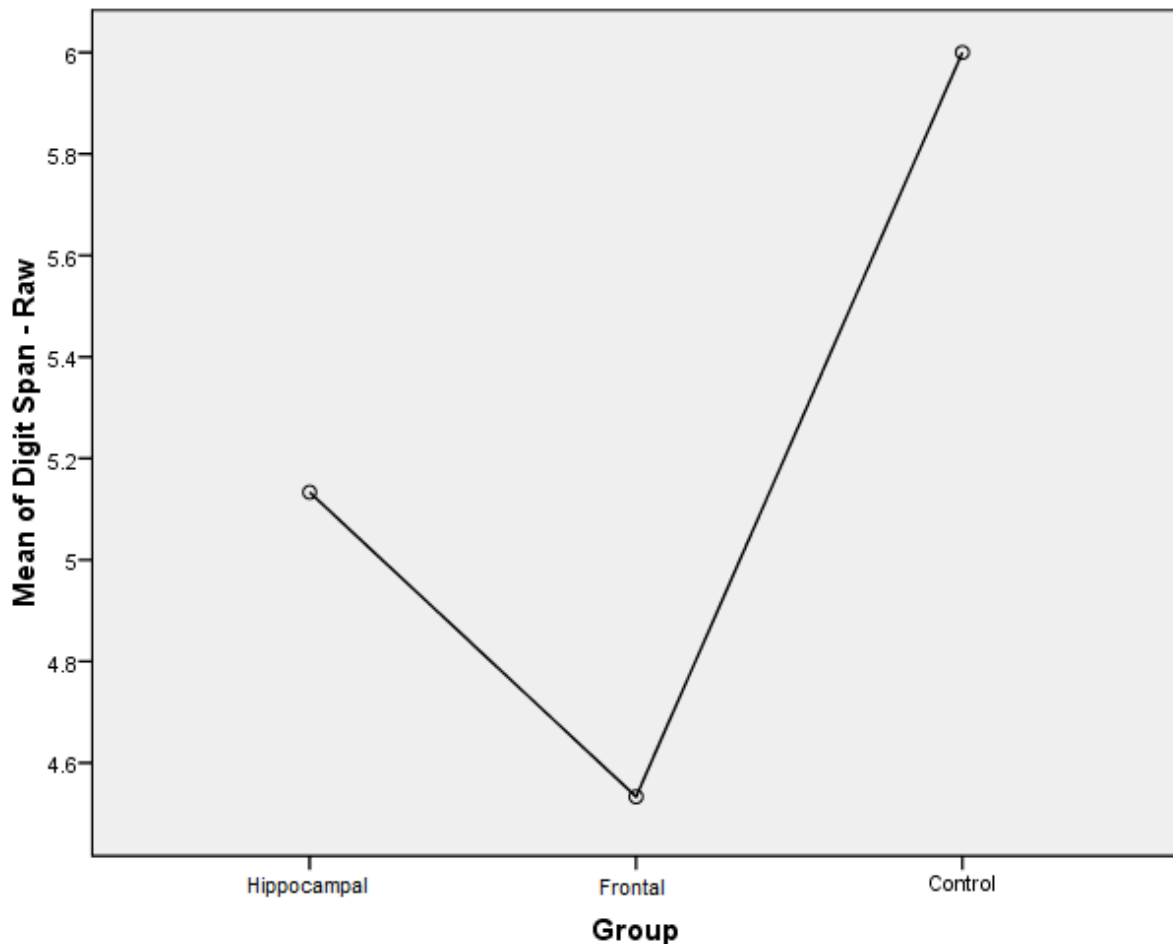


Figure 6.2. Cell Mean plot showing total Digit Span performances by group

Language Function section

To ascertain whether the Language Function section of the GSNSB could differentiate between left hemisphere and right hemisphere lesions, as well as between controls, one-way ANOVA was used (see Table 6.11 for descriptive statistics). ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p < 0.001$); given that ANOVA is a robust tool and the sample sizes were equal, the analysis was continued. A significant main effect, $F(2, 42) = 85.74, p < 0.001$ was found. Here, effect size was evaluated by calculating *eta*-squared, which was 0.8035, thereby demonstrating that participant group (left hemisphere lesion, right hemisphere lesion and control) accounted for 80.35 percent of the variation in Language Function section performances.

Table 6.11

Mean score for Language Function section

		Pathology Group		
		Left Hemisphere lesion	Right Hemisphere lesion	Control
Language section score	<i>M</i>	10.20	25.87	29.07
	<i>SD</i>	6.44	3.07	1.62
	<i>N</i>	15	15	15

A post-hoc test (Tukey's HSD) was performed to identify where the difference occurred. When compared, a large significant difference ($p < 0.001$) was found between participants with left and right hemisphere lesions. Findings also showed that a large significant difference ($p < 0.001$) was present between the control group and the left hemisphere lesion group. Finally, when examined, the comparison between the control group and the right hemisphere lesion group was not significant ($p = 0.107$). In other words, there was a significant difference in performance on this section by the left hemisphere lesion group in comparison to the right hemisphere lesion group and the controls. Figure 6.3 shows total Language Function section performance by group.

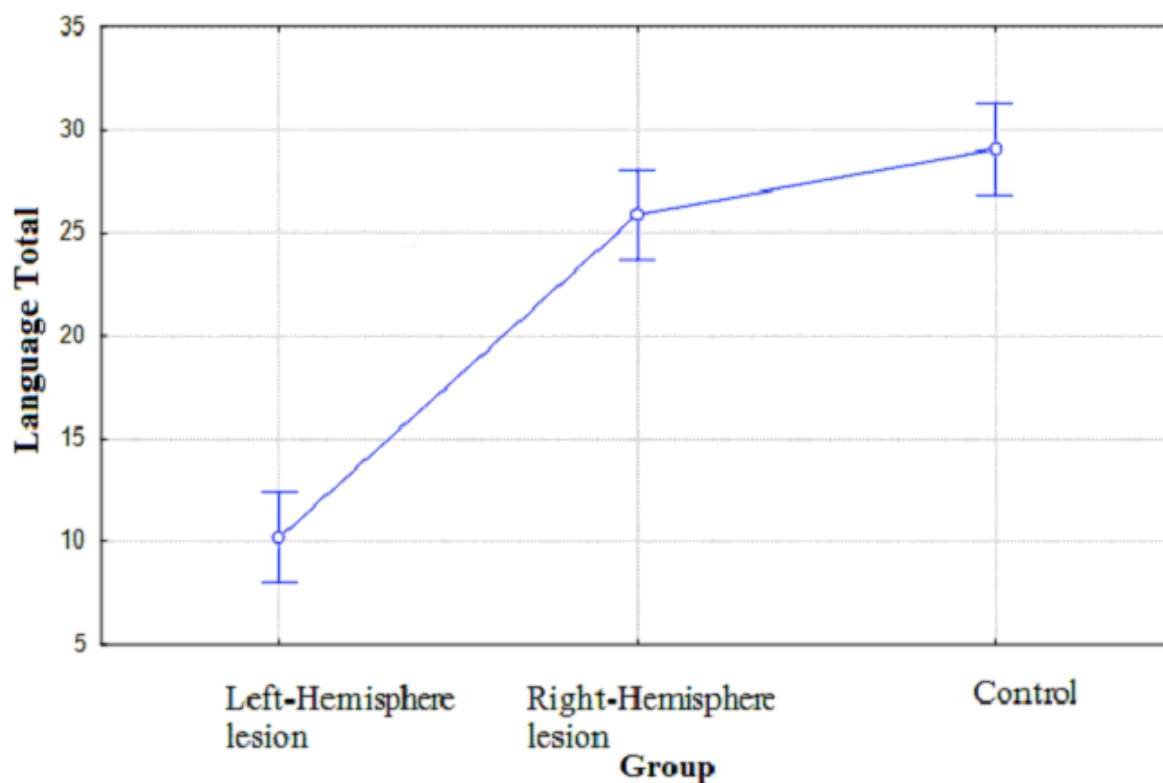


Figure 6.3. Cell mean plot showing total Language Function section performances by group

In order to investigate whether variables other than lesion-site might have influenced the total score on the Language Function section, factorial ANOVA was used (see Table 6.12 for descriptive statistics). Assumptions of normality and homogeneity of variance were upheld. No significant main effect was found for either education, $F(1, 39) = 2.11$, $p = 0.154$, or for language, $F(2, 39) = 0.01$, $p = 0.988$. Additionally, no significant interaction effect, $F(2, 39) = 0.94$, $p = 0.399$, was demonstrated between education and first language. These results showed that neither of these variables significantly affected the total Language Function section performances.

Table 6.12

Mean score for Language Function section

Factor A: Education	Factor B: Language			A marginal means
	Afrikaans	English	isiXhosa	
8 years or more	<i>M</i>	27.33	23.00	24.32
	<i>SD</i>	4.63	10.32	
	<i>N</i>	6	9	
less than 8 years	<i>M</i>	17.50	20.90	19.81
	<i>SD</i>	9.90	9.36	
	<i>N</i>	10	10	
B marginal means		21.19	21.89	22.20

In order to ascertain if there was a possible interaction between lesion-site and age for the total Language Function section score, a factorial ANOVA was again performed (see Table 6.13 for descriptive statistics). ANOVA's assumption of homogeneity of variance was violated; however, because the assumption of normality was upheld, the analysis was continued because the size of the lesion groups was equal, as were the age groups. Given that the assumption of homogeneity was violated, participants were divided into sub-groups as a requirement so that the ANOVA procedure could continue. For these groups, the participants were separated into those who were aged less than 50 and those who were 50 or older. The cut-off age of 50 was chosen as this was the age that divided the group the most equally. Although a significant main effect was demonstrated for pathology type, $F(2, 39) = 83.52, p < 0.001$, there was none for age, $F(1, 39) = 2.87, p = 0.098$. Additionally, no significant

interaction effect, $F(2, 39) = 0.54, p = 0.587$, was seen. In summary, lesion group significantly affected performance on the Language Function section performance. The participants' age, however, did not significantly affect performance on this particular section. Finally, no significant interaction was found between pathology group and age.

Table 6.13

Mean score for Language Function section

Factor A: Age Group	Factor B: Pathology Group			A marginal means	
	Left Hemisphere	Right-Hemisphere	Control		
50 years or older	<i>M</i>	8.38	25.11	28.60	19.82
	<i>SD</i>	6.80	3.69	2.61	
	<i>N</i>	8	9	5	
younger than 50 years	<i>M</i>	12.29	27.00	29.30	23.52
	<i>SD</i>	5.77	1.41	0.95	
	<i>N</i>	7	6	10	
B marginal means		10.20	25.87	29.07	

Naming Test

To see if the Naming Test was able to distinguish a left hemisphere lesion from a right hemisphere lesion and from a normal control, one-way ANOVA was used (see Table 6.14 for descriptive statistics). The assumptions of ANOVA were followed and normality was upheld, but homogeneity of variance was violated; however, the analysis proceeded, as sample sizes were equal and ANOVA is a robust technique. For lesion type, a significant effect was

demonstrated, $F(2, 42) = 52.68, p < 0.001$, indicating that lesion-site was a key determinant of Naming Test score. To investigate where the differences between the two lesion groups and the controls lay, a Tukey's HSD was performed. It was found that a significant difference ($p < 0.001$) existed between the left hemisphere lesion group and the right hemisphere lesion group. A significant difference ($p < 0.001$) was also found between the control group and the left hemisphere lesion group. An effect size of 0.7419 indicated that 74.19 percent of the variation in the participants' Naming Test performance could be accounted for by the group they were from.

Table 6.14

Mean score for Naming Test

Factor A: Age Group	Factor B: Pathology Group			A marginal means
	Left-Hemisphere	Right-Hemisphere	Control	
50 years or older	<i>M</i> 5.38	22.00	25.80	16.82
	<i>SD</i> 6.72	2.92	3.27	
	<i>N</i> 8	9	5	
younger than 50 years	<i>M</i> 11.29	23.67	27.30	21.48
	<i>SD</i> 8.90	3.08	2.41	
	<i>N</i> 7	6	10	
B marginal means	8.13	22.67	26.80	

In order to investigate whether the variables of age and lesion-site, and their possible interaction, affected the Naming Test scores, factorial ANOVA was used (see Table 6.14 for

descriptive statistics). As before, to allow for this procedure to continue, the age variable was broken down into two groups — participants aged below 50 and participants who were 50 or older. ANOVA's assumption of homogeneity was violated, but the assumption of normality was upheld. On analysis of lesion-site, a significant main effect, $F(2, 39) = 50.25, p < 0.001$, was found. However, with age, no significant effect, $F(1, 39) = 1.57, p = 0.218$, was demonstrated. Likewise, no significant interaction effect, $F(2, 39) = 1.57, p = 0.221$, was found between lesion-site and age.

To investigate the affect of the participants' first language and level of education on the Naming Test, another factorial ANOVA was performed (See Table 6.15 for descriptive statistics). ANOVA's assumptions of homogeneity of variance and normality were upheld. A significant main effect was found for level of education, $F(1, 39) = 4.11, p = 0.049$. No significant effect was found for language, $F(2, 39) = 0.02, p = 0.985$. Additionally, no significant interaction effect, $F(2, 39) = 0.58, p = 0.564$ was demonstrated between these two variables. These results indicated that only the participants' level of education significantly affected how well they faired on the Naming Test. An effect of 0.1228 indicated that 12.28 percent of the variation in the Naming Test performances could be explained by the participants' level of education.

Table 6.15

Mean score for Naming Test

Factor A: Education	Factor B: Language			A marginal means	
	Afrikaans	English	isiXhosa		
less than 8 years	<i>M</i>	15.00	18.40	16.67	16.69
	<i>SD</i>	9.15	9.80	11.66	
	<i>N</i>	10	10	6	
8 years or more	<i>M</i>	25.00	21.33	22.00	22.63
	<i>SD</i>	4.10	11.88	2.94	
	<i>N</i>	6	9	4	
B marginal means		18.75	19.79	18.80	

Spatial Cognition section

To ascertain whether the Spatial Cognition section of the GSNSB could differentiate between right hemisphere and left hemisphere lesions, and between controls, one-way ANOVA was used (see Table 6.16 for descriptive statistics). ANOVA's assumption of normality was upheld, as was the assumption of homogeneity of variance. A significant effect was demonstrated for lesion type, $F(2, 42) = 29.78, p < 0.001$, indicating that there was a large discrepancy between how these three groups performed on this section. An effect size of 0.7658 was established — in other words, 76.58 percent of the variation in the participants' overall performances on this section could be explained by the group they were from.

Table 6.16

Mean score for Spatial Cognition section

		Pathology Group		
		Left- Hemisphere	Right-Hemisphere	Control
Spatial Cognition section score	<i>M</i>	7.87	5.13	13.00
	<i>SD</i>	4.00	2.33	1.65
	<i>N</i>	15	15	15

It was also found, when a further Tukey's HSD analysis was performed, that significant differences were present in terms of performance between the left hemisphere lesion and right hemisphere lesion groups ($p = 0.031$), and between the control and right hemisphere lesion group ($p < 0.001$). Finally, a significant difference was also found between how the control group and the left hemisphere lesion group performed ($p < 0.001$). Figure 6.4 shows total Spatial Cognition section performance by group.

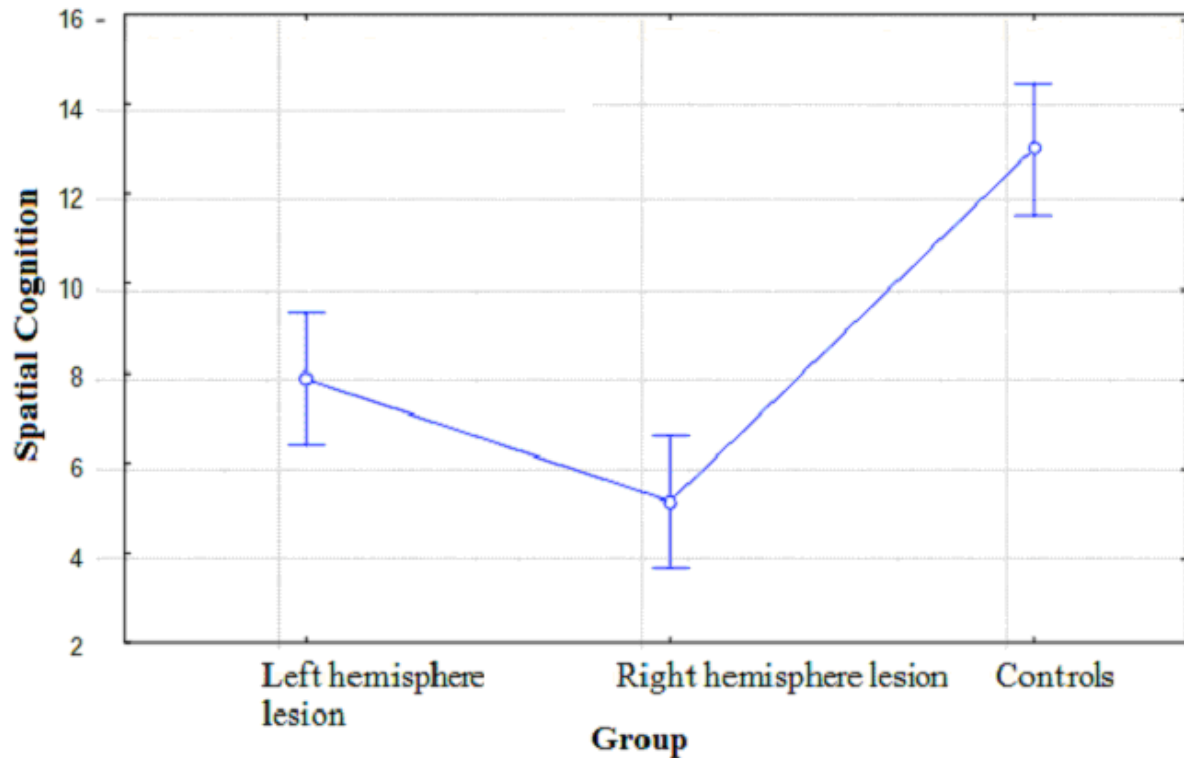


Figure 6.4. Cell mean plot showing total Spatial Cognition section performances by group

In order to investigate if the variables of the participants' level of education and first language affected their scores on the Spatial Cognition section, factorial ANOVA was run (see Table 6.17 for descriptive statistics), with the assumptions of normality and homogeneity of variance being upheld. The results showed that a significant main effect was established for education level, $F(1, 39) = 4.55, p = 0.039$, but not for language, $F(2, 39) = 0.69, p = 0.506$. Finally, no significant interaction effect, $F(2, 39) = 2.10, p = 0.136$, was found between level of education and first language. Therefore, these results showed that the participants' level of education significantly affected how they performed on this overall section. A calculated effect size of 0.2599 meant that 25.99 percent of the variation in the participants' performances on this section could be explained by their level of education.

