A Preliminary Standardisation of the Wisconsin Card Sorting Test for Setswana-Speaking University Students

by

C. Gadd

under the supervision of

Dr W.D. Phipps

Submitted in partial fulfilment of the requirements for the degree

MASTER OF SCIENCE (CLINICAL PSYCHOLOGY)

in the

Department of Clinical Psychology

in the

School of Social Sciences

and the

Faculty of Medicine

at the

University of Limpopo

(Medunsa Campus)

2011

Plagiarism Declaration

I, Christi Gadd, declare that the research entitled, *A Preliminary Standardisation of the Wisconsin Card Sorting Test for Setswana-Speaking University Students*, to the best of my knowledge and belief, is my own work and contains no plagiarism. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

All sources have been acknowledged and all references contained within it have been correctly cited and the original authors acknowledged making use of the American Psychological Association 6th edition referencing guidelines.

I have not previously submitted or intend to submit in future this work or any version of it for assessment in any other unit or award offered by the University of Limpopo, its partner institutions, or any other institution.

Signed: _________________
Ms. C. Gadd

Date: _________________

Abstract

Psychological and neuropsychological assessment in South Africa currently faces various challenges, among which, is the prominent need for tests that are standardised for the multi-cultural South African context. The absence of adequate standardisations and normative data currently hamper the confidence with which tests can be used in South Africa. This research project aimed to construct a preliminary standardisation of the Wisconsin Card Sorting Test (WCST) for Setswana-speaking university students. The testing instructions were translated to Setswana and 93 participants were assessed using a computerised version of the WCST. Hypothesis tests showed that some sub-scores significantly differed from the US norms and others did not. The distribution of WCST performance in the study population does, therefore, not follow the same distribution as the US distribution and a need was identified to construct new normative data.

Regression analysis indicated that not gender, age, or level of education influenced participants' WCST scores for all sub-scores, with the exception of the *Trials to complete first category* score that was influenced by age. One normative table was thus constructed for the entire 18- to 29-year-old age group for the variables *Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Failure to maintain set, Learning to learn* and *Percent conceptual level responses,* whilst separate norm tables for the 18- to 19-year-old and 20- to 29-year-old age groups were constructed for the *Trials to complete first category* sub-score.

Furthermore, the WCST displayed adequate internal consistency in the study population. These norms and psychometric properties are, however, subject to certain limitations and it is thus recommended that a full standardisation of the WCST be constructed for the South African population. Various unanswered questions were, however, identified in how tests should be standardised for the extremely diverse South African context. It was concluded that a great deal of academic discourse is still required in order to make fair assessment available to every South African citizen.

[326 words]

Acknowledgements

It takes a village to complete a research project! How do I express my thanks to all those without whom this research would not have been possible? My life is richer because of you. First of all, to my Creator for blessing me with the capacity to follow this career, let my life song sing to You.

Phillip, dankie dat jy geluister het, moed ingepraat het en elke opwinding en teleurstelling met my gedeel het: "Together we'll be giants and our own cure… We'll escape to our perfect place."

Mamma en Pappa, dankie dat julle my droomloopbaan moontlik gemaak het. Dankie vir jare se ondersteuning en liefde.

Aan my toekomstige ma en pa, dankie dat julle my gevoer en gehuisves het tydens die sewe maer jare. Dankie vir julle belangstelling, ondersteuning en liefde.

Leandri, wat 'n voorreg om nog 'n sak sout saam met jou te eet. Dankie vir jou luister, lees, gemak, praat, stilwees, saamwerk, saam-kla, empatie en onvoorwaardelike vriendskap die afgelope paar jare. Ek waardeer jou.

Sian, thank you for Thursday liftclub. Sharing this experience with you has made it much more fun. Thank you for your brainstorming, proofreading, finding "unfindable" information, wiping tears and late-night writing sessions. I'm honoured to share this journey with you.

Michael and Erika, thank you for completing another leg of the journey with me. Thank you for being a BBM away in times of crisis and of joy. Long live the Foundation for Sustainable Tuna Development. *"Ek lewe... gemaak om na liefde te strewe!"*

Estian, jy is my taalredigeringridder. Maite, thank you for helping with the translation. Margi, thank you for dealing with the frantic researcher with kindness and patiently explaining all the statistics over and over again.

To Digby Ormond-Brown and Ormond Software Enterprises for the software that was essential to the data collection.

Warwick, thank you for your time and going beyond the call of duty to assist me with this research. Thank you for your wisdom, commitment, feedback, empathy, confirmation and understanding. In addition to research skills, I have learnt from you what the definition of clear communication is. I feel privileged to have had you as a supervisor.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Table of contents

List of figures

List of tables

Chapter 1: Introduction

People have been trying to objectively measure and predict human behaviour since ancient times (Foxcroft & Roodt, 2005). In lieu of this, psychological assessment has proved to be a significant aid. However, since psychological assessment as a measure of behaviour has been misused during the apartheid era, it has been widely criticised within the South African context (Foxcroft & Roodt, 2005). Foxcroft and Roodt (2005) argue that this has resulted in the development of a great need for more appropriate measures to be cultivated, which can then be used with all cultural groups in South Africa.

1.1 Psychological Assessment

"Psychological assessment is a process oriented activity aimed at gathering a wide array of information by using assessment measures (tests) and information from many other sources" (Foxcroft & Roodt, 2005, pp. 3–4). These psychological tests or assessment measures are tools that allow us to measure human behaviour. The field of psychometrics (as a subfield of psychology) is concerned with ensuring that these tests that are used are valid and reliable, by compiling theory and research to certify that they are applicable to the measurement of psychological characteristics (Foxcroft & Roodt, 2005; Zillmer, Spiers & Culbertson, 2004). According to Foxcroft and Roodt, this entails the systematic and scientific development of tests as well as the construction of psychometric properties for each of these tests.

1.2 Neuropsychological Assessment

Neuropsychology as another subfield of the discipline of psychology, has its roots in neurology, psychiatry and psychology, and explores the relationship between the brain and human behaviour (Kolb & Wishaw, 2008). Its aim is to scientifically study human behaviour based on the functions of the brain, including the identification of impairments in behaviour that can result from brain trauma or disease (Kolb & Wishaw, 2008).