Table 6.17

Mean score for Spatial Cognition section

Factor A: Education	Factor B: Language			A marginal means
	Afrikaans	English	isiXhosa	
less than 8 years	<i>M</i> 7.00	5.70	10.50	7.30
	<i>SD</i> 4.27	4.37	1.64	
	<i>N</i> 10	10	6	
8 years or more	<i>M</i> 11.33	10.56	9.25	10.53
	<i>SD</i> 3.98	4.07	3.77	
	<i>N</i> 6	9	4	
B marginal means	8.63	8.00	10.00	

Spatial Acalculia Test

To see whether the participants from the right hemisphere lesion, left hemisphere lesion and control groups performed differently on the test for Spatial Acalculia, a chi-squared test was used. A chi-squared statistic of $\chi^2 = (2, N = 45) = 18.41, p < 0.001$, was found, showing that the group from which participants came did indeed influence significantly how they performed on this test. Here, only one right hemisphere patient and one left hemisphere patient passed the test, while ten of the controls passed. To explore this finding in more detail, a chi-square test of independence was used, with the purpose of examining the participants' performance on the test in relation to group. Here, a significant relationship

between these variables, $\chi^2 (1, N = 45) = 1.25, p = 0.264$, was established, indicating that the left and right hemisphere lesion participants were less likely to score higher on the Spatial Acalculia test than the controls.

Focusing on the controls, who had performed by far the best of the three groups on the test, a chi-squared test was used to see how their level of education affected their performance on the test. A chi-squared statistic of $\chi^2 = (2, N = 45) = 1.25, p = 0.264$ was found, indicating that the control group's level of education did not influence how they fared on the test. The first languages of the controls were also examined to establish a possible influence on test performance, again using the chi-squared test. A chi-squared statistic of $\chi^2 = (2, N = 45) = 2.4, p = 0.301$, was found, indicating that first language did not affect how they fared on this particular test.

At this point in the study, it was decided to adjust the cut-off score to try to reduce the controls' high failure rate on the Spatial Acalculia Test, as one in three had failed. To achieve this, the cut-off scores contained in the GSNSB were reduced from two correct answers to one. Once this was done, another chi-squared statistic was used to see whether the reduced cut-off scores could discern between the control participants and the right hemisphere lesion patients. A chi-squared statistic of $\chi^2 = (2, N = 45) = 25.79, p < 0.001$, was established, indicating that how the participants performed on the Spatial Acalculia Test in fact depended on which of the three groups they were from. Overall, this adjustment of the cut-off scores resulted in only two controls failing, as opposed to the original five. Additionally, one more right hemisphere lesion patient passed. However, no further improvement was seen with the left hemisphere lesion group.

3-D Analysis Test

To investigate the results of the 3-D Analysis Test, in order to determine whether the left hemisphere lesion, the right hemisphere lesion, and the control group performed differently from one another, a chi-squared test was again performed (as the data from the GSNSB scores was again nominal in nature). Here, a chi-squared statistic of $\chi^2 = (2, N = 45) = 21.71, p < 0.001$, was found, showing that the group from which the participants came did influence their 3-D Analysis Test performance. In total, 13 of the controls passed the test. All 15 of the right hemisphere lesion patients failed, while only four left hemisphere lesion patients failed.

To determine whether the scores of the left hemisphere lesion patients on the 3-D Analysis Test were influenced by their level of education, a chi-squared test was again used. A chi-squared statistic of $\chi^2 = (1, N = 45) = 0.17, p = 0.680$, was attained, indicating that level of education did not influence how they scored on the test. Similarly, in order to see the possible influence of first language on the 3-D Analysis Test, a chi-squared test was used. A chi-squared statistic of $\chi^2 = (2, N = 45) = 0.51, p = 0.774$, indicated that, as was the case with education, the first language spoken by the left hemisphere lesion group was not related to how they fared on the 3-D Analysis Test.

Executive Function section

To ascertain whether the Executive Function section of the GSNSB could differentiate between patients with frontal and hippocampal lesions, and controls, one-way ANOVA was used (see Table 6.18 for descriptive statistics). ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p = 0.008$). A significant effect was demonstrated for lesion type, $F(2, 42) = 39.50 (p < 0.001)$, indicating

that there was a large discrepancy between how these three groups performed on this overall section. An effect size of 0.63 was established — in other words, 63 percent of the variation of the participants' overall performances on this section could be explained by the group they were from. Figure 6.5 shows participants' performances on the Executive Function section by group.

It was also found, when a further Tukey's HSD analysis was performed, that significant differences in terms of performance were present between the frontal lesion and hippocampal lesion groups ($p = 0.013$), and between the control and frontal lesion groups ($p < 0.001$). Finally, a significant difference was also found between how the control group and the hippocampal lesion group performed ($p < 0.001$).

Table 6.18

Mean scores for Executive Function section

		Pathology Group		
		Hippocampal	Frontal	Control
Score for Executive section	<i>M</i>	5.40	3.27	9.33
	<i>SD</i>	2.85	0.96	1.59
	<i>N</i>	15	15	15

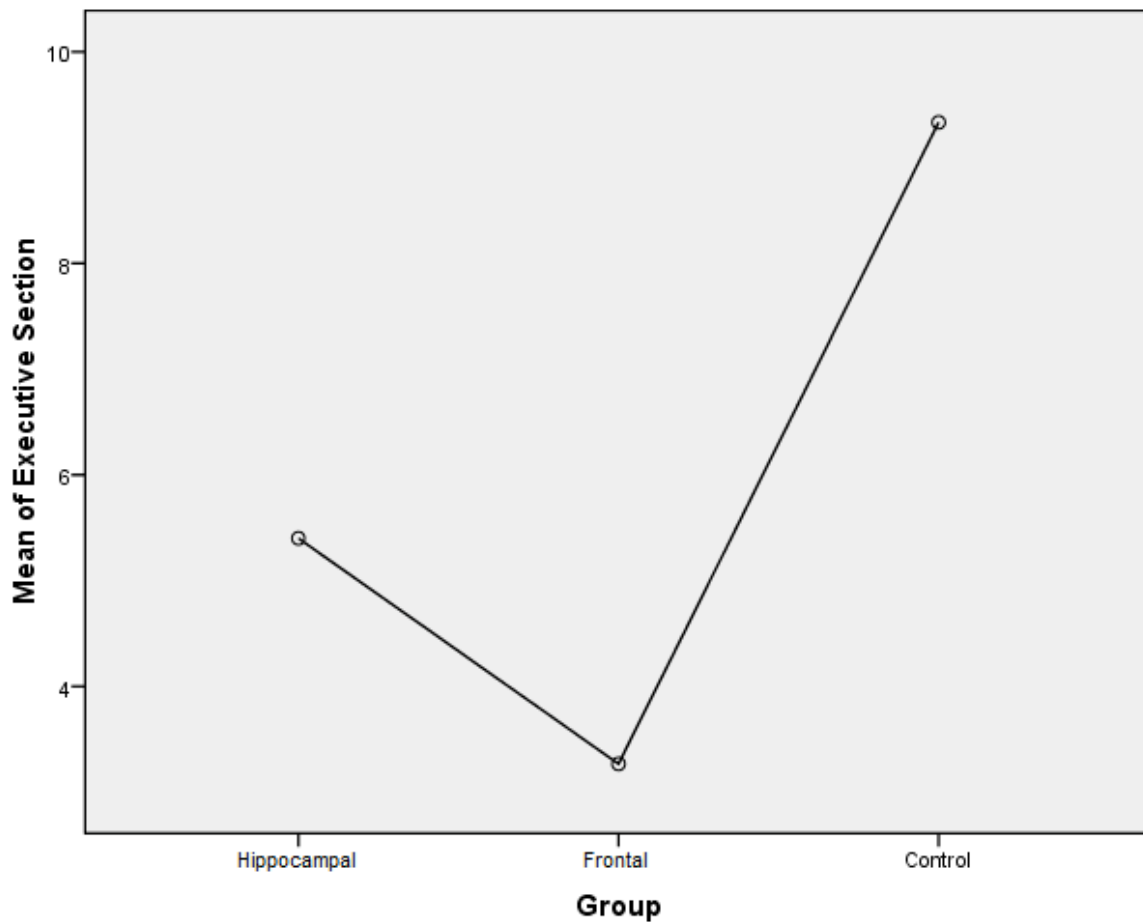


Figure 6.5. Cell mean plot showing mean performances on the Executive section by group

In order to investigate if the variables of level of education and first language affected participants' scores on the Executive Function section, factorial ANOVA was run (see Table 6.19 for descriptive statistics), with the assumptions of normality and homogeneity of variance both being upheld. The results showed that no significant main effects were established for either education level, $F = (1, 39) 0.31, p = 0.861$, or for language, $F = (2, 39) 3.23, p = 0.726$. Finally, no significant interaction effect, $F = (1, 39) 1.312, p = 0.281$, was found between level of education and first language. Therefore, these results showed that the participants' level of education and their first language did not significantly affect how they performed on this overall section.

Table 6.19

Mean scores for Executive Function section

Factor A: Education		Factor B: Language			A marginal means
		English	Afrikaans	isiXhosa	
9 years or more	<i>M</i>	7.09	7	3.00	6.84
	<i>SD</i>	3.081	3.830	.	
	<i>N</i>	11	7	1	
less than 9 years	<i>M</i>	4.56	5.40	6.43	5.38
	<i>SD</i>	3.283	3.239	2.225	
	<i>N</i>	9	10	7	
B marginal means		5.95	6.06	6.00	

The 18 Book Problem

To determine whether the participants from the frontal lesion, hippocampal lesion and control groups performed differently from one another on the 18 Book Problem, a chi-squared test was used. A chi-squared statistic of $\chi^2 = (4, N = 45) = 8.419, p = 0.760$, was calculated, showing that the group from which participants came did not influence significantly how they performed on this test. In total, 14 frontal lesion patients, 11 hippocampal lesion patients and seven controls failed the test.

FAS/BHP/NPS Test

To establish whether the *FAS/BHP/NPS* Test was able to distinguish a frontal lesion from a hippocampal lesion and from a control, one-way ANOVA was used (see Table 6.20 for descriptive statistics). The assumptions of ANOVA were followed and normality was upheld,

but homogeneity of variance was violated ($p = 0.008$). However, the analysis proceeded, as sample sizes were equal and ANOVA is a robust technique. For lesion type, a significant effect was demonstrated, $F(2, 42) = 9.394, p < 0.001$, indicating that lesion-site was a key determinant of FAS/BHP/NPS Test score.

Table 6.20

Mean scores for FAS/BHP/NPS Test

		Pathology Group		
		Hippocampal	Frontal	Control
FAS/BHP/NPS Test score	<i>M</i>	17.6	10.8	29
	<i>SD</i>	15.445	4.799	11.982
	<i>N</i>	15	15	15

To investigate where the differences between the two lesion groups and the controls lay, a Tukey's HSD was performed. Figure 6.6 shows how the three groups performed on the test. It was found that no significant difference ($p = 0.256$) existed between the frontal lesion group and the hippocampal lesion group. A significant difference ($p = 0.027$) was found between the hippocampal lesion group and the control group. Finally, a large significant difference was found between the frontal lesion group and the control group ($p < 0.001$). An effect size of 0.69 indicated that 69 percent of the variation in the participants'

FAS/BHP/NPS Test performance could be accounted for by the group they were from.

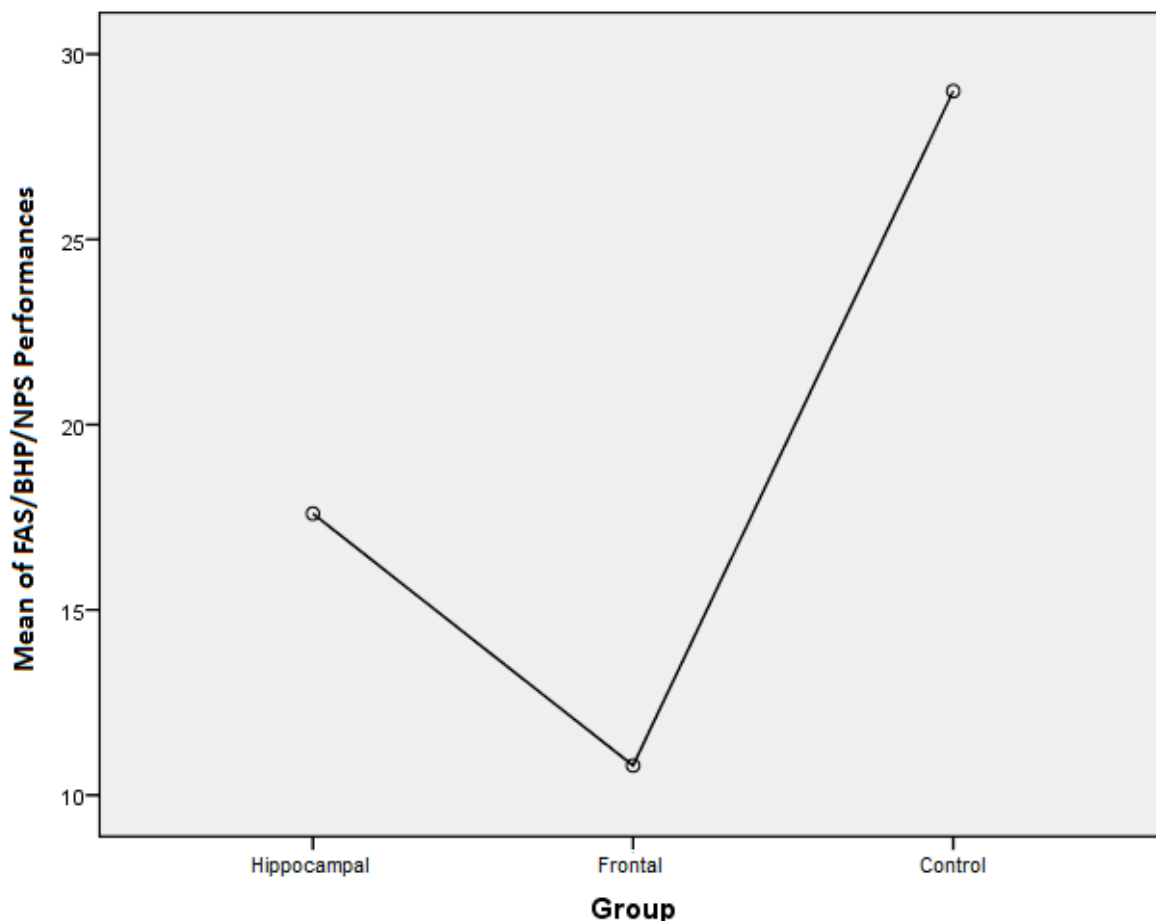


Figure 6.6. Cell mean plot of total FAS/BHP/NPS Test performances by group

In order to investigate whether the variables of level of education and first language affected participants' scores on the FAS/BHP/NPS Test, factorial ANOVA was run (see Table 6.21 for descriptive statistics), the assumptions of normality and homogeneity of variance both being upheld. Once more, participants were divided between those with less than nine years of education and those with nine or more years. The results showed that no significant main effects were established for either education level, $F(1, 39) = 0.290, p = 0.593$, or for language, $F(2, 39) = 1.299, p = 0.284$. Finally, no significant interaction effect, $F(2, 39) = 1.318, p = 0.279$, was found between level of education and first language. Therefore, these results showed that the participants' level of education and their first language did not significantly affect how they performed on the FAS/BHP/NPS Test.

Table 6.21

Mean scores for the FAS/BHP/NPS Test

Factor A: Education		Factor B: Language			A marginal means
		English	Afrikaans	isiXhosa	
9 years or more	<i>M</i>	25.36	26.29	3	24.53
	<i>SD</i>	13.626	14.130	.	
	<i>N</i>	11	7	1	
less than 9 years	<i>M</i>	17.22	13.40	15.14	15.19
	<i>SD</i>	15.643	10.814	9.788	
	<i>N</i>	9	10	7	
B marginal means		21.70	18.71	13.62	

Secondary Evaluation of the Groote Schuur Neurocognitive Screening Battery's Validity

Memory Function Section

As part of the secondary level of analysis, an additional one-way ANOVA was performed to determine whether the total Memory Function section of the GSNSB could discriminate between hippocampal lesion, left hemisphere lesion, right hemisphere lesion and frontal lesion patients. ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p < 0.001$). However, given that ANOVA is a robust technique and sample sizes were equal, the analysis was continued. Table 6.22 provides the descriptive statistics (means and standard deviations) of the analysis. A significant main

effect of, $F(4, 70) = 22.63, p < 0.001$, was found. The effect size was determined by calculating *eta*-squared, which was found to be 0.5390. Therefore, lesion-site accounted for 53.90 percent of the variation seen in the Memory Function section performances.

Table 6.22

Mean score for the Memory Function section

		Pathology Group			
		Hippocampal lesion	Left Hemisphere lesion	Right Hemisphere lesion	Frontal lesion
Memory Function section score	<i>M</i>	5.87	5.73	15.40	13.27
	<i>SD</i>	4.32	7.41	1.84	4.35
	<i>N</i>	15	15	15	15

A post-hoc test (Tukey's HSD) was performed to determine the location of the difference. Comparisons between those participants with hippocampal lesions and those with left hemisphere lesions showed no significant difference ($p = 0.999$). However, the comparison between the hippocampal lesion group and the right hemisphere lesion group was highly significant ($p < 0.001$). The comparison between the left hemisphere lesion group and the right hemisphere lesion group was significantly different ($p < 0.001$). The comparison between the left hemisphere lesion group and the frontal lesion group was significantly different ($p < 0.001$). The comparison between the right hemisphere lesion group and the frontal lesion group was not significant ($p = 0.674$). The comparison between the

hippocampal lesion group and the frontal lesion group is reported in the 'Primary Evaluation of the Groote Schuur Neurocognitive Screening Battery's Validity' section above.

Therefore there was a significant difference between how the hippocampal lesion group performed on the Memory section compared to the right hemisphere lesion group, but not compared to the left hemisphere lesion group. There was also a significant difference between how the left hemisphere lesion group performed relative to both the right hemisphere lesion and the frontal lesion groups.

Language Function Section

As a secondary level of analysis, a one-way ANOVA was performed to determine whether the total Language Function section of the GSNSB could discriminate between left hemisphere lesion, hippocampal lesion, right hemisphere lesion and frontal lesion patients. ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p = 0.002$). However, because ANOVA is a robust technique and sample sizes were equal, the analysis continued. Table 6.23 provides the descriptive statistics (means and standard deviations) for the analysis. A significant main effect was found, $F(4, 70) = 40.06, p < 0.001$. The effect size was determined by calculating *eta*-squared, which was found to be 0.6785. Therefore, lesion-site accounted for 67.85 percent of the variation in Language Function section performances.

Table 6.23

Mean score for the Language Function section

		Pathology Group			
		Left Hemisphere lesion	Hippocampal lesion	Right Hemisphere lesion	Frontal lesion
Language Function section score	<i>M</i>	10.20	25.13	25.87	26.07
	<i>SD</i>	6.44	5.64	3.07	4.33
	<i>N</i>	15	15	15	15

A post-hoc test (Tukey's HSD) was performed to determine the location of the difference. Comparisons between those patients with left hemisphere lesions and hippocampal lesions showed a strong significant difference ($p < 0.001$). Comparisons between the left hemisphere lesion group and the frontal lesion group were also highly significant ($p < 0.001$). The comparison between the hippocampal lesion and the frontal lesion groups, however, was not significant ($p = 0.980$). The comparison between the left hemisphere lesion group and the right hemisphere lesion group is reported in the 'Primary Evaluation of the Groote Schuur Neurocognitive Screening Battery's Validity' section above. The comparison between the hippocampal lesion group and the right hemisphere lesion group was not significant ($p = 0.992$). The comparison between the frontal lesion group and the right hemisphere lesion group was not significant ($p = 0.999$). Therefore there was a significant difference between

how the left hemisphere lesion group performed on the Language Function section compared to both the frontal and hippocampal lesion groups.

Spatial Cognition Section

A one-way ANOVA was performed to determine whether the total Spatial Cognition section of the GSNSB could discriminate between right hemisphere lesions, frontal lesions, left hemisphere lesions and hippocampal lesions. ANOVA's assumptions of normality and homogeneity of variance were upheld. Table 6.24 provides the descriptive statistics (means and standard deviations) for the analysis. A significant main effect, $F(4, 70) = 19.50, p < 0.001$, was established. The effect size was determined by calculating *eta*-squared, which was found to be 0.4999. Therefore, lesion-site accounted for 49.99 percent of the variation in Spatial Cognition section performances.

Table 6.24

Mean score for the Spatial Cognition section

		Pathology Group			
		Right Hemisphere lesion	Hippocampal Hemisphere lesion	Left Hemisphere lesion	Frontal lesion
Spatial Cognition section score	<i>M</i>	5.13	9.87	7.87	10.60
	<i>SD</i>	2.33	2.59	4.00	1.72
	<i>N</i>	15	15	15	15

A post-hoc test (Tukey's HSD) was performed to determine the location of the difference. Comparisons between those patients with right hemisphere lesions and hippocampal lesions showed a strong significant difference ($p < 0.001$). Comparisons between the right hemisphere lesion group and the frontal lesion group were also highly significant ($p < 0.001$). Comparisons between the hippocampal lesion group and the frontal lesion group were, however, not significant ($p = 0.938$). The comparison between the right hemisphere lesion group and the left hemisphere lesion group is reported in the 'Primary Evaluation of the Groote Schuur Neurocognitive Screening Battery's Validity' section above. The comparison between the hippocampal lesion group and the left hemisphere lesion group was not significant ($p = 0.229$). The comparison between the frontal lesion group and the left hemisphere lesion group was marginally significant ($p = 0.041$).

Therefore there was a significant difference between how the right hemisphere lesion group performed on the Spatial Cognition section when compared with the hippocampal lesion and frontal lesion groups. There was also a significant difference between how the left hemisphere lesion group performed relative to the frontal lesion group.

Executive Function Section

Finally, a one-way ANOVA was performed to determine whether the total Executive Function section of the GSNSB could discriminate between the frontal lesion, left hemisphere lesion, hippocampal lesion and right hemisphere lesion groups. ANOVA's assumption of normality was upheld, but the assumption of homogeneity of variance was violated ($p < 0.003$). However, because ANOVA is a robust technique and sample sizes were equal, the analysis was continued. Table 6.25 provides the descriptive statistics (means and standard deviations) for the analysis. A significant main effect, $F(4, 70) = 18.52, p < 0.001$,

was found. The effect size was determined by calculating *eta*-squared, which was found to be 0.4864. Therefore, lesion-site accounted for 48.64 percent of the variation in the Executive Function section performances.

Table 6.25

Mean score for the Executive Function section

		Pathology Group			
		Frontal lesion	Left Hemisphere lesion	Hippocampal lesion	Right Hemisphere lesion
Executive Function section score	<i>M</i>	3.27	3.67	5.40	5.60
	<i>SD</i>	0.96	2.72	2.85	2.10
	<i>N</i>	15	15	15	15

A post-hoc test (Tukey's HSD) was performed to determine the location of the difference. Comparisons between those patients with frontal and those with left hemisphere lesions showed no significant difference ($p = 0.986$). Comparisons between the frontal lesion group and the right hemisphere lesion group were, however, significant ($p = 0.033$). Comparisons between the left hemisphere lesion group and the right hemisphere lesion groups were found not to be significant ($p = 0.114$). The comparisons between the left hemisphere lesion group and the hippocampal lesion group were not significant ($p = 0.193$), as was the correlation between the right hemisphere lesion group and the hippocampal lesion group ($p = 0.999$). The comparison between the frontal lesion group and the hippocampal lesion group is

reported in the ‘Primary Evaluation of the Groote Schuur Neurocognitive Screening Battery’s Validity’ section above.

Therefore there was a significant difference between how the frontal lesion group performed on the Executive Function section when compared to the right hemisphere lesion group, but there was no significant difference when compared to the left hemisphere group.

Analysis of the Orientation Section

To investigate whether a relationship existed between the ‘Orientation’ section and the other sections of the GSNSB, a correlation coefficient was calculated between the ‘Orientation’ section score and each of the other four GSNSB section scores. The results from the Pearson’s product moment correlation coefficients (r) calculated for all four comparisons can be seen in Table 6.26.

Table 6.26

Correlation Coefficients for the Orientation section in relation to the Memory, Language, Spatial Cognition and Executive Function sections

Section compared to Orientation section	Correlation Coefficient
<i>Memory Function section</i>	0.87
<i>Language Function section</i>	0.74
<i>Spatial Cognition section</i>	0.42
<i>Executive Function section</i>	0.59

With these results, a high positive correlation indicated that a strong positive relationship between the particular section and the 'Orientation' sections existed — in other words, an individual with a high score for the 'Orientation' section was likely to also score highly in the other section of the GSNSB and, vice versa. The results showed that both Memory Function section and Language Function section performances correlated strongly with 'Orientation' section performance.

Discussion

The overall purpose of this study was to investigate the reliability and validity of the newly developed GSNSB. Numerous authors have emphasised the importance of having a battery validated before it is used clinically to ensure that the tool produces consistent and dependable results (Anastasia & Urbina, 1997; Hebben & Milberg, 2002; Lezak et al., 2004; Mitrushina et al., 2005; Rosenfeld, Sands & VanGorp, 1997; Russell & Russell, 2003; Russell et al., 2005; Spreen & Strauss, 1991). Given the significant progress that had already been achieved through the adaptation and development of new, culturally specific, South African neurocognitive tests, it was imperative that the overall efficacy of the GSNSB also be demonstrated. This phase of the study therefore represents a major step towards the provision of effective neurocognitive screening for the South African medical community.

Reliability

The first requirement was to investigate the reliability of the screening tool both in terms of its test-retest reliability and its inter-rater reliability. Due to the careful design of the process to test reliability, both these tasks were achieved simultaneously as the reassessments were completed by the assessor who had not done the initial assessment. Here, the results were most pleasing, indicating that both the GSNSB's re-test and inter-rater reliability were sound

— the correlations of 0.91 for the Orientation section, 0.98 for the Memory Function section, 0.99 for the Language Function section, 0.98 for the Spatial Cognition section and 0.96 for the Executive Function section were indicative of excellent overall reliability. The correlation coefficient of 0.99 for the total GSNSB scores also served as confirmation of this finding, indicating that overall the participants' performances on the GSNSB over the two assessments hardly differed, regardless of the section being administered. This finding has important implications for the GSNSB's future use, as it indicates that despite the potential variability in neurocognitive functions, the confidence the clinician can have in the tool is high.

These findings were especially convincing because of the high correlation coefficients achieved — a Pearson's product-moment correlation coefficient value over 0.7 is generally considered as a strong correlation, while 0.9 and over is considered extremely strong (Durrheim, 1999; Howell, 1999). Furthermore, establishing the reliability of neuropsychological tools and tests is traditionally widely accepted as being difficult to achieve effectively because many neurocognitive syndromes and deficits resolve quickly over time (Bleiberg et al., 1997; Walsh & Darby, 1999).

Primary Evaluation of Validity

The goal of establishing the validity of the GSNSB was achieved through the use of both qualitative and quantitative measures. Again, this approach of adopting converging lines of evidence was deemed the most appropriate because the GSNSB itself was designed using both qualitative and quantitative methods, thus allowing the one approach to compensate for the weaknesses of the other and vice versa. This also enabled a thorough analysis to be performed on a variety of possible variables (namely age, education and language), other than

lesion-site, that could have influenced performance on the GSNSB. As has been highlighted by various authors, qualitative observations provide insight into the nosology, the underlying mechanisms and the pathology that produce neurocognitive deficit, as well as the relationship of such deficits relative to overall brain functioning (Christensen & Caetano; Luria, 1966, 1973; Luria & Majovski, 1977; Walsh, 1991; Walsh & Darby, 1999). In contrast, quantitative analysis allows the patient's neurocognitive deficits to be measured and quantified, enabling his/her performance to be objectively compared to those of others.

The overall objective of establishing validity was to demonstrate in a clinically meaningful manner whether or not the screening tool was able to differentiate between patients with various lesions and neurocognitively intact individuals. Specifically, the primary focus was to ensure that the GSNSB was able to discriminate between frontal versus hippocampal lesions, and between left hemisphere versus right hemisphere lesions.

Validity through Qualitative Analysis

In terms of the overall qualitative analysis of the GSNSB, the level of agreement between the two assessors when allocating the participants to groups was pleasing. Twelve errors were made out of the total of 180 allocations (representing a 93.34 percent success rate). This indicates that qualitatively the constellation of tests assembled throughout the GSNSB was reasonably effective in discriminating between patients from the four lesion groups and from neurocognitively intact controls.

These results serve as a positive affirmation that the design of the GSNSB, in terms of its use of theoretical underpinnings and its utilisation of the hypothetico-deductive approach, has proven effective. This is an especially important finding, given that the design and overall

effectiveness of the decision-trees included in the GSNSB rest on the qualitative hypothetico-deductive insights gained from these approaches. A number of the qualitative observations made regarding the participants' performances provide further affirmation. Overall, the findings show that the scoring instructions and decision-tree procedures used in the GSNSB were easy to interpret. Furthermore, the constellations of neurocognitive tests assembled in each section were effective in eliciting the deficits typical of certain lesions, when viewed from a qualitative point of view.

In support of this, qualitatively discernable differences in test performance were noted in the Memory Function section. None of the control participants found the 4 Hidden Objects Test challenging, while the hippocampal lesion patients had major difficulties. The frontal lesion patients performed worse than the controls, although better than the hippocampal lesion patients. Here, the frontal lesion patients did not struggle from the outset, as the hippocampal ones had, but rather as the test increased in complexity. Logically, this observation suggests that as the task taxed their retrieval mechanisms more and more — in other words, once multiple locations were introduced to the task — it began to challenge their functional capacity. These observations are consistent with the literature in that executive impairment frequently results in 'frontal amnesia', and therefore such patients would be expected to underperform on memory tasks (Luria, 1973; Solms & Turnbull, 2002; Walsh & Darby, 1999). The observation that these patients fared better than those with hippocampal lesions is not surprising, given that primary memory processes (that is, the encoding of memory) remain intact in the case of frontal lesions (Luria, 1973; Solms & Turnbull, 2002; Walsh & Darby, 1999).

Qualitative observations showed that many of the participants performed poorly on the Auditory Span Test regardless of group, despite the test being able to successfully differentiate between the three participant groups. This test was qualitatively not as good at discerning between patients of different pathologies. Many of the patients failed this test, most likely because only one trial per sequence was provided. Therefore, no clear qualitative patterns of performance across patient groups were observed. However, because this test was worth only '2' of the total of '18' points allocated for the section, this difficulty did not affect the patterns of deficit observed across the patient groups in the Memory Function section as a whole, which was qualitatively able to discern between hippocampal- and frontal-lesion patients and controls when scores were not considered.

On the Language Function section, qualitatively noticeable differences in test performance were seen between the left hemisphere lesion patients relative to the right hemisphere lesion patients and the controls. On the Washing Line Picture Test, while many of the left hemisphere lesion patients produced a non-fluent performance, all the controls and the majority of the right hemisphere lesion patients performed fluently. The few right hemisphere patients who did not perform perfectly on the test did so because their unilateral neglect meant that they did not attend to the left half of the picture. The decision-tree for this particular test should thus include instructions/steps to exclude a primary visual deficit (such as a hemianopia, for example) or unilateral neglect, before the test is begun.

From the qualitative observations concerning the Naming Test, the majority of the controls and the right hemisphere lesion patients performed well. A number of the left hemisphere lesion patients had word-finding difficulties with this test, and a number of paraphasias, both

semantic and literal, were observed. For example, one patient named the *spear* item a *spade*, and the *goal* was named a *window*.

The effect of education on performance in the Language Function section was observed when some of the controls fared poorly on the writing task when asked to produce a spontaneous sentence of their choice. It appeared that this stemmed from their low level of education, which resulted in misspellings and grammatical errors (for example, *there* written as *their*). However, these were only minor problems and, in general, the controls performed well throughout this section. With regard to the repetition task, it was pleasing to note that many left hemisphere lesion patients struggled particularly with the longer sentences, while the controls performed well. The right hemisphere lesion patients who performed poorly on the test did so in a qualitatively different manner, making errors due to their emotionally inappropriate responses, rather than errors due to aphasic difficulties. An example of this qualitative difference is that the right hemisphere lesion patients made rude comments about their doctor when asked to repeat the sentence “*this doctor does not visit all the patients in the ward*”, while the left hemisphere lesion patients failed because they lost the sentence and were unable to repeat it fully. Overall, this section as a whole was, from a qualitative standpoint, more than adequate in discerning left hemisphere lesions.

Qualitative observations of the Spatial Cognition section indicated that it was able to differentiate those patients with right hemisphere lesions from those with left hemisphere lesions and from controls. However, the patients with left hemisphere lesions were seen to perform poorly on the section as a whole relative to the controls, although not as poorly as they did on the Language Function section — indicating that aphasic deficits were the primary reason for their test failings. Specifically, the left hemisphere lesion patients

performed poorly on the 3-D Analysis Test, either because they struggled to understand the test instructions or due to perseveration, which was present in some of the cases. This finding is, however, perhaps not surprising and is consistent with observations made in the literature that patients with aphasic difficulties typically struggle on tests other than those simply testing language, as a consequence of failing to comprehend test instructions and procedures, as language is the medium through which these are explained (Benson, 1979; Hebben & Milberg, 2002). In addition, patients with left hemisphere lesions are also prone to experience problems with spatial cognition, but these performances differ qualitatively from those resulting from right hemisphere lesions. Here, rather than a disintegration of the spatial components of the task, as is the case with right hemisphere lesions, patients with left hemisphere lesions retain the overall gestalt of the image, but lack the detail. Consequently, such patients may fail tests of spatial cognition (Benton, 1979; Delis & Bihrlé, 1989; Devinsky & D'Esposito, 1992; Solms & Turnbull, 2002; Walsh & Darby, 1999). These key qualitative insights underscore the need to incorporate steps to differentiate between a left hemisphere deficit of spatial cognition and a right hemisphere deficit into the GSNSB's decision-trees in the future.

Qualitative observations from the Spatial Acalculia Test showed that the majority of participants performed poorly, but again in qualitatively distinct ways. The right hemisphere lesion patients struggled due either to unilateral neglect resulting in their ignoring the left part of the sums, or due to spatial difficulties. Alternatively, the left hemisphere lesion patients found difficulty in understanding the test instructions and what was required of them. The controls struggled with the complexity of the sums, even though they understood what the task required in general terms. The Spatial Acalculia Test was the only test in the Spatial Cognition section that the controls struggled with.

In general, the Executive Function section can be viewed as valid on qualitative grounds, given the success rate of the qualitative allocation process. However, a large number of false positives were present, with eight of the 15 control participants failing the section according to the original cut-off score (that is, the one used prior to the validation process). Here, it was observed that the 18 Book Problem could account for this high error rate, as seven of the controls failed this test. If performances on this test are excluded, then only one of the controls failed this section. The raters reported that, of the tests of executive function, the FAS/NPS/BHP Test served as the most accurate in guiding them towards the correct qualitative allocation of patients, as it was evident that the frontal lesion patients performed poorly on this particular test relative to the other groups. It was qualitatively observed that the frontal lesion patients were the only ones to consistently repeat words, regularly give proper nouns and generally produce very few words.