Neuropsychological assessment necessitates the use of psychological tests and clinical observation in the assessment of neurological impairment (Kolb & Wishaw, 2008).

1.2.1 Neuropsychological tests. The behavioural manifestations of brain damage can be extremely heterogeneous (Groth-Marnat, 2009). Neuropsychological tests facilitate the visibility of neurological disorders to the clinician from the behaviour of the persons being tested (Kolb & Wishaw, 2008). This neuropsychological assessment determines the presence of neurological impairment by making use of quantitative cut-off scores within the tests (Groth-Marnat, 2009). Groth-Marnat states that this requires the consideration of norms constructed for different ages, genders and educational levels. The process of neuropsychological assessment thus depends, to a large extent, on the reliability and validity of the neuropsychological tests that are used (Sherman, Strauss & Spreen, 2006).

Neuropsychological tests can be organised into five measurement categories or functions that cover the primary areas of cognitive functioning (Groth-Marnat, 2009). According to Groth-Marnat these five areas include *attention*, *language*, *memory*, *spatial skills* and *executive* functions.

1.3 Executive Function

Executive function involves a person's capacity to "effectively regulate and direct self-behaviour" (Groth-Marnat, 2009, p. 499). Sherman et al. (2006) define the executive function as "a multidimensional construct referring to a variety of loosely related higher order cognitive processes including initiation, planning, hypothesis generation, cognitive flexibility, decision making, regulation, judgement, feedback utilization, and self-perception that are necessary for effective and contextually appropriate behaviour" (p. 171). Executive function therefore includes planning, flexibility of thought, judgment, and goal-directed behaviour (Sherman et al., 2006). Consequently, Sherman et al. (2006) describe these functions as the ability to make decisions when dealing with unfamiliar situations.

Executive difficulties are most typically caused by frontal lobe damage, but can also result from damage to the subcortical regions (especially thalamic structures), or diffuse damage to the brain (Groth-Marnat, 2009). According to Purdon and Waldie (2001), the occurrence of deficits in executive functioning is an area often considered one of the most prominent barriers to vocational reintegration. Persons being tested can present with a variety of difficulties in executive functioning that ranges from a semi-vegetative state (in which activity is hardly ever initiated, despite unharmed cognitive abilities) to executive difficulties (as evident by a lack of awareness of their impact on others) and an inability to effectively regulate behaviour (Groth-Marnat, 2009).

1.3.1 Measuring executive function. Executive difficulties are often overlooked during formal assessment since they can occur whilst other cognitive abilities are intact (Groth-Marnat, 2009; Sherman et al., 2006). Formal assessments are usually very structured, leaving little room for inappropriate behaviour and limits the person in showing difficulties in initiation, planning and judgment of behaviour (Groth-Marnat, 2009; Sherman et al., 2006).

In addition to being overlooked, executive dysfunction may also be frequently misdiagnosed as depression, since apathy, lack of affect, and lack of direction can also occur as a result of brain impairment (Groth-Marnat, 2009). Executive function is thus one of the primary areas of neuropsychological assessment that is considered when compiling a neuropsychological test battery (Pepping, 2003). For this purpose, Pepping suggests making use of the Wisconsin Card Sorting Test.

1.4 The Wisconsin Card Sorting Test (WCST)

The WCST is a neuropsychological test which is extremely sensitive to frontal lobe lesions (Heaton, Chelune, Talley, Kay & Curtiss, 1993; Kolb & Wishaw, 2008; Purdon & Waldie, 2001). It provides an indication of frontal lobe activity by measuring executive functioning (Heaton et al., 1993; Kolb & Wishaw, 2008) and cerebral blood-flow studies have found activation in the prefrontal lobe whilst subjects are engaging in the WCST (Smith-Seemiller, Franzen & Bowers, 1997). The WCST assesses abstract reasoning and mental flexibility (Smith-Seemiller et al., 1997) and provides an indication of a person's ability to change cognitive strategies in response to changes in their environment (Heaton et al., 1993).

1.4.1 The WCST in South Africa. The WCST is a well-validated neuropsychological test that is widely used in assessing executive functioning within the realm of neuropsychological assessment (Purdon & Waldie, 2001; Smith-Seemiller et al., 1997). This statement, however, does not entirely resonate with the South African context, since there are not any standardisations or norms available for the WCST in South Africa. The WCST makes use of cut-off scores to determine the level of executive functioning and this presents a difficulty as cerebral organisation differs between gender, handedness, age, education and/or experience (Kolb & Wishaw, 2008).

Thus, the aim of this study is to preliminary standardise the WCST for a specific South African population.

1.5 Motivation for the Proposed Research.

Before using any psychological test, it is of key importance to firstly evaluate the effectiveness of the test (Groth-Marnat, 2009). Groth-Marnat points out that one aspect to consider when evaluating a psychological test is whether the test has been adequately standardised. This includes answering the following questions:

- "Is the population to be tested similar to the population the test was standardized for?
- Was the size of the standardization sample adequate?
- Have specialized subgroup norms been established?
- How adequately do the instructions permit standardized administration?" (Groth-Marnat, 2009, p. 9).

When considering the WCST within the South African context, the answer to all four these questions are: No.

Pepping (2003) suggest that when assessing patients from various cultural, ethnic or linguistic groups, it is necessary to be familiar with literature on this topic and to seek out instruments that are standardised for the specific population in question. Groth-Marnat (2009) suggests consulting relevant literature when using an instrument to determine whether supplementary norms have been developed for specific groups after publication of the test manual. This information, unfortunately, does not exist for the South African context.

Various studies at the University of Limpopo (Medunsa Campus) have, however, been conducted to preliminary standardise some psychological tests. Among these are studies by Burns (1996), Endres (1996) and Phipps (1997) that were, among others, part of a project run by the Department of Clinical Psychology. Their aim was to standardise the WISC-R for a population that consisted of eight-year-old children from Ga-Rankuwa. This notion supports the feasibility of conducting preliminary standardisations and the need for standardisation of psychological test instruments in South Africa. This study will therefore aim to broaden previously conducted research projects by adding a neuropsychological test to the already established list of tests that was standardised for the specified Ga-Rankuwa population.