Overall, many of the right hemisphere lesion patients were observed to be emotionally inappropriate throughout their testing, which adversely affected their cooperation and concentration, especially on the more difficult tasks. They were also found to be easily distracted and reluctant to engage with some of the more challenging tests. Likewise, many of the left hemisphere lesion patients performed poorly throughout all the tests, due to the obvious handicap that problems understanding tests and test instructions produce. These key additional qualitative observations as to how patients with specific pathologies might approach the task should in future be included in the decision-tree instructions so that the assessor is forewarned about potential problems with the administration of the GSNSB to such patients.

Validity through Quantitative Analysis

Memory Function Section

When focusing on the validation of the Memory Function section, the central goal was to examine whether the assembled tests were able to successfully distinguish between patients with hippocampal lesions and frontal lesions, and neurocognitively intact individuals. In other words, was the performance of the hippocampal lesion patients significantly different from the other two groups? The results indicated that this section was indeed able to differentiate between the three groups, with the hippocampal lesion patients performing at far lower levels than the other two groups. In total, lesion-site explained 65 percent of the variation seen in performance. It was also found that frontal lesion patients performed significantly differently from the controls. Once more, this finding is largely to be expected, given that patients are entitled to have ‘memory’ problems on an executive basis due to deficits in the *retrieval* of memory (Luria, 1973; Solms & Turnbull, 2002; Walsh & Darby, 1999).

A pleasing result was the fact that when statistically investigated, it was found that neither first language nor education significantly influenced performance in the Memory Function section of the GSNSB. The possible influence of age was also found not to significantly influence performance in this section.

The validity of the Auditory Span Test as an individual test contained in the Memory Function section was also closely examined. It was found that this test effectively distinguished between frontal lesion patients and neurocognitively intact individuals. However, no significant difference was demonstrated between how the hippocampal lesion

patients and the neurocognitively intact individuals performed. This latter finding is most likely accounted for by the fact that short-term/working memory, which the Auditory Span Test assesses, is largely an executive (frontal) function (Kolb & Whishaw, 2003; Solms & Turnbull, 2002). Patients with hippocampal lesions, which affect the encoding of memory, should therefore still have intact working memory, as this is not a function of the hippocampi (Evans, 2004; Walsh & Darby, 1999). In light of both the statistical and qualitative findings, it was evident that the Auditory Span Test needed further improvement, as five of the control participants scored '1' out of 2' for the test (that is, a span of five digits) and three failed it outright. A possible reason for this finding was that only one trial per number sequence was provided. Therefore, the function was not sufficiently probed in each individual and it is possible that the test was therefore ended prematurely, before the participant's true ability on the test could be ascertained. It was decided that more trials per sequence needed to be provided. The overall validity of the Memory Function section was, however, not seen to be adversely affected by these shortcomings in the current format of the Auditory Span Test, as this test is worth only a small percentage of the section's total score.

Language Function Section

The aim of validating this section was to investigate whether the constellation of tests assembled was able to meaningfully discern left hemisphere lesion patients from right hemisphere lesion patients and from neurocognitively intact individuals. In other words, did the left hemisphere lesion patients perform significantly differently from the other two groups? The results showed, most convincingly, that the GSNSB was able to discern those with left hemisphere lesions from those without, as was evidenced by the findings that the left hemisphere group scored much lower on this section than the other two groups. The *eta*-squared effect size supported this conclusion, indicating that 80.35 percent of the variation in

the participants' Language Function section performances could be accounted for by the group they were from.

The other results pertaining to the Language Function section showed that none of the variables of 'age', 'education level' or 'first language' significantly affected the participants' performances. This result is most satisfying and is a testament to the success of the previous piloting and re-piloting work on the neurocognitive tests undertaken in order to reduce the influence of these variables on test outcome. In this light, the overall outcome can be deemed a success. It must, however, be remembered that due to the sample size and the nature of the statistical procedures used, the variables of 'education level' and 'age' were subdivided into only two categories (those with less than eight years of education and those with eight or more years of education; those aged less than 50 versus those aged 50 or older, respectively). If the influence of these variables is to be examined in more detail, future studies will be required examining a broader range of education levels and ages and, if possible, comparing a much older sample with a much younger one. Given the nature of clinical neuropsychological research, this task will be difficult, as it is not easy to find suitable patients in large numbers because of the scarcity of certain pathologies and the rate at which many deficits resolve. However, for the purposes of the initial validation of the GSNSB and given the objectives of this study, this present body of work is more than adequate.

After investigating the effectiveness of the whole section, the validity of some of the individual tests was also examined. As had been the case with the Language Function section as a whole, the results also showed that the Naming Test effectively differentiated between patients with left hemisphere lesions and those with right hemisphere lesions, as well as healthy controls. Here, the results were highly significant, with a probability value of ($p <$

0.001) indicating that, on the whole, a clinically effective, culturally appropriate test had been created.

The factorial ANOVA performed on the Naming Test found that level of education significantly affected the participants' performances, while age and first language did not. The significant difference of ($p = 0.049$) was, however, relatively small; the findings also showed that despite 'education level' having a significant effect, all the participants still performed within the range of normal function. Again, as mentioned above, it is proposed that further studies examining the influence of education level in more detail be undertaken. The finding regarding the influence of education level on confrontation naming is consistent with the findings of other studies. For example, Delouche et al.(1996), Hawkins and Bender (2002) and Saxton et al. (2000) all investigated the Boston Naming Test across participants of varying education level, finding that those with lower levels of education performed poorer and with more variability than well-educated individuals.

Spatial Cognition Section

The primary aim of validating this section was to investigate whether the constellation of tests assembled was able to meaningfully discern right hemisphere lesion patients from left hemisphere lesion patients and from healthy individuals. In other words, did the right hemisphere lesion patients perform significantly differently from the other two groups? Once again, the results showed convincingly that this section of the GSNSB was successful in differentiating between these three groups — a significant difference ($p = 0.031$) was found between how the right hemisphere lesion patients performed compared to patients with left hemisphere lesions. A significant difference ($p < 0.001$) was also found between how the right hemisphere lesion patients and the healthy controls performed on the section.

A third significant difference ($p < 0.001$) was found between how the healthy controls performed relative to the left hemisphere lesion patients. This latter finding indicated that although the section was able to distinguish between right and left lesions and controls, the left hemisphere lesion patients did not perform similarly to the neurocognitively intact individuals. This finding is not particularly surprising, as difficulties with language inevitably result in left hemisphere patients failing a wide array of tests other than tests of language.

A second possible reason for this finding, in keeping with the above-mentioned qualitative observations, is that left hemisphere lesion patients are also entitled to experience deficits of spatial cognition as a result of possible left parietal lesions (Benton, 1979; Delis & Bihrlé, 1989; Devinsky & D'Esposito, 2003; Walsh & Darby, 1999). Such deficits are not qualitatively the same as spatial difficulties resulting from right hemisphere lesions. This illustrates the importance of not only adopting converging lines of evidence in understanding a patient's test performance, but also of ensuring that qualitative observations inform the testing process (that is, the decision-tree approach). The findings thus highlight the need to include additional steps in the decision-trees in order to exclude possible left hemisphere spatial deficit before proceeding with the assessment of possible right hemisphere deficits in the current Spatial Cognition section.

The results of the factorial ANOVA revealed that 'education' as a variable significantly affected ($p = 0.039$) the participants' performances on this section of the GSNSB, while 'first language' did not. This finding is again consistent with the available literature, which has shown that tests of spatial cognition are sensitive to education level (Lezak, 1995). The implication of this finding is that the GSNSB's decision-tree approach needs to incorporate

procedures accommodating the possible affect of education on the test before it is fully interpreted.

After investigating the effectiveness of the section as a whole, the validity of some of the individual tests was also examined. Firstly, the results from the 3-D Analysis Test showed that this test was effective in differentiating between right hemisphere lesion patients and healthy individuals, as was evident from the finding that just two controls failed, while all the right hemisphere patients failed. The left hemisphere lesion patients performed poorly on this test too — only four passed — relative to the healthy controls. Once again, this finding can be explained by aphasic patients experiencing great difficulty with understanding test instructions in general and, as a result, struggling on most neurocognitive tests. Added to this disadvantage is the fact that visuospatial tests such as the 3-D Analysis Test and the original Cube Analysis Test are widely regarded as being difficult (Lezak, 1995). In fact, the qualitative feedback from the controls revealed that the majority reported finding this particular test difficult. Given that a floor effect seemed evident with this particular test, further work will need to be done to either simplify the test still further or to provide an alternative test in the GSNSB, making the 3-D Analysis Test an optional test for those deemed able to cope with it.

Further encouraging findings resulting from the chi-squared performed on the 3-D Analysis Test revealed that neither first language nor level of education appeared to influence how participants fared on this test. This result, which confirmed that a culturally appropriate neurocognitive test had been successfully created for the South African context, is in pleasing contrast to observations by authors such as Rosselli and Ardila (2003) that neurocognitive tests of visuospatial function are typically culturally biased. This result also supports the fact

that the approach adopted in developing the adapted neurocognitive tests was a sound one, and that the goal of eliminating as far as possible the reliance on clinical judgement when examining test performances has been satisfactorily achieved (Russell et al., 2005). This was a key criterion in ensuring that the screening tool developed is fit for use as a 'transferable technology' (Nell, 2000).

Another closely examined test was the Spatial Acalculia Test. The findings had shown that all participants had struggled on this test, with five controls having failed. A significant difference ($p < 0.001$) was found between how the controls performed relative to the left hemisphere and right hemisphere lesion patients, although no significant difference was demonstrated between these latter two groups. This finding indicated that the test could differentiate a right hemisphere lesion patient from a healthy individual, but not a right hemisphere lesion patient from a left hemisphere lesion patient. The fact that so many of the healthy controls had struggled on the test was troubling, especially since neurocognitively intact individuals should not find these tests difficult as they are designed to tax only those with neurocognitive deficits. The results of the chi-squared analysis performed to investigate the possible influence of the participants' level of education and language showed that neither of these variables affected their performances. This was a pleasing result, consistent with the other tests adapted for this section of the GSNSB.

Given the participants' poor performances, the cut-off scores of the test were lowered from two or more correct answers out of three to one or more correct answers to determine if the controls would then fair better. Once this alteration had been made, only two of the controls failed the test, which was also still able to differentiate patients with right hemisphere lesions

from controls. This finding suggests that the original items contained in the GSNSB were too difficult.

The second difficulty highlighted by the results was that the left hemisphere lesion patients performed poorly on the Spatial Acalculia Test. There are a number of possible reasons for this. Firstly, as with all neurocognitive tests, aphasic patients who struggle to comprehend the test instructions or the arithmetic operations will fail a test on the basis of their aphasia rather than because they have a deficit of visuospatial functioning. It is possible for left hemisphere lesion patients to present with primary acalculia or anarithmetria. “Acalculia may occur with any lesion of the left cerebral hemisphere that produces aphasia, but is mostly likely to be associated with lesions of the posterior temporal or parietal region” (Hebben & Milberg, 2002, p. 211). These primary deficits present with an inability to compute numerals and figures on an alexic basis, and with arithmetical sums (anarithmetria), rather than with the spatial components of calculations as seen in spatial acalculia (Ardila, Lopez & Solano, 1989; Levin & Spiers, 1979; McNeil, 2004; Walsh & Darby, 1999). Although it is possible to qualitatively differentiate these two clinical phenomena — again highlighting the importance of adopting a hypothetico-deductive approach to understanding test performance — the GSNSB in its initial form clearly did not adequately do this, as the decision-tree used only required the assessor to: “*Establish that the patient can do simple addition and subtraction first...*” without specifying that these initial sums should not be spatially loaded. Here, spatial loading refers to the fact that even arithmetical calculations done in one’s head require a certain degree of spatial ability once the task requires numbers to be moved over decimal places. Simple calculations such as ‘ $4 + 2$ ’ do not have spatial loading — it is verbal examples such as this that are required in order to test for a primary acalculia or anarithmetria.

This is an important finding (and a good case example) for the future development of the GSNSB. It highlights not only how multiple possible determinants of test failure need to be taken into account when further designing the GSNSB's decision-trees, but also the contributions of this research in identifying such determinants. With regard to the Spatial Acaculia test in the GSNSB, the decision-tree requires further modification to include simple sums (not spatially loaded) involving addition and subtraction, and written down as examples in order to be verbally administered. This will allow for the primary calculation deficits first to be excluded before sums involving spatial loading are introduced. In addition, these spatially loaded tasks must be expanded to include some slightly simpler items than the current examples, which are clearly too difficult for many people.

Executive Function Section

The aim of validating this section was to investigate whether the constellation of tests assembled could meaningfully discern frontal lesion patients from hippocampal lesion patients and from neurocognitively intact individuals. In other words, did the frontal lesion patients perform significantly differently from the other two groups? The results showed, most convincingly, that the GSNSB was indeed able to discern patients with frontal lesions from those without, evident from the findings that the frontal lesion group scored much lower on this section than the other two groups. The *eta*-squared effect size supported this conclusion, indicating that 63 percent of the variation in the participants' Executive Function section performances was accounted for by the group they were from.

The other statistical results from this section showed that neither 'education level' nor 'first language' significantly affected the participants' performances. Again, this result is most satisfying and a testament to the success of the piloting and re-piloting work undertaken on

the neurocognitive tests in order to reduce the influence of such variables on the test outcome. However, it must be remembered that due to the sample size and the nature of the statistical procedures used, the variables of 'education level' and 'age' were divided into only two categories (those with less than eight years of education and those with eight or more years of education; those aged less than 50 versus those aged 50 or more, respectively). Therefore, if the influence of these variables is to be examined in more detail, future studies should examine a broader range of education levels and ages and, if possible, compare a much older sample with a much younger one. Given the nature of neuropsychological research, this will be a difficult task, as it is not easy to find suitable patients in large numbers because of the scarcity of certain pathologies and the rate at which many deficits resolve.

Two of the tests contained in the Executive Function section were also examined individually. Significant differences were found in the FAS//BHP/NPS Test between the performance of the frontal lesion patients in relation to the controls, and between the performance of the hippocampal lesion patients in relation to the controls. No significant difference was established in the performance of the frontal lesion and hippocampal lesion patients in relation to each other. However, the test still displayed good validity and the effect size found indicated that 69 percent of the variations in performance on this test could be explained by lesion-site. In addition, the majority of the hippocampal lesion patients performed in the 'adequate performance' range (that is, between 15 and 25 words). This result is most likely explained by the hippocampal patients' tendency to forget instructions and become confused, resulting in poor performance.

The second test examined as part of this section was the 18 Book Problem, which proved problematic. No significant differences were demonstrated between the frontal lesion, the

hippocampal lesion and the control groups. On inspection, this finding can be accounted for by the high failure rate (seven failures) in the control group, indicating that the test was too difficult for the average South African. Therefore, the validity of this test was poor. The overall results for the Executive Function section would be even more positive if this test were to be replaced by one that could better discriminate between the presence versus the absence of pathology.

Secondary Evaluation of Validity

The results from the secondary level of analysis carried out during the overall validation process were also pleasing, providing further evidence of the GSNSB's ability to differentiate between a variety of lesions. Despite the fact that the primary aim of demonstrating validity was to show that the GSNSB was able to differentiate between frontal and hippocampal lesions, and between left hemisphere and right hemisphere lesions, it was also important to ascertain whether the GSNSB was able to discriminate between other combinations of lesions.

The results showed that the Memory Function section was able to differentiate between a hippocampal lesion and a right hemisphere lesion, but not between a hippocampal lesion and a left hemisphere lesion. There was also a significant difference between how the left hemisphere lesion group performed compared to both the right hemisphere and frontal lesion groups, meaning that these groups were not performing uniformly on this section. These findings are not surprising given that many of the aphasic patients performed poorly on the GSNSB as a whole — the left hemisphere lesion group being the only group in the study to perform poorly on the majority of the sections. In addition, roughly 54 percent of the

variation in the participants' performances on this section could be explained by the site of their lesion.

The results from the Language Function section revealed that this section was able to differentiate between a left hemisphere and a hippocampal lesion, and between a left hemisphere and a frontal lesion, with roughly 68 percent of the variation in the participants' performances on this section being accounted for by the site of their lesion. Pleasingly, the frontal, right hemisphere and hippocampal lesion groups did not perform significantly differently from one another on this section.

The results from the Spatial Cognition section revealed that the GSNSB was able to differentiate between both a right hemisphere lesion and a hippocampal lesion, and between a right hemisphere lesion and a frontal lesion, with roughly 50 percent of the variation seen in the participants' performances on this section being explained by the site of their lesion.

There was also a significant difference between how the left hemisphere lesion and frontal lesion patients performed on this section; again, given the left hemisphere patients' overall poor performance throughout the GSNSB, this result is not surprising.

Finally, the results from the Executive Function section demonstrated that this section was able to differentiate between a frontal and a right hemisphere lesion, but not between a frontal and a left hemisphere lesion. In total, roughly 48 percent of the variation seen in the participants' performance on this section could be accounted for by their lesion-site.

Pleasingly, the left hemisphere, right hemisphere and hippocampal lesion groups did not perform significantly differently from one another on this section.

In summary, these findings demonstrate that the GSNSB was able to differentiate between all of the lesion-sites examined as part of this secondary analysis, with the exception of left hemisphere lesions in the case of the Memory, Spatial Cognition and Executive Function sections. Again, this finding is not surprising when seen in the light of the poor performances of the left hemisphere lesion patients throughout the GSNSB. Such performances meant that the left hemisphere patients were failing on the majority of the memory, spatial and executive tests, which made it impossible to discriminate at a statistical level between their performances and those of the patients with the primary amnesic and frontal deficits. Such a differentiation can therefore only be made on qualitative grounds, when one factors in how the patient approached the test and *how* he/she failed. This finding underscores the importance of including the decision-tree approach in the GSNSB.

The investigation as to whether the participants' performance on the 'Orientation' section of the GSNSB in any way covaried with their performances on any of the GSNSB's other four core sections revealed that Orientation score most strongly predicted performance on the Memory Function section. Here, a strong correlation coefficient of 0.87 was found. This finding is pleasing, given that a patient who is amnesic is often found not to be fully orientated to person, place and time. A number of studies investigating the relationship between amnesia and orientation have reached this conclusion (Schnider, von Daniken & Gutbrod, 1996; Small, Viitanen & Bacekman, 1997; Sweet, Suchy, Leahy, Abramowitz, & Nowinski, 1999).

The other results revealed that there was a strong relationship between Orientation score and Language Function section score, with a correlation coefficient of 0.74. The relationship between Orientation score and Spatial Cognition and Executive Function section scores was

less significant, with correlation coefficients of 0.42 and 0.59 respectively. Altogether, these findings meant that, generally, the higher a patient's performance on one of the GSNSB's sections, the higher his/her Orientation score was likely to be. Once more, the high correlation between the Orientation section and the Language Function section is also predicted, given the overall poor performances of the left hemisphere lesion patients on the GSNSB. In other words, the left hemisphere lesion patients performed poorly on both the Orientation and Language Function sections because, by and large, they were failing any test given to them as they struggled to either understand the instructions or to respond verbally. Consequently, if their aphasia was relatively mild, their performance on the Orientation section simply improved in proportion to their ability to pass the aphasia tests given to them.

Finally, the relatively strong positive correlation between the Orientation section and the Executive Function section is also not surprising, as executive dysfunction affects neurocognitive function in a global way; all tests are likely to be affected to a varying extent, depending on the severity of the impairment. Again, similar findings are present in the literature, where Brookes (1976) and Zenicus, Wesolowski and Rodriguez (1998), for example, highlight that patients often present as disoriented in the context of executive impairment. In addition to this, if the patient has frontal 'amnesia', he/she will struggle with the orientation task, given that no prompting/cueing is provided.

Conclusions

Overall, the initial demonstration of the reliability and validity of the newly developed GSNSB as a screening tool has proven a success. Crucially, the GSNSB has been shown to possess both good reliability and validity. Sound inter-rater and test-retest reliability have been demonstrated. In terms of its validity, the GSNSB possesses an ability to differentiate

between frontal versus hippocampal lesions, and between left hemisphere lesions versus right hemisphere lesions. The ability to make these differentiations is the benchmark in research practice in terms of demonstrating validity. At this juncture, it must be remembered that the GSNSB is new and, following on from the findings, recommendations and limitations of this doctoral study, further development is still required before the GSNSB can be published and utilised in the context where it is so greatly needed. Despite this, significant progress towards this goal has now formally been achieved.

This phase of the research has been invaluable in identifying key areas for future development. Due to the magnitude of the task of effectively addressing the issues of 'culture', 'language' and 'education', no one single study can solve every issue. Nonetheless, this study has been singularly successful in significantly reducing the influence of these variables on neurocognitive test performance. With reference to cross-cultural research, Van der Vijver and Leung (1997) emphasise that the key to validity is to obtain equivalent measures. "For measures to be equivalent, individuals with the same or similar standing on a construct ... but belonging to different groups, such as Xhosa- and Afrikaans- speaking, should obtain the same or similar scores on the different language versions of the items or measure" (Kanjee, 2005, p. 59). In this light, the findings of this study are most pleasing in that not only has the validity of the GSNSB been demonstrated with respect to clinical lesions but, crucially, it has also been shown that, given that all participants were assessed in their first language, the GSNSB is valid across cultural groups. This provides an endorsement of both the test adaptations and the translation work.

Future studies are required to investigate in more detail the effect of age on neurocognitive testing in the South African context. Due to the relatively small sample sizes adopted in this

study, it is recommended that further studies be conducted using a sample comprising a wider range of ages. So, too, the effects of ‘level’ and ‘quality’ of education on test outcome merit further investigation, focusing in more detail on the effects of various levels of education, rather than simply high school versus junior school. The deprivation of education that many South Africans have experienced makes this a very poignant and relevant topic — one that neuropsychological investigation in South Africa needs to embrace.

Another area for future research is a more detailed appraisal of the phenomenon of ‘acculturation’ and the impact this has on neurocognitive tests. Here, the present study has highlighted that the extreme cultural and linguistic diversity in South Africa has resulted in a population that is far from homogenous. Consequently, the effects that an urban versus a rural upbringing might have on an individual are diverse and complex (Foxcroft & Roodt; 2005; Grieve, 2005; Nell, 2000; Shuttleworth-Jordan, 1996; Uzzell, 2007) — as is exposure to western medicine and practice. Further studies are recommended to examine the effect that level of acculturation has on test performance by investigating closely the combination of upbringing, region of origin, formal education and relative exposure to urbanisation.

Recommendations for the further development of some of the newly developed neurocognitive tests, and the accompanying instructions and decision-trees, have also been made. Before the GSNSB can be used clinically, a number of important modifications need to be made, and further researched and validated. For example, the Spatial Acalculia Test requires additional spatially loaded sums, which are simpler than the current examples. The current examples can be retained, but these should either be optional, ‘more difficult’ tests, or should carry less weighting in the overall scoring procedure. In addition, simple, non-spatially loaded addition and subtraction sums (such as ‘ $2 + 4$ ’ and ‘ $7 - 3$ ’) need to be

introduced as initial tests, preceding the spatially loaded items, so that primary acalculia can first be excluded. To accurately achieve this goal, scores should be given for the patient's approach to the task, as well as for obtaining the correct answer. This would also allow for the educational ability of the patient to be examined, before he/she is asked to tackle the more complex tasks.

Another pertinent finding is the need to incorporate multiple-choice questions into the GSNSB's decision-tree in the 'Mesial' section, in order to discriminate between a memory 'encoding' versus a memory 'retrieval' problem, as the latter requires examining whether the patient can benefit from prompting. The results of the validation study also found that the '18 Book Problem' was a poor discriminator of executive function among the poorly educated sample, as was evidenced by the poor performance of the control participants on this task. Consequently, an alternative pre-frontal test is required.

CHAPTER SEVEN: THE IMPLEMENTATION OF POST- VALIDATION CHANGES TO THE GROOTE SCHUUR NEUROCOGNITIVE SCREENING BATTERY

A comprehensive test battery contains measures of both higher and lower cognitive domains in order to identify the point of processing at which functions break down. In addition, the clinician must assemble a test battery that permits assessment of the same cognitive domain with multiple measures to explore the reliability of the deficit.

(Hebben & Milberg, 2002, p. 91)

The importance of using effective decision-trees to guide the qualitative assessment process is well established in the literature, and has been outlined in previous chapters with reference to the work of various authors (Christensen, 1979; Christensen & Caetano, 1999; Glozman, 1999; Hebben & Milberg, 2002; Luria, 1966). The refinement of the decision-trees — critical to the success of the qualitative approach, and central to the design of the GSNSB in ensuring that multiple determinants of test failure are eliminated — is now outlined in the following chapter.

Rationale

This chapter outlines the post-validation changes made to the Groote Schuur Neurocognitive Screening Battery (GSNSB), and explains why these modifications were made, based on the results and qualitative findings of the validation. All post-validation changes made to the scoring instructions, decision-trees, and overall scoring procedure are highlighted in ‘red’ in the Post-validation version of the GSNSB (see Appendix E). These changes were informed

by three sources: the qualitative results of the study (from the qualitative allocation process), the quantitative findings, and the feedback from experiences in administering the GSNSB.

These changes were the final contribution of the present study. It is acknowledged that further research and development will be needed on the GSNSB before its eventual publication but, at the same time, it is hoped that, overall, this contribution will provide a clear and meaningful framework for future research. The purpose of implementing these changes here was to consolidate the work already done by this study, while at the same time providing a logical departure point for the next phase of development.

Changes Made to the Memory Function Section

The post-validation changes to the Memory Function section involved the Auditory Span Test, the Township Fire Story and the Rey Complex Figure. The results of the Auditory Span Test indicated that many of the participants had struggled to perform well, with eight of the control participants failing to register full marks, and three of these scoring zero. Upon investigation, this finding stemmed from the fact that only one trial per number sequence was provided. From the neuropsychologists' clinical experiences in administering the Digit Span Test, it is clear that it often takes many trials before one is able to demonstrate that the patient is in fact able to get the sequence correct. Coupled with this, a digit span of 'six' is considered by the neuropsychologists to be clinically 'normal' in the South African context, whereas elsewhere 'seven' is usually viewed as normal (Joynt & Shoulson; Solms & Turnbull, 2002).

Given the findings from the validation study that the control sample was averaging '6' digits rather than '7', that three of the controls had failed the test, and five had scored only '1 out of

2' (that is, a digit span of '5') — coupled with the neuropsychologists' established clinical practice — the Auditory Span Test's cut-off scores were adjusted. Whereas a digit span of '7' was previously required to score '2 out of 2', now '6 or 7' digits correctly achieved receives full marks, while '5' digits now scores '1 out of 2', and '4' digits or less scores '0'. Previously, '6' digits had scored '1 out of 2' and '5' digits '0 out of 2' (see Appendix E for all changes made).

The number of trials per sequence was also increased to two trials each for the sequences up to four digits, and three trials for the sequences containing five or more digits. The latter sequences were given three trials each as opposed to two because they are more complex than the first four trials, as it was observed that the majority of participants began to struggle around the five-digit sequence. These changes provide the patient much more opportunity to pass the test, while at the same time affording the assessor a more accurate account of the patient's working memory ability.

Changes were also made to the Auditory Span Test's administration instructions. These involved describing the increasing number of test trials, and informing the assessor in more detail than previously provided how to administer the test — instructing him/her to pause briefly between numbers, and to ensure that the patient does not repeat the sequence until the assessor has finished saying it.

Extensive changes were also made to the Township Fire Story, a test designated as an 'optional test' in the Memory Function section and therefore not part of the GSNSB's core scoring procedure for this particular section. Here, it was found that both the scoring instructions for the test and its decision-trees required far greater elaboration. These details

could not be provided in the main body of the GSNSB due to issues concerning space and the overall design structure, and were thus added to the relevant appendix instead. These changes to the instructions, as well as to the decision-tree procedure, were necessitated by two key findings. The first of these was the need to include multiple-choice questions in the decision-tree procedure, necessitating substantial additions to the decision-tree (see ‘Changes to the Executive Function section’ below). Secondly, the instructions in the GSNSB Prototype were deemed to be insufficiently detailed. For example, details such as the instruction not to prompt the patient after the first recall trial had been omitted.

To solve these two issues, four sequential ‘steps’, each with its own decision-tree procedure (and part of the overall decision-tree procedure for this test), were created for the administration procedure (see Appendix E). The basic administration procedure was first spelled out in far greater detail for ‘Step 1’. Next, for ‘Step 2’, a breakdown of the test’s tentative cut-off scores was provided to enable the assessor to decide whether the patient is a possible amnesic or not using the decision-tree. This specific decision-tree was designed in order that the patient’s performance on the Township Fire Story could be viewed in relation to his/her performance on other pertinent sections of the GSNSB, to provide converging lines of evidence. ‘Step 3’ was designed to facilitate in the decision of whether the patient has an ‘encoding’ or a ‘retrieval’ problem, while ‘Step 4’ was specifically created to help score the newly included multiple-choice questions in Section 5’s ‘Mesial’ subsection. Overall, this decision-tree allows the assessor to use the Township Fire Story in a flexible, coherent manner.

As with the Township Fire Story, all the administration and scoring instructions for the Rey Complex Figure, specific to this Memory Function section, were also moved to the

appendices and elaborated on substantially (see Appendix E). Vital instructions that had initially been excluded (such as, for example, the need to ensure that the patient is not told that it is a memory test), were included and elaborated on.

Finally, minor changes were made to the ‘normal’ scores displayed in the scoring blocks — these scores were changed to those achieved on average by the control participants. It is important to note that the changes made to these normal scores were tentative, and are only intended to serve as rough guidelines — as an initial starting point to build on for future research. This is especially valid at this early stage of the GSNSB’s development when relatively few participants have been assessed. This also applies to all the ‘normal’ scores adjusted throughout the GSNSB, as mentioned in the sections to follow.

Changes Made to the Language Function Section

Given the overall success of the initial Language Function section results, only minor changes were made (see Appendix E). It was observed that many of the participants could not name the watch ‘winder’ in the ‘bedside naming’ subsection of the ‘Naming’ subsection. Therefore, this item was replaced with the word ‘mattress’, which was deemed to be more familiar and appropriate for the assessment of basic naming ability, especially in the hospital ward, where a large number of assessments will take place. One other specific change to this section was to the instructions for the assessment of ‘Naming’. The instruction that it must first be ensured that the patient’s vision is intact before beginning the test was added to the decision-tree. The only other changes made to this section were to adjust the ‘normal’ scores in the decision-tree scoring boxes according to how the controls had performed on average, along with a minor adjustment to the Naming Test referring to where its administration procedure and scores could be found in the appendices.

Changes Made to the Spatial Cognition Section

The major findings relating to the Spatial Cognition section necessitated changes to the Rey Complex Figure, the 3-D Analysis Test and the Spatial Acalculia Test. The results of the validation study had shown that the Rey Complex Figure was a more reliable measure of the participants' spatial cognition and perception ability than the 3-D Analysis Test, mainly as a consequence of the high error rates from the participants (including the controls) on the 3-D Analysis Test. Therefore, it was decided to give the scores for the Rey Complex Figure more weighting as a part of the 'Visuospatial Analysis' subsection. The changes made were to double the score allocation for the copying of the Figure. Whereas previously the scores had been '2 out of 2', '1 out of 2' and '0 out of 2' for a pass, a defective attempt and a wholly inadequate attempt respectively, the scores were now changed to '6', '3' and '0' respectively (see Appendix E). This makes the Rey Complex Figure worth more of the total score for the 'Visuospatial Analysis' subsection.

The results of the 3-D Analysis Test had shown that overall, although the test could differentiate between controls and right middle cerebral artery (MCA) patients, the patient sample as a whole performed poorly, despite the fact that the majority of the control participants had fared well. This finding seemed indicative of a possible floor-effect. The best solution to this problem in terms of the GSNSB's further development will be to create and include an alternative test, rendering the 3-D Analysis Test an 'optional' test, for use on individuals with nine or more years of education. Tests of cognitive spatial ability are widely acknowledged to be problematic when used on individuals with low levels of education (Lezak, 1995).

Given that the creation of an alternative test was not feasible at this late point in the current research, the cut-off scores of the 3-D Analysis Test were in the mean time lowered, in an attempt to compensate for the high patient failure rate. Now, it is only required to get anywhere between six and ten items correct to achieve a full score of ‘2 out of 2’ (see Appendix E). This reduction in the scores, in conjunction with increasing the relative weighting of the scores for the other tests in this section, will hopefully reduce the impact of the floor-effect on the overall performance for this subsection.

The results of the original Spatial Acalculia Test showed that it was too difficult a test — the majority of controls performed poorly — and that it also required additional items, to be included before the spatially-loaded written sums, in order to exclude primary acalculia and to ensure upfront that the patient has basic numeracy. Consequently, extensive changes were made to this test (see Appendix E). Firstly, simple sums, which do not involve spatial-loading, were added. These sums are not scored because they serve only to answer a ‘yes/no’ question in the decision-tree as to whether a primary acalculia is present and whether the patient has basic numeracy. Instructions were added to advise the assessor to discontinue the testing at this point if the patient fails these initial four sums.

Once these questions had been included as the first step in the decision-tree, a second series of four spatially-loaded sums was created. These four sums are intended to be simpler than the three sums in the GSNSB Prototype, but at the same time require the patient to calculate across decimal places in order to introduce the spatial-loading component to the task. To further introduce a progression of complexity to this testing procedure, the assessor is instructed to administer the first two of these sums verbally, whereas the latter two are to be written by the patient. These sums are scored out of four, with one point for each correct

answer. In addition to this, a qualitative scoring procedure was added to the decision-tree. Here, the patient is scored on the latter two written sums through the assessor's qualitative observation of whether the patient has been able to correctly position the numbers in relation to the sum. This observation is scored out of one.

The third and final step created for the decision-tree of the Spatial Acalculia Test took the same form as the initial three sums constituting the original test. However, the three sums were simplified slightly but are still scored out of three — one point for each correct answer. In addition to this, a qualitative scoring procedure was added so that the assessor can score how the patient spatially positioned the numbers vertically and horizontally, and whether he/she rearranged any of the numbers erroneously. The addition of this qualitative scoring procedure both here and in the previous step allows the patient to be scored on his/her spatial abilities, thereby being compensated with some points even if he/she is unable reach the final answer correctly.

The final changes made to the Spatial Cognition section of this Post-validation version of the GSNSB were to modify the overall scoring summary. Here, because the Rey Complex Figure's scores had been inflated, and additional scores for the Spatial Acalculia Test had been added, the total score for this section of the GSNSB was increased from 15 to 23. Part of this process also involved changing the relevant subsection scores and subtotals throughout this section of the GSNSB, as the 'Visuospatial Analysis' subsection now had a different total and therefore a different weighting in the overall Spatial Cognition section. Finally, as none of the initial 'normal' scores required adjustment, the ones originally provided were retained.

Changes Made to the Executive Function Section

The findings from the FAS Test and its isiXhosa and Afrikaans equivalents, the NPS and BHP Tests, were that in general the neurocognitively intact control participants performed well, while at the same time the test was able to successfully differentiate between frontal lesion patients and control participants, but not between frontal lesion patients and hippocampal lesion patients. Despite these positive findings, and the finding that level of education did not influence performance on this test, it was decided nonetheless to also include a simpler, equivalent test, as an ‘optional’ test in this section (see Appendix E). The primary reason for this particular change to the GSNSB was that many South Africans have experienced a poor quality of education, and the available literature suggests that education significantly influences Controlled Oral Word Association Test (COWAT) performance (Lezak, 2004; Loonstra, Tarlow & Sellers, 2001; Spreen & Strauss, 1991; Tombaugh, Kozak & Rees, 1999).

Where there is some concern regarding the level/quality of education that the particular patient has received, this alternative to the FAS/BHP/NPS Test can now be administered. This simple task is not scored, as is the case with the other ‘optional’ tests contained in the GSNSB. The task requires the patient to name as many animals as he/she can in one minute; this has been found to be a less challenging test for poorly educated, illiterate or semiliterate individuals, as has been confirmed in studies previously undertaken internationally and in the South African context (Lezak, 1995; Nell, 2000; Tombaugh et al., 1999). The animals’ names can begin with any letter, rendering the task a simpler proposition than the FAS Test.