1.6 Research Outline

This study will conduct a preliminary standardisation of the WCST for a Setswana-speaking population of university students. In Chapter 2 a literature review will be conducted to explore the WCST and its relation to the frontal lobe and executive functioning, as well as its lack of standardisation in South Africa.

Chapter 3 will describe the outline of the research investigation. This will include the proposed research design and the actual research methodology that was followed. Chapter 4 will report on the results of the investigation that includes a description of the findings, conclusions and suggestions for future research.

And lastly, Chapter 5 will conclude the research by returning to the discussion of standardisation of the WCST for the specific South African population.

Chapter 2: Literature Review

This chapter will outline the major theoretical concepts involved in the proposed research project. The chapter will begin with a discussion on the principles of psychological assessment. This will be followed by the key principles involved in neuropsychological assessment, with special emphasis on the assessment of executive functions. Subsequently, a discussion on the Wisconsin Card Sorting Test (WCST) as an instrument to measure executive functions will be provided. Thereafter, the process and key concepts of standardisation will be elaborated on. This discussion will focus on the need for standardisation of the WCST and the suitability thereof for a South African population.

2.1 Psychological Assessment

According to Foxcroft and Roodt (2005), people have been trying to measure and predict human behaviour since ancient times. One area that has contributed to the measurement and prediction of behaviour is the field of psychological assessment.

It provides us with certain tools to measure behaviour. Foxcroft and Roodt (2005) describe psychological assessment as “a process oriented activity aimed at gathering a wide array of information by using assessment measures (tests) and information from many other sources” (pp. 3–4).

The process of engaging in psychological assessment can provide valuable information that could be used to assist individuals, groups and organisations in making informed decisions about their behaviour (Foxcroft & Roodt, 2005; Haynes, Richard & Kubany, 2005). This assessment can aid in identifying personal strengths and weaknesses, keep track of developmental progress, assist in determining the suitability of candidates for a job or study area, identify training or education needs, and even assist in making a clinical diagnosis (Foxcroft & Roodt, 2005; Haynes et al., 2005). In addition, it can also assist in the formulation of therapy goals, determining the effectiveness of an intervention, and to gather research data to assist in the growth of psychological knowledge or to influence policy decisions (Foxcroft & Roodt, 2005; Haynes et al., 2005). Haynes et al. (2005) therefore stress the importance of these instruments in the collection of data to make informed clinical judgments and decisions.

According to Foxcroft, Paterson, Le Roux and Herbst (2004) there are various challenges that psychological assessment currently faces in South Africa. Among which, is a need for quality tests that can yield valid and reliable results. Since the misuse of psychological assessment during the apartheid era, which led to a disproportionate representation of one sector of the population, it rendered tests invalid and unreliable for the majority of the South African population (Foxcroft & Roodt, 2005). Psychological assessment has therefore been widely criticised within the South African context (Foxcroft & Roodt, 2005). Consequently, a great need for appropriate measures to be developed in this country has arisen – specifically with the aim of ensuring that psychological instruments can be used with all cultural groups in South Africa (Foxcroft & Roodt, 2005).

However, the field of psychometrics (as a subfield of psychological assessment) is concerned with the validity, reliability and other psychometric characteristics of these tests (Foxcroft & Roodt, 2005; Zillmer, Spiers & Culbertson, 2004). This is achieved by compiling theory and research to ensure that the tests are applicable to the measurement of psychological characteristics (Foxcroft & Roodt, 2005; Zillmer, Spiers & Culbertson, 2004). Foxcroft and Roodt (2005) maintain that this entails the systematic and scientific development of tests, and the construction of psychometric properties for these tests.

Through the scientific development of tests, the subfield of psychological assessment has contributed to the recognition and development of the field of psychology as a whole. An example of its feasible success was seen during World War I, where "the success of psychological testing procedures to assess and select individuals to become officers and undertake special assignments in World War I was the impetus for some of the earliest recognition of psychology as a scientific field" (Zillmer et al., 2004, p. 66). Since then, the field of psychological assessment has greatly expanded. One such expansion includes the use of psychological assessment measures in the assessment of neurological damage and deficits. The following section will provide an overview of neuropsychological assessment, the use of neuropsychological tests, and considerations when interpreting these tests.

2.2 Neuropsychological Assessment

"Man's *[people's]* interest in the relationship between the brain and behaviour extends back at least 2 500 years, when Pythagoras argued that the brain was the site of human reasoning" [italics added] (Sbordone & Saul, 2000, p. 7). Today, the field neuropsychology continues in the quest of understanding the brain–behaviour relationship.

Neuropsychology is an expanding field of specialisation within the context of clinical psychology (Goldstein & McNeil, 2004). Neuropsychologists are primarily clinical psychologists that specialise in neuropsychological conceptualisations and methods (Zillmer et al., 2004). Neuropsychology draws from various fields that include anatomy, biology, biophysics, ethology, neurology, psychiatry, psychology, neurophysiology, neurochemistry, and neuropharmacology (Kolb & Wishaw, 2008; Siddiqui, Chatterjee, Kumar, Siddiqui & Goyal, 2008). Competence in the field thus requires competence in the fields of clinical practice, psychometrics, neuro-anatomy as well as neuropathology, and then the behavioural impacts thereof (Grieve, 2001).

A major aim of neuropsychology (and the neurosciences) is the explanation of how the various levels and components of the brain interact to control behaviour and generate emotion, language, cognition and conscious awareness (Cairns, 2004). Accordingly, the main focus of neuropsychology is to scientifically study human behaviour based on the function of the human brain (Kolb & Wishaw, 2008). In other words, neuropsychology explores the relationship between brain function and human behaviour (Kolb & Wishaw, 2008; Sbordone & Saul, 2000); also referred to as the brain–behaviour relationship (Zillmer et al., 2004).