The results from the validation study for the Township Fire Story (as previously mentioned in the ‘Changes made to the Memory Function section’), provided valuable insight into the need to include multiple-choice questions in the ‘Mesial’ subsection. This would help to differentiate between a memory ‘encoding’ and a ‘retrieval’ problem, the latter resulting from executive dysfunction. Steps ‘3’ and ‘4’, as previously mentioned, were specifically included in the decision-tree for this purpose (see Appendix E). Given that the Township Fire Story was already included in the appendices, it was decided, due to its length and the detailed instructions, to retain it as an appendix for the Executive Function section, and to refer the assessor to the relevant appendix when testing the ‘Mesial’ subsection.

Therefore, the changes made to the ‘Mesial’ subsection involved the addition of instructions referring to the appendices and to the Memory Function section (for definitions of the types of ‘amnesias’), as well as the inclusion of four multiple-choice questions. These questions were made to score out of four, with one point for each correct answer. Consequently, the overall scores for this section were also increased to accommodate the questions — the ‘Mesial’ section is now worth six rather than two points — and the overall total score for the Executive Function section was increased from 11 to 16 points (one extra point also being added to the Fist/Side/Palm Test, see below).

When the Township Fire Story is administered as part of the Memory Function section, the decision-tree steps now allow the assessment of a patient to end prior the final Step if at the end of ‘Step 2’ the patient is deemed not to be amnesic. However, when the test is administered as part of the Executive Function section, all four steps must be completed. This is because the key components of memory retrieval are only investigated at steps ‘3’ and ‘4’ of the overall decision-tree for this particular test. This point is made clear to assessors as part

of the changes made to the administration instructions of the ‘Mesial’ section, reminding them to “complete ALL FOUR STEPS” in the decision-tree.

It was also necessary to include a substantial amount of technical detail when these changes to the ‘Mesial’ subsection were made. Specifically, the instructions for steps ‘3’ and ‘4’, and the scoring procedure added for the latter step, required careful planning and elaboration. The scoring procedure is reversed when a patient is identified in ‘Step 3’ of the decision-tree as having a clear memory problem while at the same time benefitting from prompting. The reason for the need to reverse the patient’s score in this instance is because benefitting from prompting in the context of a memory difficulty (if demonstrated at steps ‘1’ and ‘2’) is indicative of executive impairment. Therefore, a score of ‘4 out of 4’ in this context actually represents the presence of a deficit and thus requires reversal (see Appendix E for the full administration and scoring procedure). Finally, the instructions to the questions eliciting confabulation were also slightly modified in the light of having to accommodate the new multiple-choice questions.

Another change to the Executive Function section was to the Fist/Side/Palm Test, where it was decided to increase the score weighting of the test from one point to two (see Appendix E). This was necessary because this particular test is an excellent determinant of executive function and one of the more robust indicators of executive impairment (Christensen, 1979; Luria, 1973). It was therefore decided that it should be worth more points relative to the other tests in this subsection — the Tapping Rhythm Test and the Repeated Pattern Drawing Test — which are only worth one point each. This change required only a slight modification to the scoring instructions for the test, resulting in the criteria for scoring full marks, as opposed to ‘1 out of 2’ and ‘0 out of 2’, to be modified.

The final post-validation change made to the Executive Function section was the modification of the '18 Book Problem'. The validation study results had shown that this test was too difficult for the majority of the participants (both patients and controls) and that when a good test performance had been produced, it was highly dependent on level of education. In total, eight of the neurocognitively intact controls failed this test. In light of these findings, it was decided to make the 18 Book Problem an 'optional' test in this section, while at the same time including a simpler problem. The simpler problem was named the '6 Apples Problem' and is scored using the pre-existing scoring procedure from the original 18 Book Problem (see Appendix E). The decision-tree for this task was redesigned accordingly so that the assessor is instructed to decide which of the two tests to administer based on the patient's level of education. Patients with low levels of education should be given the '6 Apples Problem'. If the 18 Book Problem is used, it is scored in place of the '6 Apples Problem', using the scoring procedure that has been modified in terms of its instructions to accommodate both tests.

Changes Made to the Summary of Scores Section

The 'Total' score for the GSNSB, given the additional scores that have now been added and the reweighting of certain scores, has been changed to total exactly 100 as opposed to the original 85 (see Appendix E). This increased total resulted from the Spatial Cognition section now being scored out of 25 (instead of the original 15 points), and the Executive Function section now being scored out of 16 (instead of the original 11 points). These score modifications provide a 'balance' to the overall GSNSB, as all four primary sections (excluding the Orientation section) now have more evenly matched total scores. The total scores of the Memory Function and Language Function sections have not changed. Finally,

the overall ‘normal’ scores for the Orientation, Memory Function and Language Function sections were adjusted (from 8 to 10, 14 to 17, and 26 to 29 respectively) to conform with the averages achieved by the controls on these sections. New equivalent ‘normal’ scores for the Spatial Cognition and Executive Function sections could not yet be provided, as the total scores for these two sections were changed post-validation.

Conclusion

In summary, it is hoped that the post-validation changes to the GSNSB have served to effectively consolidate the work already done in this study, while at the same time forming a firm foundation for future developmental work on this tool. The changes made naturally require further validation. The avenues identified by this study for further exploration must also be addressed before the GSNSB can be published and utilised. The need for this tool in the South African context — given the present dearth of neuropsychological resources, and the limitations of the MMSE — cannot be overestimated.

At this final junction it must also be reiterated that the GSNSB has been designed solely as a screening tool for use in the South African context. This tool is not intended to substitute the role of the clinician, but rather to serve as a screening tool to provide initial insights into a patient’s neurocognitive deficits so that he/she can ultimately be referred to the appropriate specialist.

Given the lack of neuropsychological expertise in South Africa, it is hoped that this tool will eventually serve as ‘transferable technology’ in providing sufficient insight into neurocognitive dysfunction that some meaningful contribution to the patient’s treatment and care can still be made in the absence of a specialist to refer to. The need for ‘transferable

technology' of this sort in South Africa is an urgent reality (Nell, 2000). It is also hoped that in this role the GSNSB will serve as an educational and training tool in providing the South African medical community in general with basic knowledge of neuropsychological functioning.

Finally, if the newly designed neurocognitive tests are researched and developed further, it is envisaged that they might one day be used individually by the neuropsychologist in the clinical setting. It was, after all, in this very context that the clinical insights playing such an important role in the development of these tests were gained.

University of Cape Town

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Anderson, S. J. (2001). On the importance of collecting local neuropsychological normative data. *South African Journal of Psychology*, *31*, 29-34.

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 143-150.

Ardila, A. (2007). Toward the development of a cross-linguistic naming test. *Archives of Clinical Neuropsychology*, *22*, 297-307.

Ardila, A. (2007). The impact of culture on neuropsychological test performance. In B. Uzzell, M. Ponton & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology*. (pp. 23-44). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Ardila, A., & Keating, K. (2007). Cognitive abilities in different cultural contexts. In *International handbook of cross-cultural neuropsychology*. (pp. 109-125). Mahwah: Lawrence Erlbaum Associates, Publishers.

Ardila, A., Lopez, M. V., & Solano, E. (1989). Semantic aphasia reconsidered. In A. Ardila & F. Ostrosky-Solis (Eds.), *Brain organization of language and cognitive processes*. (pp. 177-193). New York: Plenum Press.

Ardila, A., & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology, 5*, 307-320.

Ardila, A., Rosselli, M., & Rosas, P. (1989). Neuropsychological assessment of illiterates: Visuospatial and memory abilities. *Brain and Cognition, 11*, 147-166.

Artiola i Fortuny, L., & Mullaney, H. (1997). Neuropsychology with Spanish-speakers: Language use and proficiency issues for test development. *Journal of Clinical and Experimental Neuropsychology, 19*, 615-622.

Artiola i Fortuny, L., & Mullaney, H. A. (1998). Assessing patients whose language you do not know: Can the absurd be ethical? *The Clinical Neuropsychologist, 12*, 113-126.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 14*, 575-589.

Banos, J. H., & Franklin, L. M. (2002). Factor structure of the Mini-Mental State Examination in adult psychiatric patients. *Psychological Assessment, 14*, 397-400.

Barker-Collo, S. L. (2001). The 60-item Boston Naming Test: Cultural bias and possible adaptations for New Zealand. *Aphasiology, 15*, 85-92.

Benson, D. F. (1979). Aphasia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 17-47). New York: Oxford University Press.

Berry, D. T. R., Allen, R. S., & Schmitt, F. A. (1991). Rey-Osterrieth Complex Figure: Psychometric characteristics in a geriatric sample. *Clinical Neuropsychologist*, 5, 143-153.

Berry, J. W. (1965). Temne and Eskimo perceptual skills. *International Journal of Psychology*, 1, 207-229.

Betancourt, H., & Lopez, S. R. (1993). The study of culture, ethnicity, and race in American psychology. *American Psychologist*, 48, 629-637.

Black, F. W., & Strub, R. L. (1994). The bedside and office mental status examination. In D. Touyz, D. Bryne & A. Gilandas (Eds.), *Neuropsychology in clinical practice* (pp. 38-60). Sydney: Academic Press, Inc. Harcourt Brace & Company, Publishers.

Bleiberg, J., Garmoe, W. S., & Halpern, E. L. (1997). Consistency of within-day and across-day performance after mild brain injury. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 10, 247-253.

Blessed, G., Tomlinson, B. E., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects. *British Journal of Psychiatry*, 114, 797-811.

Bradshaw, D., Nannan, N., Laubscher, R., Groenewald, P., Joubert, J., Nojilana, B., et al. (2000). *South African national burden of disease study: Western Cape Province.*: South African Medical Research Council: Burden of Disease Research Unit.

Brauer Boonea, K., Victor, T. L., Wen, J., Razani, J., & Pont'on, M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, 22, 355-365.

Bravo, G., & Hebbert, R. (1997). Age- and education- specific reference values for the Mini-Mental State Examination derived from a non-demented elderly population. *International Journal of Geriatric Psychiatry*, 12, 1008-1018.

Brickman, A. M., Cabo, R., & Manly, J. J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology*, 13, 91-100.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & W. Berry (Eds.), *Field-methods in cross-cultural research*. (pp. 137-164). Berkeley: Sage Publications.

Brooks, D. N. (1976). Wechsler Memory Scale performance and its relationship to brain damage after severe closed head injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, 39, 593-601.

Butters, N., & Miliotis, P. (1979). Amnesic disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 403-451). New York Oxford University Press.

Cahn, D. A., Salmon, D. P., Butters, N., Wiederholt, W. C., Corey-Bloom, J., & Edelstein, S. L. (1995). Detection of dementia of the Alzheimer type in a population-based sample:

Neuropsychological test performance. *Journal of the International Neuropsychological Society*, 1, 252-260.

Chincotta, D., & Underwood, G. (1997). Speech rate estimates, language of schooling and bilingual digit span. *European Journal of Cognitive Psychology*, 9, 325-348.

Christensen, A. L. (1979). *Luria's Neuropsychological Investigation*. Copenhagen: Munksgaard.

Christensen, A. L., & Caetano, C. (1999). Luria's Neuropsychological Evaluation in the Nordic countries. *Neuropsychology Review*, 9, 71-78.

Connor, M., & Bryer, A. (2006). Stroke in South Africa. In K. Steyn, J. Fourie & N. Temple (Eds.), *Chronic Diseases of Lifestyle in South Africa since 1995 - 2005* (pp. 195 - 203): Medical Research Council.

Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 121-140). Chichester: John Wiley & Sons Ltd.

Cruice, M. N., Worrall, L. E., & Hickson, L. M. H. (2000). Boston Naming Test results for healthy older Australians: A longitudinal and cross-sectional study. *Aphasiology*, 14, 143-155.

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *The Journal of the American Medical Association*, 269, 2386-2391.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry*, 78, 790-799.

da Costa Pinto, A. (1991). Reading rates and digit span in bilinguals: the superiority of mother tongue. *International Journal of Psychology*, 26, 471-483.

Damasio, A. (1979). The frontal lobes. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology* (pp. 339-376). New York: Oxford University Press.

Delis, D. C., & Bihrlle, A. M. (1989). Fractionation of spatial cognition following focal and diffuse brain damage. In A. Ardila & F. Ostrosky-Solis (Eds.). New York: Plenum Press.

Delouche, G., Hannequin, D., Dordain, M., & Perrier, D. (1996). Picture confrontation oral naming: Performance differences between aphasics and normals. *Brain and Language*, 53, 105-120.

Department of Health. (2003). *Annual Report 2002/2003: South Africa*. Retrieved January 17, 2005, from <http://www.info.gov.za/annualreport/2003/statssa03.pdf>

Devinsky, O. (1992). *Behavioral neuropsychology: 100 Maxims in neurology*. New York: Edward Arnold.

Devinsky, O., & D'Esposito, M. (2003). *Neurology of cognitive and behavioral disorders*. New York: Oxford University Press.

Durrheim, K. (1999). Quantitative measurement. In M. Terre Blanche & K. Durrheim (Eds.), *Research in practice: Applied methods for the social sciences*. Cape Town: University of Cape Town Press Ltd.

Edwards, D. J. A., & Louw, D. A. (1997). Psychology as a science and a profession. In D. A. Louw & D. J. A. Edwards (Eds.), *Psychology: An introduction for students in Southern Africa* (2nd ed., pp. 1-59). Johannesburg: Heinemann.

Espino, D. V., Lichtenstein, M. J., Palmer, R. F., & Hazuda, H. P. (2001). Ethnic differences in Mini-Mental State Examination scores: Where you live makes a difference. *Journal of the American Geriatrics Society*, 49, 538-548.

Evans, J. J. (2004). Disorders of memory. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians*. (pp. 143-163). Chichester: John Wiley & Sons Ltd.

Fan, J. Z., Lezak, M. D., Yuan, G., G., & Hu, C. H. (1988). A comparison of the sensitivity of different techniques for eliciting visuospatial neglect. *Journal of Clinical and Experimental Neuropsychology*, 10, (Abstract).

Feher, E. P., Mahurin, R. K., Doody, R. S., & Cooke, N. (1992). Establishing the limits of the Mini-Mental State: Examination of "subtests." *Archives of Neurology*, *49*, 87-92.

Feher, E. P., & Martin, R. C. (1992). Cognitive assessment of long-term memory disorders. In D. I. Margolin (Ed.), *Cognitive neuropsychology in clinical practice* (pp. 168-202). New York: Oxford University Press.

Fisher, N. J., Tierney, M. C., Snow, W. G., & Szalai, J. P. (1990). Odd/even short forms of the Boston Naming Test: Preliminary geriatric norms. *Clinical Neuropsychologist*, *13*, 359-364.

FitzGerald, M. J. T., & Folan-Curran, J. (2002). *Clinical neuroanatomy and related neuroscience* (4th ed.). Edinburgh: W.B. Saunders.

Fletcher-Janzen, E., Strickland, T. L., & Reynolds, C. (2000). Preface. In E. Fletcher-Janzen, T. L. Strickland & C. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology*. (pp. 3-7). New York: Kluwer Academic.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.

Foxcroft, C., & Roodt, G. (2005). What the future holds for psychological assessment. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 242-257). Oxford: Oxford University Press.

Fritz, V. U. (1997). Stroke incidence in South Africa. *South African Medical Journal*, 87(No.5), 584 - 585.

Garb, H. N. (1998). *Studying the clinician: Judgement research and psychological assessment*. Washington, DC: American Psychological Association.

Garcia-Morales, P., Gich-Fulla, J., Guardia-Olmos, J., & Pena-Casanova, J. (1998). Digit span, automatic speech and orientation: Amplified norms of the Barcelona Test. *Neurologia*, 13, 271-276.

Gilbert, L., & Tollman, S. (2007). Epidemiological, social, and cultural aspects of illness: A case study of brain injuries, stroke, and HIV/AIDS in South Africa. In B. P. Uzzell, M. O. Ponton & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology*. (pp. 319-340). Mahwah: Lawrence Erlbaum Associates, Publishers.

Glozman, J. M. (1999). Quantitative and qualitative integration of Lurian procedures. *Neuropsychology Review*, 9, 23-32.

Glozman, J. M., & Tupper, D. E. (1995). Converging impressions in Russian and American neuropsychology: discussion of a clinical case. *Applied Neuropsychology*, 2, 15-23.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The Boston Diagnostic Aphasia Examination*. Baltimore: Lippincott, Williams & Wilkins.

Griessel, L. (2005). Administering psychological assessment measures. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 83-98). Oxford: Oxford University Press.

Grieve, K. W. (2005). Factors affecting assessment results. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 224-241). Oxford: Oxford University Press.

Gurland, B. J., Wilder, D. E., Teresi, J., & Barrett, V. W. (1992). Screening Scales for Dementia: Toward reconciliation of conflicting cross-cultural finding. *International Journal of Geriatric Psychiatry*, 7, 105-113.

Halstead, W. C., & Wepman, J. M. (1959). The Halsted-Wepman Aphasia Screening Test. *Journal of Speech and Hearing Disorders*, 14, 9-15.

Haralambos, M., & Holborn, M. (1991). *Sociology: Themes and perspectives*. (3rd ed.). London: Collins Educational.

Harris, M. (1983). *Culture, people, nature: An introduction to general anthropology*. New York: Harper and Row.

Harvey, R. J. (2003). Epidemiology of presenile dementia. In J. R. Hodges (Ed.), *Early-onset dementia* (pp. 1-23). Oxford: Oxford University Press.

Hawkins, K. A., & Bender, S. (2002). Norms and the relationship of Boston Naming Test performance to vocabulary and education: A review. *Aphasiology*, *16*, 1143-1153.

Heart and Stroke Foundation of South Africa. (2007). *Heart disease in South Africa*.

Retrieved February 17, 2006, from <http://www.heartfoundation.co.za/docs/heartmonth/HeartDiseaseinSA.pdf>

Heaton, R. K., Avitable, N., Grant, I., & Matthews, C. G. (1999). Further cross-validation of regression-based neuropsychological norms with an update of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*, *21*, 571-582.

Hebben, N., & Milberg, W. (2002). *Essentials of neuropsychological assessment*. New York: John Wiley & Sons Inc.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1979). Neglect and related disorders. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology*. (pp. 243-293). New York: Oxford University Press.

Herman, J. (1992). *Trauma and Recovery*. New York: Basic Books.

Hicksen, J. K., S. (1991). The Mission and Role of Psychology: The Case of South Africa. *International of Psychology*, *26*(6), 783-793.

Hitch, G. H., Halliday, M. S., & Littler, J. E. (1989). Item identification time and rehearsal rate as predictors of memory span in children. *The Quarterly Journal of Experimental Psychology*, *41A*, 321-337.

Hodges, J. R., & Miller, B. (2001). Frontotemporal dementia (Pick's disease). In J. R. Hodges (Ed.), *Early-onset dementia*. (pp. 284-303). Oxford: Oxford University Press.

Horner, M. D., Teichner, G., Kortte, K. B., & Harvey, R. T. (2002). Construct validity of the Babcock Story Recall Test. *Applied Neuropsychology*, *9*, 114-116.

Howell, D. C. (1999). *Fundamental statistics for the behavioral sciences*. (4th ed.). Pacific Grove: Duxbury.

Huppert, F. A., Brayne, C., Gill, C., Paykel, E. S., & Beardsall, L. (1995). CAMCOG - a concise neuropsychological test to assist dementia diagnosis: socio-demographic determinants in an elderly population sample. *British Journal of Clinical Psychology*, *34*, 529-541.

Jacobs, D. M., Sano, M., Dooneief, G., Marder, K., Bell, K., & Stern, Y. (1995). Neuropsychological detection and characterization of preclinical Alzheimer's disease. *Neurology*, *45*, 957-962.

Jervis, L. L., Beals, J., Fickenscher, A., & Arciniegas, D. B. (2007). Performance on the Mini-Mental State Examination and Mattis Dementia Rating Scale among American Indians. *Journal of Neuropsychiatry and Clinical Neurosciences*, *19*, 173-178.

Joyce, P. (1989). *The South African family encyclopaedia*. Cape Town: Struik.

Joynt, R. J., & Shoulson, I. (1979). Dementia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology*. (pp. 453-480). New York: Oxford University Press.

Juby, A., Tench, S., & Baker, V. (2002). The value of clock drawing in identifying executive cognitive dysfunction in people with a normal mini-mental state examination. *Canadian Medical Association Journal*, *168*, 859-864.

Kanjee, A. (2005). Cross-cultural test adaptation and translation. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 57-66). Oxford: Oxford University Press.

Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement and practice*. Washington, DC: American Psychological Association.

Kaplan, E., Fein, D., Morris, R., & Delis, D. (1991). *WAIS-R as a neuropsychological instrument*. San Antonio, TX: The Psychological Corporation.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology, 44*, 231-248.

Kennepohl, S. (1999). Toward cultural neuropsychology: An alternative view and preliminary model. *Brain and Cognition, 41*, 345-362.

Kertesz, A. (1989). Anatomical and physiological correlations and neuroimaging techniques in language disorders. In A. Ardila & F. Ostrosky-Solis (Eds.), *Brain organization of language and cognitive processes*. New York: Plenum Press.

Kiernan, R. J., Mueller, J., Langston, J. W., & VanDyke, C. (1987). The Neurobehavioural Cognitive Status Examination. *Annals of Internal Medicine, 481-485*.

Kim, H., & Na, C. L. (1999). Normative data on the Korean version of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 21*, 127-133.

Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of Human Neuropsychology* (5th ed.). Alberta: Worth Publishers.

Kotik-Friedgut, B. (2006). Development of the Lurian approach: A cultural neurolinguistic perspective. *Neuropsychology Review, 16*, 43-52.

Kraybill, M. L., Larson, E. B., Tsuang, D. W., Teri, L., McCormick, W. C., Bowen, J. D., et al. (2005). Cognitive differences in dementia patients with autopsy-verified AD, Lewy body pathology, or both. *Neurology*, *64*, 2069-2073.

Kukull, W. A., Larson, E. B., & Teri, L. (1994). The Mini-Mental State Examination score and the clinical diagnosis of dementia. *Journal of Clinical Epidemiology*, *47*, 1061-1067.

Lansing, A. E., Ivnik, R. J., Cullum, C. M., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology*, *14*, 481-487.

Lezak, M. D. (1982). The Problem of Assessing Executive Functions. *International Journal of Psychology*, *17*, 281-297.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Loewenstein, D. A., Arguelles, T., Arguelles, S., & Linn-Fuentes, P. (1994). Potential cultural bias in the neuropsychological assessment of the older adult. *Journal of Clinical and Experimental Neuropsychology*, *16*, 623-629.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove: Lawrence Erlbaum.

Loonstra, A. S., Tarlow, A. R., & Sellers, A. H. (2001). COWAT metanorms across age, education and gender. *Applied Neuropsychology*, 8, 161-166.

Lopez, M. N., Charter, R. A., Mostafavi, B., Nibut, L. P., & Smith, W. E. (2005). Psychometric properties of the Folstein Mini-Mental State Examination. *Assessment*, 12, 137-144.

Louw, J. F., D. (2004). Race and Psychology in South Africa. In A. S. Winston (Ed.), *Defining Difference: Race and Racism in the History of Psychology* (pp. 171-197). Washington D.C.: American Psychological Association.

Luiz, D., Stroud, L., & Jansen, J. (2005). Assessment of young children, physically disabled individuals, mentally challenged learners, and individuals with chronic conditions. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 101-117). Oxford: Oxford University Press.

Luria, A. R. (1966). *Higher Cortical Functions in Man*. London: Tavistock Publications.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology*. Harmondsworth: Penguin Books Ltd.

Luria, A. R. (1999). Outline for the neuropsychological examination of patients with local brain lesions. *Neuropsychology Review*, 9, 9-22.

Luria, A. R., & Majovski, L. (1977). Basic approaches used in American and Soviet clinical neuropsychology. *American Psychologist*, *32*, 959 - 968.

Mackenzie, C., Brady, M., Norrie, J., & Poedjianto, N. (2007). Picture description in neurologically normal adults: Concepts and topic coherence. *Aphasiology*, *21*, 340-354.

Manly, J. J. (2005). Book review: Handbook of cross-cultural neuropsychology. *Journal of Cross-cultural Neuropsychology*, *36*, 278-279.

Mapou, R. L. (1988). Testing to detect brain damage: An alternative to what may no longer be useful. *Journal of Clinical and Experimental Neuropsychology*, *10*, 271-278.

Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence*. New York: Oxford University Press.

McFie, J. (1961). The effect of education of African performance on a group of intellectual tests. *British Journal of Educational Psychology*, *31*, 232-239.

McKenna, P. (2004). Disorders of language and communication. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians*. (pp. 165-184).

McNeil, J. E. (2004). Disorders of number processing and calculation. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians*. (pp. 253-272). Chichester: John Wiley & Sons, Ltd.

Medical Research Council. (1999). *Cape Metropolitan study. Cape Town: National Trauma Research Programme.*

Mellers, J. D. C. (2004). Neurological Investigations. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology* (pp. 57-78). Chichester: John Wiley and Sons Ltd.

Milberg, W. P., & Hebben, N. (2006). The historical antecedents of the Boston Process Approach. In A. M. Poreh (Ed.), *The quantified process approach to neuropsychological assessment* (pp. 17-25). Philadelphia: PA.

Milberg, W. P. (1996). Issues in the assessment of cognitive function in dementia. *Brain and Cognition*, 31, 114-132.

Miller, B. L., Darby, A., Benson, F., Cummings, J. L., & Miller, M. H. (1997). Aggressive, socially disruptive and anti-social behaviour associated with fronto-temporal dementia. *British Journal of Psychiatry*, 170, 150-154.

Ministry of Health. (1997). *South African Government and Department of Health White Paper for the Transformation of the Health System in South Africa.* South Africa.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of Normative Data for Neuropsychological Assessment* (2nd ed.). New York: Oxford University Press, Inc.

Mkoka, S., Vaughan, J., Wylie, T., Yelland, H., & Jelsma, J. (2003). The pitfalls of translation — a case study based on the translation of the EQ-5D into Xhosa. *South African Medical Journal*, *93*, 265-266.

Monsch, A. U., Foldi, N. S., Ermini-Funfschilling, D. E., & Berres, M. (1995). Improving the diagnostic accuracy of the Mini-Mental State Examination. *Acta Neurologica Scandinavica* *92*, 145-150.

Morris, R. G. (2004). Neuropsychology of older adults. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians*. (pp. 301-318). Chichester: John Wiley & Sons Ltd.

Nathan, J., Wilkinson, D., Stammers, S., & Low, J. (2001). The role of tests of frontal executive functions in the detection of mild dementia. *International Journal of Geriatric Psychiatry*, *16*, 18-26.

Nell, V. (1997). Science and Politics Meet at Last: The Insurance Industry and Neuropsychological Norms. *South African Journal of Psychology*, *27*(1), 43-51.

Nell, V. (2000). *Cross-Cultural Neuropsychological Assessment: Theory and Practice*. New Jersey: Lawrence Erlbaum Associates.

Nestor, P., & Hodges, J. R. (2001). The clinical approach to assessing patients with early onset dementia. In J. R. Hodges (Ed.), *Early-onset Dementia: a multidisciplinary approach* (pp. 23-46).

Nilsson, F. M. (2007). Mini Mental State Examination (MMSE) – probably one of the most cited papers in health science. *Acta Psychiatrica Scandinavica*, *116*, 156-157.

Ostrosky-Solis, F., Lopez-Arango, G., & Ardila, A. (2000). Sensitivity and Specificity of the Mini-Mental State Examination in a Spanish-Speaking Population. *Applied Neuropsychology*, *7*(1), 25-31.

Pacaud, S. (1989). Performance in relation to age and educational level: A monumental research. *Experimental Aging Research*, *15*, 123-136.

Park, K. W., Pavlik, V. N., Rountree, S. D., Darby, E. J., & Doody, R. S. (2007). Is functional decline necessary for a diagnosis of Alzheimer's disease? *Dementia and Geriatric Cognitive Disorders*, *24*, 375-379.

Perez-Arce, P., & Puente, A. (1996). Neuropsychological assessment of ethnic minorities. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological tests*. Delray Beach: GR Press.

Pharos. (1998). *English–Xhosa, Xhosa–English Dictionary*. Cape Town: Pharos

Philcox, D., Kies, B., Lee Pan, E., Eastman, R., Bryer, A., & Silber, M. (2000). *Clinical neurology for students*. (2nd ed.): Neurology Unit of the Department of Medicine, University of Cape Town.

Pimental, P. A., & Ross, C. (2003). ROCF productions in right- and left-hemisphere lesion patients. In J. A. Knight (Ed.), *The handbook of Rey-Osterrieth Complex Figure usage: Clinical and research applications*. Lutz, FL: Psychological Assessment Resources.

Ponton, M. O., Satz, P., & Herrera, L. (1996). Normative data stratified by age and education for the Neuropsychological Screening Battery for Hispanics (NeSBHIS): Initial report. *Journal of the International Neuropsychological Society*, 2, 96-104.

Puenta, A. E., & Ardila, A. (2000). Neuropsychological assessment of Hispanics. In E. Fletcher-Janzen & T. L. Strickland (Eds.), *Handbook of cross-cultural neuropsychology: Critical issues in neuropsychology*. (pp. 87-104). Dordrecht: Kluwer Academic Publishers.

Rapp, S. R., Espeland, M. A., Hogan, P., Jones, B. N., & Dugan, E. (2003). Baseline experience with Modified Mini Mental State Exam: The women's health initiative memory study. *Aging and Mental Health*, 7, 217-223.

Rapport, L. J., Charter, R. A., Dutra, R. L., Farchione, T. J., & Kingsley, J. J. (1997). Psychometric properties of the Rey-Osterrieth Complex Figure: Lezak-Osterrieth versus Denman scoring systems. *Clinical Neuropsychologist*, 11, 46-53.

Retzlaff, P. D., & Gibertini, M. (1994). Neuropsychometric issues and problems. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment*. Hillsdale, NJ: Erlbaum.

Rosenfeld, B., Sands, S. A., & VanGorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*, 349-359.

Ross, T. P., Calhoun, E., Cox, T., Wenner, C., Kono, W., & Pleasant, M. (2007). The reliability and validity of qualitative scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology, 22*, 475-488.

Ross, T. P., Lichtenberg, P. A., & Christensen, B. K. (1995). Normative data on the Boston Naming Test for elderly adults in a demographically diverse medical sample. *The Clinical Neuropsychologist, 9*, 321-325.

Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition, 52*, 326-333.

Rosselli, M., Ardila, A., & Rosas, P. (1990). Neuropsychological assessment in illiterates II: Language and praxic abilities. *Brain and Cognition, 12*, 281-296.

Russell, E. W., & Russell, S. L. K. (2003). Twenty ways and more of diagnosing brain damage when there is none. *Journal of Controversial Medical Claims, 10*, 1-14.

Russell, E. W., Russell, S. L. K., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology, 20*, 785-794.

Salazar, G. D., Garcia, M. P., & Puenta, A. E. (2007). Clinical neuropsychology of Spanish speakers: The challenge and pitfalls of a neuropsychology of a heterogenous population. In B. Uzzell, M. O. Ponton & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology*. (pp. 283-302). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Salmon, D. P., & Hodges, J. R. (2003). Neuropsychological assessment of early onset dementia. In J. R. Hodges (Ed.), *Early-onset Dementia: a multidisciplinary approach* (pp. 47-73). Oxford: Oxford University Press.

Saxton, J., Ratcliff, G., Munro, C. A., Coffey, E. C., Becker, J. T., & Fried, L. (2000). Normative data on the Boston Naming Test and two equivalent 30-item short forms. *Clinical Neuropsychologist*, *14*, 526-534.

Schnider, A., von Daniken, C., & Gutbrod, K. (1996). Disorientation in amnesia: A confusion of memory traces. *Brain*, *119*, 1627-1632.

Schwamm, L. H., VanDyke, C., & Kiernan, R. J. (1987). The Neurobehavioral Cognitive Status Examination. *Annals of Internal Medicine*, *107*, 486-491.

Shuttleworth-Jordan, A. B. (1996). On not reinventing the wheel: A clinical perspective on culturally relevant test usage in South Africa. *South African Journal of Psychology*, *26*, 96-102.

Simard, M. (1998). The Mini-Mental State Examination: Strengths and weaknesses of a clinical instrument. *The Canadian Alzheimer's Disease Review* (Dec), 10-12.

Small, B. J., Viitanen, M., & Bacekman, L. (1997). Mini-Mental State Examination items scores as predictors of Alzheimer's disease: Incidence data from the Kungsholmen Project, Stockholm. *Journals of Gerontology*, 52, 299-304.

Solms, M., & Turnbull, O. (2002). *The brain and the inner world: An introduction to the neuroscience of subjective experience*. New York: Other Press.

Spreeen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Spreeen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests*. (2nd ed.). New York: Oxford University Press.

Statistics South Africa. (2004). *Statistics South Africa: Annual Report 2003/2004*. Retrieved January 29, 2005, from <http://www.info.gov.za/annualreport/2003/statssa03.pdf>

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, 53, 401-433.

Sugarman, R. (2007). Visuospatial assessment in cross-cultural and nonwestern settings. In B. P. Uzzell, M. O. Ponton & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology*. (pp. 227-251). Mahwah: Lawrence Erlbaum Associates, Publishers.

Sumerall, S. W., Timmons, P. L., James, A. L., Ewing, M. J. M., & Oehlert, M. E. (1997). Expanded norms for the Controlled Oral Word Association Test. *Journal of Clinical Psychology, 53*, 517-521.

Swartz, L. (1998). *Culture and mental health: A southern African view*. Oxford: Oxford University Press.

Sweet, J. J., Suchy, Y., Leahy, B., Abramowitz, C., & Nowinski, C. J. (1999). Normative clinical relationships between orientation and memory: Age as an important moderator variable. *The Clinical Neuropsychologist, 13*, 495-508.

Temple, C. M. (2002). Oral fluency and narrative production in children with Turner's syndrome. *Neuropsychologia, 40*, 1419-1427.

Teng, E. L. (1996). Cross-cultural testing and the cognitive abilities screening instrument. In G. Yeo & Gallagher-Thompson (Eds.), *Ethnicity and the dementias*. (pp. 77-85). Washington, DC: Taylor & Francis.

Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (MMS) examination. *Journal of Clinical Psychiatry, 48*, 314-318.

Teng, E. L., Chui, H. C., & Gong, A. (1990). Comparisons between the Mini-Mental State Exam and its modified version. *ANONYMOUS, Psychogeriatrics: Biomedical and social advances, 189-192*.

Teng, E. L., Chui, H. C., Schneider, L. S., & Metzger, L. E. (1987). Alzheimer's dementia: Performance on the Mini-Mental State Exam. *Journal of Consulting and Clinical Psychology, 55*.

Teuber, H. L. (1955). Physiological psychology. *Annual Review of Psychology, 6*, 267-296.

Teuber, H. L. (1975). Recovery of function after brain injury. In *Ciba Foundation Symposium 34, Outcome of Severe Damage to the Central Nervous System* (pp. 159–190). Amsterdam: Elsevier.

Tollman, S. G., & Msengana, N. B. (1990). Neuropsychological assessment: Problems in evaluating the higher mental functioning of Zulu-speaking people using traditional Western techniques. *South African Journal of Psychology, 20*, 20-24.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*, 167-177.

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubble, A. M. (1996). Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A Psychometric Comparison and Normative Data. *Psychological Assessment, 8*, 48-59.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A comprehensive review. *Journal of the American Geriatrics Society, 40*, 48-59.

Tsang, H. L., & Lee, T. M. (2003). The effect of ageing on confrontational naming ability. *Archives of Clinical Neuropsychology, 18*, 81-89.

UNAIDS 2006 Report on the global AIDS epidemic. (2006). Retrieved May 9, 2005, from http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp

UNAIDS Table of country-specific HIV/AIDS estimates and data, end 2003 (July 2004). Retrieved May 9, 2005, from http://www.unaids.org/bangkok2004/GAR2004_html/GAR2004_00_en.htm

Uzzell, B. P., Ponton, M. O., & Ardila, A. (2007). Grasping the cross-cultural reality. In B. P. Uzzell, M. O. Ponton & A. Ardila (Eds.), *International handbook of cross-cultural neuropsychology*. (pp. 1-19). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Van der Vijver, F., & Hambleton, R. K. (1996). Translating tests: some practice guidelines. *European Psychologist, 1*, 89-99.

Van der Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. New York: Sage.