Neuropsychological assessment aims to identify the behavioural impairment in behaviour that results from brain trauma or disease (Kolb & Wishaw, 2008; Sbordone & Saul, 2000). It thus examines the behavioural implications of dysfunctional brain structures (Lezak, Howieson & Loring, 2004; Sbordone & Saul, 2000): Sbordone and Saul (2000) refers to neuropsychological assessments as a

> […] complex process that the neuropsychologist goes through to reach opinion regarding the effect of a specific injury on the patient's cognitive and emotional functioning and on the patient's quality of life and ability to work, attend school, and perform activities of daily living (p. 133).

Neuropsychological assessment thus consists of using psychological tests and clinical observation in the assessment of neurological impairment (Kolb & Wishaw, 2008; Sbordone & Saul, 2000). It aims to identify the person's functional strengths and weaknesses, and to explain their behaviour according to their neuropsychological functioning (Goldstein & McNeil, 2004).

According to Lezak et al. (2004) neuropsychological assessment might serve one of these six purposes: diagnosis; patient care and planning; identifying treatment needs; evaluating the effectiveness of treatment; research; and forensic investigations. Kolb and

Wishaw (2008) and Orsini, Van Gorp and Boone (1988) also note that a thorough neuropsychological evaluation can provide the following benefits:

- Determining the effects of a brain injury on social, cognitive and emotional functioning.
- Describing the extent and quality of cognitive, emotional and/or motor impairment.
- Providing indicators of the person's recovery potential and course of recovery after injury.
- Providing a calculated indication of premorbid or baseline functioning.
- Assisting in the formulation of a differential diagnosis.
- Facilitating the localisation of the damage.
- Assisting in the planning of vocational, educational and rehabilitation programmes.
- Indicating a person's cognitive strengths and weaknesses.
- Monitoring of change through repeated behavioural assessments (this allows for the measurement of the impact of treatment, aging, progression or improvement in illness and developmental growth).
- Distinguishing between central nervous system (CNS) based disorders and disorders that are "functional" in origin allowing for the selection of the appropriate treatment methods.

Neuropsychological assessment has increasingly grown in popularity and scope (Orsini et al., 1988). And it no longer merely involves the assessment of cognitive abilities, but is increasingly playing a role in the rehabilitation and treatment planning of persons with cerebral pathology (Goldstein & McNeil, 2004). Neuropsychological assessment now also contributes to the understanding of the neuropsychological impact of psychiatric disorders and allows for the conceptualisation of psychiatric disorders in neuropsychological terms (Goldstein & McNeil, 2004). Goldstein and McNeil (2004) further state that neuropsychology has expanded its scope beyond the testing room to explore the implications of cognitive impairment in everyday life tasks. It therefore aims to become increasingly ecologically valid and environmentally based. As a result, neuropsychologists can no longer be labelled as "brain damage testers" or "lesion detectors" (Zillmer et al., 2004).

In conclusion of this section, literature indicates that neuropsychological assessment can be conceptualised as a method of studying the brain by observing its behavioural product (Zillmer et al., 2004). For this purpose, neuropsychologists employ a variety of methods to investigate brain function. Zillmer et al. list the application of neuropsychological tests as one of these methods. In order to facilitate a comprehensive understanding of the field of neuropsychological assessment, it is imperative to evaluate the use of neuropsychological tests.

2.2.1 Neuropsychological tests. As previously discussed, neuropsychology has greatly contributed to the field of psychology through the development of neuropsychological tests, which allow for the evaluation of brain functioning (Zillmer et al., 2004). "As with other psychological assessments, neuropsychological evaluations involve the intensive study of behaviour by means of standardized tests that provide relatively sensitive indices of brain-behaviour relationships" (Zillmer et al., 2004, p. 66). Because the behavioural manifestations of brain damage can be extremely heterogeneous (Groth-Marnat, 2009), neuropsychological tests may therefore allow for neurological disorders to become visible to the clinician, through the behavioural manifestations of the person being tested (Kolb & Wishaw, 2008). According to Zillmer et al. (2004), one cannot determine if a brain function is impaired without first testing that function. In this sense, neuropsychology greatly depends on the use of psychological tests that enable the investigation of the brain–behaviour relationship (Sbordone & Saul, 2000).

It is clear that neuropsychological tests involve the administration of standardised, scientifically validated tests to evaluate the behavioural functions associated with brain structures (Kolb & Wishaw, 2008; Sbordone & Saul, 2000; Zillmer et al., 2004). This entails the systematic administration of specific procedures in a structured environment, in order to identify a person's behavioural and cognitive strengths and limitations (Sbordone & Saul, 2000; Zillmer et al., 2004).

Neuropsychological tests are traditionally defined as instruments that are sensitive to brain damage, but recently its definition has expanded to also refer to a psychological instrument that reflects changes in brain in relation to, changes in testing behaviour and test performance (Zillmer et al., 2004).

However, measuring a complete neuropsychological construct or set of skills with one test is highly unlikely (Zillmer et al., 2004). For this reason, many neuropsychologists prefer to administer a set of different neuropsychological tests known as a test battery. In doing so, it enables the measurement of different brain–behaviour functions (Zillmer et al., 2004). Foxcroft and Roodt (2005) propose that test batteries should be altered, and tests selected, according to the needs of the person being tested in terms of age, level of ability or disability, and capacity in relation to the purposes of assessment. One aspect to consider in the selection of an appropriate neuropsychological test (to include or exclude from a battery), is whether the test is suitably standardised to allow for accurate comparisons to the standardisation sample (Goldstein & McNeil, 2004).

Neuropsychological tests cover a wide spectrum of skills ranging from motor speed, strength and coordination, to language functioning, memory, learning, attention, concentration, cognitive flexibility, constructional ability, abstract reasoning, problem solving, executive functions, judgment, academic functioning, intelligence, and personality

(Sbordone & Saul, 2000; Zillmer et al., 2004). These tests have been arranged into five measurement categories that cover the primary areas of cognitive functioning, namely *attention*, *language*, *memory*, *spatial skill*, and *executive functions* (Groth-Marnat, 2009).

The field of psychological assessment has therefore been substantially explored with specific focus on the subfield of neuropsychological assessment. Consequently, the discussion now shifts to focus more specifically, on the concept of *executive function* as one measurement area within the realm of neuropsychology.

2.3 Executive Function

2.3.1 Defining executive function. "A toddler forming a sentence. A kindergartener reciting the alphabet in order. A ninth-grader staying in sync with her clique. A college senior interviewing for a job. A writer formulating his next newspaper column. An entrepreneur strategizing the company's initial public stock offering. Despite the vast array of ages, tasks, and situations, these events all involve executive function, the mental process of planning and organizing flexible, strategic, appropriate actions" (Moran & Gardner, 2007, p. 19).

The study of executive function is concerned with how behaviour is executed (Lezak et al., 2004). It is a broad concept that includes sub-concepts such as anticipation, goal selection, planning, organisation, motivation, selection of a plan, and even self-evaluation of final performance (Goldstein & Green, 1995). Thus, executive function is seen as to include those skills that enable a person to effectively engage in independent, goal-directed, self-serving behaviour (Lezak et al., 2004). In defining executive function and its components, many different definitions emerge.

Kolb and Wishaw (2008) refer to executive function as a mental skill that allows a person to effectively deal with the problem solving requirements of everyday life. Sbordone and Saul (2000) similarly define it as an intricate process by which an individual negotiates problem-solving tasks from initiation to completion. This entails the ability to become aware of a problem, and evaluating and analysing that problem and its conditions.

According to Groth-Marnat (2009), executive function entails a person's capacity to "effectively regulate and direct self-behaviour" (p. 499). Goldstein and Green (1995) also describe it as comprising subtly interrelated behaviours such as planning, hypothesis testing, self-monitoring and the ability to use feedback.

Sherman, Strauss and Spreen (2006) define executive function more broadly to include "a multidimensional construct referring to a variety of loosely related higher order cognitive processes including initiation, planning, hypothesis generation, cognitive flexibility, decision making, regulation, judgement, feedback utilization, and self-perception that are necessary for effective and contextually appropriate behaviour" (p. 171). In other words, executive function thus includes planning, flexibility of thought, judgment, and goal-directed behaviour that can be summarised as the ability to make decisions and solve problems when dealing with unfamiliar situations (Sherman et al., 2006).

Moran and Gardner (2007), as with Groth-Marnat (2009), describe executive function as serving a regulatory function over a person's goal-directed behaviour by defining it as a cognitive process aimed at regulating behaviour and preparing a person for situations. "It contextualizes intended actions in light of past knowledge and experience, current situational cues, expectations of the future, and personally relevant values and purposes" (Moran & Gardner, 2007, p. 19). It is thus described as the ability to regulate behaviour within a fluctuating, unpredictable environment. Hughes and Graham (2008) again refer to executive function as a complex cognitive construct consisting of various processes that maintain controlled, goal-directed behaviour.

It is evident that different theorists define executive function in different manners, but for the purpose of this research, the definition used by Luria (1973) and Shallice (1982), as cited in Heaton, Chelune, Talley, Kay and Curtiss (1993, p. 1), will be used in order to ensure consistency with the testing instructions as set out in the testing manual. According to these two authors, executive function can be defined as the ability to develop and maintain an appropriate problem-solving strategy across changing stimulus conditions in order to achieve a future goal (Heaton et al., 1993).

2.3.2 Components of executive function. Despite the fact that a number of different definitions exist, the relative agreement subsists that "executive function" serves as an umbrella term that describes the intricate cognitive processes that maintain goal-directed behaviours (Meltzer, 2007). According to Meltzer (2007), most of these definitions include key elements such as

- goal setting and planning
- organisation of behaviours over time
- flexibility
- attention and memory systems to guide the above mentioned processes, and
- self-regulatory processes including self-monitoring and inhibitory control.

Moran and Gardener (2007) suggest describing executive function as the integration of three parameters: the *hill; skill;* and *will of behaviour*. They describe *hill* as the setting of a clear goal, the purpose towards which abilities and efforts are directed; *skill* as the learned sequences of behaviour, abilities and techniques required to achieve the goal; and *will* as the volition, effort and motivation to begin and persevere until the goal has been achieved.

Lezak et al. (2004) describe executive function in terms of four components: *volition*; *planning*; *purposive action*; and *effective performance*.

Volition refers to the intricate process by which a person decides what they need or want and how they conceptualise an idea of how they are going to meet the need or want. It requires a number of prerequisite skills, including the setting of a goal or intention, motivation and the ability to initiate behaviour as well as psychological, physical and spatial awareness (Lezak et al., 2004). When a person lacks volitional capacity they merely cannot think of anything to do. This can come across as apathy or the absence of the sense that they are a separate, unique individual (Lezak et al., 2004). They may be capable of completing activities, but only in response to being told to complete the activity (Lezak et al., 2004). Thus, these individuals can be said to be lacking in spontaneous behaviour initiation.

Planning entails the identification and organisation of stages and elements such as abilities, material, or persons needed to execute the intention or goal as set out in volition (Lezak et al., 2004). According to Lezak et al. (2004), the ability to plan depends on the ability to

- conceptualise changes from present circumstances
- deal objectively with oneself in relation to the environment
- generate alternatives, evaluate the alternatives, and make choices
- maintain sequential and hierarchical thought required for the development of a conceptual framework to guide the execution of the plan
- keep impulse control intact
- keep memory skills relatively intact, and

- sustain attention.

Purposive action refers to the transformation of a goal, intention, or plan to a dynamic activity (Lezak et al., 2004). According to Lezak et al. (2004), this necessitates an integrated, organised process by which a person initiates, maintains, switches, and ends complex sequences of behaviour. Subsequently, this programming of activity is required for the effective execution of plans despite motivation, knowledge, or capacity to execute the behaviour.

Effective performance depends on the task-performing person's ability to monitor, self-correct, and manage the force, pace and other aspects of their behaviour. This includes the ability to perceive errors as well as to correct those errors (Lezak et al., 2004).