Van Eeden, R., & De Beer, M. (2005). Assessment of cognitive function. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context*. (2nd ed., pp. 119-139). Oxford: Oxford University Press.

Walsh, K. W. (1985). *Understanding Brain Damage: A Primer of Neuropsychological Evaluation*. New York: Churchill Livingstone.

Walsh, K. W. (1987). *Neuropsychology: A clinical approach* (2nd ed.). United Kingdom: Churchill Livingstone.

Walsh, K. W. (1991). *Understanding brain damage: A primer of neuropsychological evaluation*. London: Churchill Livingstone.

Walsh, K. W., & Darby, E. J. (1999). *Neuropsychology: A clinical approach* (4th ed.). Edinburgh: Churchill Livingstone.

Welch, L. W., Doineau, D., Johnson, S., & King, D. (1996). Educational and gender normative data for the Boston Naming Test. *Brain and Language*, 53, 260-266.

Wigren, J. (1994). Narrative completion in the treatment of trauma. *Psychotherapy*, 31, 415-423.

Williams, B. W., Mack, W., & Henderson, V. W. (1989). Boston Naming Test in Alzheimer's disease. *Neuropsychologia*, 27, 1073-1079.

Wober, M. (1966). Sensotypes. In J. W. Berry & P. R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural psychology*. (pp. 119-127). London: Methuen & Co Ltd.

World Health Organisation. (2003). *World Report on Traffic Injury Prevention*. Geneva.

Worrall, L. E., Yiu, E. M.-L., Hickson, L. M. H., & Barnett, H. M. (1995). Normative data for the Boston Naming Test for Australian elderly. *Aphasiology*, *9*, 541-551.

Zenicus, A. H., Wesolowski, M. D., & Rodriguez, I. M. (1998). Improving orientation in head injured adults by repeated practice, multi-sensory input and peer participation. *Brain Injury*, *12*, 53-61.

University of Cape Town

Appendix A

University of Cape Town

Appendix B

University of Cape Town

Appendix C

University of Cape Town

Appendix D

University of Cape Town

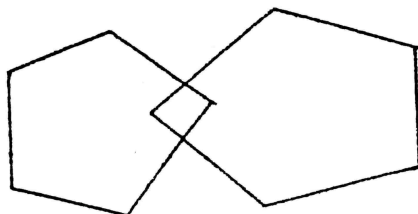
Appendix E

University of Cape Town

Appendix F

MINI-MENTAL STATE EXAMINATION

ORIENTATION .	<u>SCORE</u>
1. What is the YEAR?	_____ 1
SEASON?	_____ 1
MONTH?	_____ 1
DAY?	_____ 1
DATE?	_____ 1
2. Where are we COUNTRY?	_____ 1
PROVINCE?	_____ 1
TOWN OR CITY?	_____ 1
HOSPITAL?	_____ 1
WARD?	_____ 1
 REGISTRATION	
3. Name three objects, taking one second to say each. Then ask the patient all three after you have said them. Give one point for each correct answer. Repeat the answers until patient learns all three.	_____ 3
 ATTENTION AND CALCULATION	
4. Serial sevens. Give one point for each correct answer. Stop after five answers. ALTERNATE: Spell WORLD backwards, or HERFS	_____ 5
 RECALL	
5. Ask for names of three objects learned in Q.3. Give one point for each correct answer.	_____ 3
 LANGUAGE	
6. Point to a pencil and a watch. Have the patient name them as you point.	_____ 2
7. Have the patient follow a three-stage command: "Take this paper in your right hand. Fold the paper in half. Put the paper on the floor".	_____ 3
8. Have the patient repeat "No ifs, ands or buts". "Nog vis, nog vleys, noq voël".	_____ 1
9. Have the patient read and obey the following : "CLOSE YOUR EYES". (Write it in large letters).	_____ 1
10. Have the patient write a sentence of his or her choice. (The sentence should contain a subject and an object, and should make sense. Ignore spelling errors when scoring).	_____ 1
11. Have the patient copy the design printed below. (Give one point if all sides and angles are preserved and if the intersecting sides form a diamond shape).	_____ 1
	_____ = TOTAL



DATE:

Instructions for administration of the Mini-Mental State Examination

Orientation

- (1) Ask for the date. Then ask specifically for parts omitted, e.g., "Can you also tell me what season it is?" One point for each correct.
- (2) Ask in turn "Can you tell me the name of this hospital?" (town, county, etc.). One point for each correct.

Registration

Ask the patient if you may test his memory. Then say the names of 3 unrelated objects, clearly and slowly, about one second for each. After you have said all 3, ask him to repeat them. This first repetition determines his score (0-3) but keep saying them until he can repeat all 3, up to 6 trials. If he does not eventually learn all 3, recall cannot be meaningfully tested.

Attention and calculation

Ask the patient to begin with 100 and count backwards by 7. Stop after 5 subtractions (93, 86, 79, 72, 65).

Score the total number of correct answers.

If the patient cannot or will not perform this task, ask him to spell the word "world" backwards. The score is the number of letters in correct order. E.g. dlrow = 5, dlrow = 3.

Recall

Ask the patient if he can recall the 3 words you previously asked him to remember. Score 0-3.

Language

Naming: Show the patient a wrist watch and ask him what it is. Repeat for pencil. Score 0-2.

Repetition: Ask the patient to repeat the sentence after you. Allow only one trial. Score 0 or 1.

3-Stage command: Give the patient a piece of plain blank paper and repeat the command.

Score 1 point

for each part correctly executed.

Reading: On a blank piece of paper print the sentence "Close your eyes", in letters large enough for the patient to see clearly. Ask him to read it and do what it says. Score 1 point only if he actually closes his eyes.

Writing: Give the patient a blank piece of paper and ask him to write a sentence for you. Do not dictate a sentence; it is to be written spontaneously. It must contain a subject and verb and be sensible. Correct grammar and punctuation are not necessary.

Copying: On a clean piece of paper, draw intersecting pentagons, each side about 1 in., and ask him to copy it exactly as it is. All 10 angles must be present and 2 must intersect to score 1 point. Tremor and rotation are ignored.

Estimate the patient's level of sensorium along a continuum, from alert on the left to coma on the right.

Appendix G

AGE and EDUCATION WEIGHTED NORMS TABLE FOR MINI-MENTAL STATE EXAMINATION
Ages 18 - 85+ and Education Level 0 - 13+ YEARS

Source: Crum R, Anthony JC et al. (1993) Population-based norms for the Mini-Mental State Examination by age and educational level.

JAMA [May 12, 1993], 269(18), 2386-2391.

Compiled by: Dr. Bill Lynch - BIRU [2B2-PAD]

AGE:	EDUCATIONAL LEVEL [Yrs]:							
	0 - 4		5 - 8		9 - 12		13 or more*	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
18-24	22 [N = 17]	2.9	27 [N = 94]	2.7	29 [N = 1326]	2.2	29 [N = 783]	1.3
25-29	25 [N = 23]	2.0	27 [N = 83]	2.5	29 [N = 958]	1.3	29 [N = 1012]	0.9
30-34	25 [N = 41]	2.4	26 [N = 74]	1.8	29 [N = 822]	1.3	29 [N = 989]	1.0
35-39	23 [N = 33]	2.5	26 [N = 101]	2.8	28 [N = 668]	1.8	29 [N = 641]	1.0
40-44	23 [N = 36]	2.5	27 [N = 100]	1.8	28 [N = 489]	1.9	29 [N = 354]	1.7
45-49	23 [N = 28]	3.7	26 [N = 121]	2.5	28 [N = 423]	2.4	29 [N = 259]	1.6
50-54	23 [N = 34]	2.6	27 [N = 154]	2.4	28 [N = 462]	2.2	29 [N = 220]	1.9
55-59	22 [N = 49]	2.7	26 [N = 208]	2.9	28 [N = 525]	2.2	29 [N = 231]	1.5
60-64	23 [N = 88]	1.9	26 [N = 310]	2.3	28 [N = 626]	1.7	29 [N = 270]	1.3
65-69	22 [N = 126]	1.9	26 [N = 633]	1.7	28 [N = 814]	1.4	29 [N = 358]	1.0
70-74	22 [N = 139]	1.7	26 [N = 533]	1.8	27 [N = 550]	1.6	28 [N = 255]	1.6
75-79	21 [N = 112]	2.0	25 [N = 437]	2.1	27 [N = 315]	1.5	28 [N = 181]	1.6
80-84	20 [N = 105]	2.2	25 [N = 241]	1.9	25 [N = 163]	2.3	27 [N = 96]	0.9
85 and up	19 [N = 81]	2.9	23 [N = 134]	3.3	26 [N = 99]	2.0	27 [N = 52]	1.3
ALL AGES	22 [N = 892]	2.3	26 [N = 3223]	2.2	28 [N = 8240]	1.9	29 [N = 5701]	1.3

*College experience or higher degree

Appendix H

Screening Sheet

Please indicate whether you have had any of the following (either currently or previously):

Please be assured of the confidentiality and anonymity of any personal information that you give when participating in this study.

	<i>Please tick all that apply:</i>
A stroke	
A heart operation to treat coronary artery disease, e.g. coronary artery bypass graft surgery or stenting.	
Epilepsy or other seizures/fits	
A severe head injury	
Brain tumour or cancer	
Hydrocephalus ('water on the brain')	
Herpes encephalitis	
TB (tuberculosis)	
Diabetes (sugar disease)	
High blood pressure (hypertension)	
Multiple sclerosis	
Meningitis	
Alzheimer's disease	
Parkinson's disease	
Systemic Lupus Erythematosus (SLE)	
Syphilis	
HIV/AIDS	
Neurocysticercosis	
Recreational drugs (e.g. dagga, tik, cocaine etc)	
Any psychiatric condition (e.g. schizophrenia, bipolar)?	
Any other disease that you've had/currently have that may affect the brain?	

Appendix I

CONSENT FORM

TITLE OF PROJECT: The Groote Schuur Neurocognitive Screening Battery

*Please cross out
as necessary*

Have you read the Patient Informantion Sheet? YES/NO

Have you had an oppportunity to ask questions and discuss the study? YES/NO

Have your received satisfactory answers to all your questions? YES/NO

Have you received enough information about the study? YES/NO

Who have you spoken to? Dr/Mr/Mrs/Ms/Prof.

Do you understand that you are free to withdraw from the study:

- at any time
 - without having to give a reason for withdrawing
 - and without affecting your future treatment
- YES/NO

Do you consent to the unattributed and confidential use of these recordings for scientific purposes? YES/NO

Signed:

Date:

(NAME IN BLOCK LETTERS)

Appendix J

Patient Information Sheet

TITLE OF PROJECT:

The Groote Schuur Neurocognitive Screening Battery

- You are invited to participate in a neuropsychological study conducted at Groote Schuur Hospital. Please read this information sheet carefully and do not hesitate to ask the researcher for any additional information.
- The overall purpose of the investigation is to adapt and validate a South African neurocognitive screening battery, which comprises neuropsychological tests.
- You are asked to take part in this study by participating with different neuropsychological tests and tasks. You will be asked to attend two half-hour testing sessions a week apart.
- There are no anticipated risks involved in this research, but if you should experience mental and/or physical fatigue, or any form of psychological distress please be aware that you could inform the researcher immediately.
- It is up to you to decide whether or not you take part. If you decide to take part you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part you are still free to withdraw from the study at any time, without having to give a reason and without this affecting future treatment.
- The confidentiality of your answers and your identity will be protected. All data collected will be suitably anonymous, securely stored, made accessible only to the researcher, and destroyed at the end of the project.
- This study is an educational project, forming part of a Ph.D. degree at the University of Cape Town (UCT). The research will be carried out by researchers from UCT and will be funded by the same university.
- It has been reviewed by the UCT Psychology Department's Ethics Committee.
- If you have any questions regarding this study, or concerns regarding the manner in which the study was conducted, or would like to be informed of the results when the study is completed, please feel free to contact the principal researcher.
- Address for communications:

Professor Mark Solms:
 Department of Psychology
 University of Cape Town
 Rondebosch 7701

Ph. (021) 650-3437

Appendix K

Scoring Sheet

Patient's name: _____ no. _____

1. Orientation

1.1 Person

Normal score = 2/2

1.2 Place

Normal = 3/4

1.3 Time

Normal = 3/5

2. Memory

2.1 Auditory Span

- Record patient's response on the lines provided:

2, 7

5, 7, 2

1, 9, 6, 4

1, 4, 2, 7, 9

8, 3, 7, 4, 6, 2

7, 3, 5, 2, 4, 1, 9

7 Digits = 2/2
5 Digits = 0/2
Normal = 6 digits

2.2 Four Hidden Objects

1. Show the 4 objects (a key, pipe, flower and bangle) and ask the patient to name them – *record items left out*;

4 Objects = 2/2
3 Objects = 1/2
2 Objects = 0/2
Normal = 4 objects

2. Hide all 4 objects in one location e.g. under the sheet and immediately ask the patient what the objects were – *record items left out*;

3. Distract the patient (eg. ask them about names and ages of children);

4. Repeat the question (NB: do not tell the patient how many objects or where you hid them);

4 Objects = 2/2
3 Objects = 1/2
2 Objects = 0/2
Normal = 4 objects

5. If patient fails, repeat the process. If patient successful, proceed to step 7 – *record how many repetitions*;

6. Hide all 4 objects in different locations;

2.4 Rey Complex Figure - **Copy**

Name: _____

Date and time administered: _____

University of Cape Town

Rey Complex Copy – **Second recall - immediate**

Name: _____ Date and time administered: _____

University of Cape Town

Rey Complex Copy – **Delayed recall at thirty minutes**

Name: _____ Date and time administered: _____

University of Cape Town

3. Language tests

Washing Line test

Show patient the washing line picture and ask them to discuss what they see. Record verbatim what s/he says. Give the patient one minute to discuss the picture.

Fluency .../2
Normal = 2
Mild defect = 1
paraphasia .../2
Normal = 2
mild defect = 1

Writing

Ask patient to: write their own name

Write sentence to dictation:

Write spontaneously (a full, grammatical sentence)

Were these commands correctly followed? Was the writing the same as spoken production or normal/better than spoken production?

*Same as spoken
 production = 0/1*
*Normal or better
 than spoken
 production = 1/1*

Comprehension tests

Procedure: Utter the following verbal commands and comment on the appropriateness of the patient's responses

.../3
Normal = 3
.../5
Normal = 4

Reading tests**Show me three fingers**

.../1

Procedure: "I'm going to give you a short story. Please read it aloud."

Mary Selo story

Fluency: Make a note of time taken to read aloud story

Were all words properly enunciated? Make notes of problematic words

Other comments

.../2
 Mild defect = 1
 Normal = 2

Repetition

"Why am I sitting here?" (score =1)

"The painter painted many beautiful scenes" (cumulative score = 2)

"This doctor does not visit all the patients in the ward" (cumulative score =3)

"Why do the members of the committee not ask their representatives for aid?" (cumulative score=4)

Procedure: Compare to production

.../4
 Normal = 3

Naming

Procedure: Ask patient to name body parts and objects at the bedside:

"Elbow, ankle, wrist, knee, shoulder"

.../5
 Normal =4

"Pillow, sheet, spectacles, collar, buckle

.../5
 Normal =4
 Total

Naming Test:Drawing shown:Response from patient:Incorrect/correct

3. Visual-spatial assessment

a. Neglect

Visual:

Tactile:

Auditory:

- (a) All stimuli correctly identified consistently in all modalities
not affected by primary sensory impairment = 2/2*
*(b) Consistently identifies both unilateral (L) and (R) stimuli
correctly but frequently neglects (L) on bilateral stimulation
(in modalities without primary sensory impairment) = 1/2*
*(c) Consistently neglects (L) even on unilateral stimulation
(in modalities without primary sensory impairment) = 0/2*

4. Visual-spatial assessment

a. Anosognosia

If the patient does not spontaneously describe deficit, ask "Please describe all your current symptoms/deficits"

Score 3/3 if they can

If they do not describe deficit, ask "What about your legs/arms/hands/eyes, etc. (where applicable), are they all functioning normally?"

Score 2/3 if they can

If still denies deficit, demonstrate deficit to patient by physical examination, then ask: "Do you still think that your... is functioning normally?"

Score 1/3

Is there still a denial of deficit?

Score 0/3

5. Visual-spatial assessment

a. Hut drawing test

<i>Picture proportionate = 1</i> <i>Any sign that the left side of the picture is neglected = 0</i>
--

Hut drawing - copy

University of Cape Town

4. Executive

4.1 Controlled Oral Word Association Test

Procedure:

Ask patient to tell you as many words beginning with the letter F, A and S (or equivalent letters in Xhosa/Afrik). They must exclude proper names like the names of their friends and family or products like "Nike", must only be objects, give some examples.

Make a line for each 15 seconds to mark where the patient is. Allow *60 seconds* for each trial.

->Please record repetitions with an "r" and rule breaks

F

A

S

Total F:

Total A:

Total S:

Total: _____

(a) *More than 25 words (normal) = 2/2*

(b) *Between 15 and 25 words = 1/2*

(c) *Less than 15 words = 0/2*

4.2 Red/Green

Perfect with one or two mistakes = 2

Consistently imperfect performance = 1

Gross-evidence of imperfect impulsivity, rule breaking, stereotyped responses = 0

4.3 Fist-Side-Palm

Perfect performance after one or initial errors = 1

Inability to achieve perfect performance despite repeated trials = 0

4.4 Tapping Test

Perfect performance after initial errors = 1

Inability to achieve criterion despite repeated trials = 0

4.5 Repeated Pattern Drawing



4.6 18 Book Problem

Correct response at first attempt = 2

Correct response after initial impulsive or stereotyped response = 1

Two incorrect responses = 0

Summary of Scores

Assessment of Orientation	Subtotal	Total	Normal
Orientation to Person			
Orientation to Place			
Orientation to Time	/11	8
Assessment of Memory Function			
Digit Span			
Registration			
Simple Recall			
Complex Recall	/18	14
Assessment of Language Function			
Production			
Comprehension			
Repetition			
Naming	/30	26
Assessment of Spatial Cognition			
Visuospatial Analysis			
Neglect			
Anosognosia	/15	12
Assessment of Executive Function			
Deep White Matter			
Mesial			
Orbital / Basal			
Dorsolateral	/11	10
Total	/85	70