Hughes and Graham (2008) reduce executive function to three factors: *cognitive flexibility*; *inhibitory control*; and *planning*. As planning has already been discussed by Lezak et al. (2004), the focus on Huges and Graham's (2008) factors will only consist of a description of cognitive flexibility and inhibitory control.

Cognitive flexibility refers to the skill of exhibiting a "win–stay, lose–shift" strategy (Goldstein & Green, 1995). This includes the ability to generate hypotheses and to shift concepts. Cognitive flexibility can therefore be described as the ability to view objects or events from different angles, especially within varying contexts (Sherman et al., 2006). A subset of this ability is reactive flexibility; the ability to realign behavioural tendencies in different contexts (Sherman et al., 2006).

Inhibitory control refers to the withholding of a response (Hughes & Graham, 2008). This includes the ability to stop or alter on-going behaviour (Lezak et al., 2004). Difficulties in maintaining inhibitory control may present as impulsivity, over-reactivity or perseveration (Lezak et al., 2004).

From the discussion above, it is clear that executive function depends on a variety of interdependent skills. Yet, this wide array of skills cannot be localised to only a certain area of the brain; some cortical areas have been proven to play a greater role in executive function than others. As previously stated, neuropsychologists require knowledge of neuro-anatomy to study the executive function more closely (Sbordone & Saul, 2000). The following section will provide a simplified discussion of the brain by focussing on the anatomy of the cerebral cortex, and more specifically the frontal lobe and its relationship to executive function.

2.3.3 Anatomy of executive function.

> The brain's functions are both mysterious and remarkable. From the brain come all thoughts, beliefs, memories, behaviours and moods. The brain is the site of thinking and the control centre of the rest of the body. The brain coordinates the abilities to move, touch, smell, taste, hear and see. It enables people to form words, understand and manipulate numbers, compose and appreciate music, recognize and understand geometric shapes, communicate with others, plan ahead, and even fantasize.
> (Beers, 2003, p. 390).

In order to fully understand the anatomy of executive function, a brief description of the functioning of the brain is required.

The brain can be seen as an organ that is organised in layers (Cairns, 2004; Preston, O'Neal & Talaga, 2009). A major aim of the neurosciences is the explanation of how the various layers and components of the brain interact to control behaviour and generate emotion, language, cognition and conscious awareness (Cairns, 2004). Each layer is controlled in one or more of the three main parts of the brain: *the cerebrum* (also called the cortex); *the cerebellum*; and *the brainstem* (Beers, 2003). The literature review of this study only requires a review of the cerebrum.

The cerebrum is a dense tissue mass that contains the majority of the nerves that the nervous system is comprised of (Beers, 2003). The cerebrum embodies the greatest mass of the brain (Cairns, 2004; Zillmer et al., 2004), and is seen as the part of the brain with highest level of functioning; it is this most developed part of the brain that sets humans and primates apart from other animals (Preston et al., 2009). The cerebrum is what allows human perception, information processing, thinking, reasoning, and other higher cognitive functions that include intelligence, language, judgment, memory, problem solving, social behaviour, and most importantly, executive function (Preston et al., 2009; Sbordone & Saul, 2000). In essence, the cerebrum is responsible for higher human mental functions and the unique nature of human cognition (Zillmer et al., 2004). According to Zillmer et al. (2004) the functions of the cerebrum can thus be summarised as: the analysis of sensory input; the organisation, integration and storage of information; and the direction of motor output.

The cerebrum can then be divided into two hemispheres that are connected by dense fibres called the *corpus callosum* (Beers, 2003; Kolb & Wishaw, 2008, Sbordone & Saul, 2000; Zillmer et al., 2004). Each hemisphere is comprised of folds of white and grey matter, with the white matter being myelinated axons, and the grey matter that consists of neurons (Cairns, 2004; Kolb & Wishaw, 2008; Zillmer et al., 2004). In most instances, the two hemispheres control behaviours on the contra-lateral side of the body (Beers, 2003; Kolb & Wishaw, 2008; Sbordone & Saul, 2000; Zillmer et al., 2004).

Each hemisphere of the cerebrum is also further divided into four lobes, each associated with specific functions. The four lobes of each hemisphere are the *temporal*, *parietal*, *occipital* and *frontal* lobes (Beers, 2003; Kolb & Wishaw, 2008; Sbordone & Saul, 2000; Zillmer et al., 2004).

The *temporal lobe* is involved in the generation of memory and emotions; the fixation of current events into recent and long-term memory; the retrieval of long-term memories; and the reception and interpretation of auditory input allowing the facilitation of the understanding of sounds and images that enable individuals to recognise people and objects visually, as well as to integrate hearing and speech (Beers, 2003; Kolb & Wishaw, 2008).

The *parietal lobe* is responsible for the interpretation of sensory information from the body as well as the control of movement; it translates sensory stimuli to general perceptions (Beers, 2003; Cairns, 2004; Kolb & Wishaw, 2008). This lobe is also the storage site for spatial memories that allow a person to orientate himself in space; to have awareness of the placement of their body parts in space (Beers, 2003; Cairns, 2004; Kolb & Wishaw, 2008).

The *occipital lobe*, is the smallest of the four lobes, and is involved in the processing and interpretation of visual stimuli (Beers, 2003; Cairns, 2004; Kolb & Wishaw, 2008; Zillmer et al., 2004). This includes the forming of visual memories and integrating visual information with spatial information that is received from the parietal lobe (Beers, 2003; Kolb & Wishaw, 2008).

The *frontal lobe* is the largest of all the lobes, comprising approximately a third of the cerebrum when both hemispheres are placed together (Cairns, 2004; Sbordone & Saul, 2000). According to Preston et al. (2009), the frontal lobes are thought to be the last addition to and most developed part of the cerebrum. The frontal lobe is unique in its development in that it contains three sub-cortices within the lobe: the *premotor*, *supplementary motor* and the *prefrontal cortex* – with the prefrontal cortex forming the front half-portion of the frontal lobe (Preston et al., 2009; Zillmer et al., 2004).

The frontal lobe is a highly complex structure involved in a much wider range of functions (Heaton et al., 1993). It is considered to be the coordinator or executive of the brain that manages and controls behaviour and decision-making based on the suitability of social behaviour (Zillmer et al., 2004). The frontal lobe is associated with the inhibition of emotional reactions, the maintenance of attention and concentration, as well as the thinking of complex thoughts and the solving of problems (Preston et al., 2009). It is therefore fitting that Siddiqui et al. (2008) describe the frontal lobe as the mediator of the cognitive, affective and emotional aspects of a person's personality.

The frontal lobe is also involved in the initiation, planning, execution and regulation of complex voluntary movements and motor behaviour that is learned (Beers, 2003; Sbordone & Saul, 2000). Similarly, complex intellectual processes such as speech, thought, concentration, problem solving, and cognitive planning therefore stems from the frontal lobe. And consequently, behaviours involved in emotional expression such as facial expressions and gestures, as well as moods and feelings are controlled by the frontal lobe (Beers, 2003).

The *prefrontal cortex* is associated with the prediction of the consequences of behaviours (Zillmer et al., 2004). Moreover, it is reported to be responsible for the unique nature of a person's personality and their conscience (Zillmer et al., 2004).

Although it has been stated that each lobe is responsible for unique and specific behavioural repertoires, the higher mental functions, such as executive functions, emotional functioning, and social cognition, however, rely on the interdependence of all of the functions of the various lobes (Zillmer et al., 2004). In so doing, higher order processes require a dynamic, integrative process in which the various lobes function with one another to produce a specific function. As a result, the frontal lobes are therefore not *solely* responsible for executive function, but play a prominent role in its working (Heaton et al., 1993).

Due to irregularities of the skull's bone, the frontal and temporal lobes are especially sensitive to damage due to head injury (Orsini et al., 1988). For this reason, it is of key importance to assess the functions associated with these lobes, because damage to the frontal lobe is specifically associated with impairment of executive function (Orsini et al., 1988; Kolb & Wishaw, 2008). The following section will discuss impairment of executive function accordingly.

2.3.4 Deficits in executive function. "When executive functions are impaired, the individual may no longer be capable of satisfactory self-care, of performing remunerative or useful work independently, or of maintaining social relationships" (Lezak et al., 2004, p. 35).

Patients with deficits in executive function present a variety of difficulties that can range from a semi-vegetative state (in which activity is hardly ever initiated), to a lack of awareness of their impact on others, and to an inability to effectively regulate behaviour (Groth-Marnat, 2009). Deficits in executive function can thus be perceived as changes in personality and social conduct, which, in turn, may result in a significant decrease in the person's capacity to function independently (Cairns, 2004). This may include mutism or attenuated speech and perseveration of motor skills (Cairns, 2004). As a result, deficits in executive function can be some of the most prominent barriers to vocational reintegration (Purdon & Waldie, 2001).

Executive difficulty results most typically from frontal lobe damage, but can also result from damage to the sub-cortical regions (especially thalamic structures) or diffuse damage (Groth-Marnat, 2009). Persons with damage to the frontal lobes often present very intense emotional responses, specifically irritability and anger; they may also present difficulty with sustaining attention and concentration (Preston et al., 2009). And according to Preston et al. (2009), this could have severe implications on academic or occupational functioning.

The frontal lobe (specifically the prefrontal cortex) is also often indicated as an area of concern with attention-deficit hyperactivity disorder (ADHD) (Preston et al., 2009). Damage to the prefrontal cortex may result in impaired concentration, problem-solving ability, planning and judgment (Zillmer et al., 2004). In addition, damage to the lateral prefrontal cortex can result in a lack of motivation and a decrease in vitality (Preston et al., 2009).

What is more, damage to the frontal lobe may even result in impairment of the execution of voluntary movement on the contra-lateral side of where the damage actually occurred. This may be observed as spastic limbs on the contra-lateral side, or hyperactive reflexes on the ipsi-lateral side of the body (Sbordone & Saul, 2000).

According to Lezak et al. (2004), the clinical presentation of executive dysfunction can be summarised as follows: impaired self-control or self-direction which may be experienced as emotional flatness; increased irritability and excitability; impulsive behaviour; erratic carelessness; rigidity; difficulty making attention and behavioural shifts; difficulty initiating action; decreased motivation; impaired planning ability; as well as sequencing difficulty and perseveration.

Furthermore, executive dysfunction has been implicated in various childhood difficulties such as ADHD (as mentioned above), antisocial behaviour, and autism (Hughes &

Graham, 2008). Executive dysfunction and its relation to the frontal lobe have also been implicated in the disorganised behavioural symptoms of schizophrenia, and various other psychiatric disorders (Preston et al., 2009). Measuring the functions associated with the prefrontal cortex can thus be greatly beneficial to a person being tested (Preston et al., 2009). "Studies that deepen our understanding of normative age related improvements in executive function may therefore help to identify children with poor regulatory control who could benefit from intervention programs, and so has clear societal importance" (Hughes & Graham, 2008, p. 278). The following section will discuss the issues that regard the measurement of executive function.

2.3.5 Measuring executive function. Executive function is one of the primary areas of neuropsychological assessment that clinicians focus on when compiling a neuropsychological test battery (Pepping, 2003). However, assessing problem solving and executive function is one of the most challenging tasks for a clinician when evaluating behaviour (Goldstein & Green, 1995).

Executive difficulties are often overlooked during formal assessment since these difficulties can occur whilst other cognitive abilities are still intact (Groth-Marnat, 2009; Lezak et al, 2004; Sherman et al., 2006). Persons with frontal lobe damage or executive dysfunction may perform adequately on structured psychometric tests, but these formal assessments are usually very structured, leaving little room for inappropriate behaviour and not allowing enough opportunity to display difficulty in initiation, planning and judgment of behaviour (Groth-Marnat, 2009; Lezak et al., 2004; Orsini et al., 1988; Sherman et al., 2006). Executive dysfunction can thus frequently be misdiagnosed as depression, since apathy; lack of affect and lack of direction can similarly result from brain impairment (Groth-Marnat, 2009). Executive difficulty may even be mistaken for non-involvement, laziness or being spoilt (Lezak et al., 2004).

For this reason, it is important to include tests that are especially sensitive to frontal lobe and/or executive dysfunction in a test battery (Orsini et al., 1988). Neuropsychologists have consequently developed creative assessment procedures for evaluating executive abilities, and literally dozens of tests attempt to measure this neuropsychological domain (Zillmer et al., 2004), although no single test measures its full complexity (Goldstein & Green, 1995). Assessment of executive function thus requires a comprehensive approach that makes use of various methods, including detailed analysis of underlying impairments (Goldstein & Green, 1995).

Tests that can therefore be included in the battery to assess executive function include the Auditory Consonant Trigram Test, Rey Tangled Lines, the Porteus Maze Test, the Stroop Test, Trail Making Test A and B, Verbal Fluency, and Finger Tapping (Orsini et al., 1988); the Trail Making Test B (part of the Halstead-Reitan battery), the Tower of London-Drexel University, and the WCST (Zillmer et al, 2004).

Pepping (2003), Heaton et al. (1993), and Hughes and Graham (2008) suggest making use of the WCST for the purpose of testing, as it provides a relatively sensitive measure of executive function.

This discussion specifically focused on defining executive function and its components, as well as the anatomical associates of executive function. This led to a discussion of the deficits that are evident when executive function is impaired as a result of damage to these associates, and concluded with a look at the complexities of measuring executive function. The following section will explore the WCST as an instrument that can be employed in the measurement of executive function.

2.4 The Wisconsin Card Sorting Test (WCST)

The WCST is a neuropsychological test that is extremely sensitive to frontal lobe lesions (Heaton et al., 1993; Kolb & Wishaw, 2008; Purdon & Waldie, 2001; Zillmer et al., 2004). It consists of four stimulus cards and 128 response cards that are printed with one to four symbols each that can include stars, triangles, crosses, or circles and can be blue, green, yellow, or red – no two cards are alike (Heaton et al., 1993; Sbordone & Saul, 2000; Zillmer et al., 2004).

The stimulus cards are placed horizontally in front of the person being tested in a standardised manner (Heaton et al., 1993; Sbordone & Saul, 2000). The person is then required to sort the response cards according to an underlying sorting principle that they are expected to deduce from the examiner's "right" or "wrong" responses to the their card placements (Heaton et al., 1993; Horton & Soper, 2008; Hughes & Graham, 2008; Sbordone & Saul, 2000; Zillmer et al., 2004). After a set amount of correct responses (ten correct responses), the underlying sorting principle changes, and the person is then required to determine the new sorting principle (Sbordone & Saul, 2000).

Patients with damage to the frontal lobe may often continue sorting according one sorting principle despite feedback from the examiner who indicates that the sorting is incorrect (Sbordone & Saul, 2000). The WCST thus provides an indication of one aspect of frontal lobe activity by measuring executive function (Heaton et al., 1993; Kolb & Wishaw, 2008).

Higher mental functions – such as executive functions, emotional functioning, and social cognition however, rely on the integration of functions of the various lobes (Zillmer et al., 2004). As a result, it is important to note that labelling the WCST as a measure of frontal or prefrontal functioning is a generalisation. The frontal lobes are highly complex structures involved in a much wider range of functions than only those assessed by the WCST. Although the WCST is extremely sensitive to frontal lobe conditions, any medical or psychological condition affecting executive function may have an impact on the WCST score. For this reason, it is recommended that the WCST is used within the context of a full neuropsychological evaluation that integrates neuropsychological, medical, psychosocial and historical information (Heaton et al, 1993).

According to Heaton et al (1993), Sbordone and Saul (2000), and Zillmer et al (2004), the WCST measures the following domains: problem-solving skills; abstract reasoning ability; cognitive flexibility; the ability to maintain cognitive set; and concept formation.

Only in this manner, can the WCST be seen as a measure of executive function (Heaton et al., 1993). The WCST is thus widely used to measure abstract behaviour and shifting of set (Smith-Seemiller et al., 1997; Zillmer et al., 2004), and it provides an indication of a person's ability to change cognitive strategies in response to changes in their environment (Heaton et al., 1993).

According to Cinan and Tanor (2002), different types of executive processes are involved in the completion of the WCST. Successful completion of the WCST requires a person to display strategic planning, organised searching, feedback utilisation to shift mental set, behaviour direction towards goal achievement, and impulse control (Heaton et al, 1993).

Successful completion of the WCST also requires the person being tested to deduce the correct sorting principle from the feedback that is given by the examiner (Heaton et al, 1993). The person is then required to follow through with one sorting principle or set across various different stimulus conditions, and to ignore stimulus dimensions that are not required by the current sorting principle (Heaton et al., 1993). When the person is unable to do this, and completes five or more correct sorts followed by an error before completing ten correct sorts, the client is said to have failed to maintain the set (Heaton et al., 1993).

If the person being tested completes ten correct sorts and the sorting principle changes, the person is required to inhibit the tendency to perseverate and keep sorting according to the previous principle (Heaton et al., 1993). The person must adapt by using the feedback to deduce the new sorting principle. The WCST thus provides an indication of cognitive set shifting (Sinclair & Taylor, 2008).

Poor performance on the WCST suggests that the person being tested may have difficulty in organising their behaviour, or may have difficulty in translating rules from one situation to the next (Zillmer et al., 2004). Poor performance may, however, also be accounted for by visual difficulties, colour blindness, visual–perceptual impairment, hearing impairment, psychosis, depression, anxiety, low levels of motivation, or non-involvement (Sbordone & Saul, 2000). Sbordone and Saul (2000) therefore stress the key importance of the WCST to be interpreted within the context of a complete neuropsychological interview and assessment battery.

The WCST is set apart from other measures of abstract reasoning by providing more than an overall indication of success (Heaton et al., 1993). It also provides an indication of the source of difficulty on the task, such as ineffective initial conceptualisation, inability to maintain cognitive set, perseveration, or ineffective learning across the test (Heaton et al., 1993). The sub-scores that provide a more detailed picture of performance on the WCST will now be discussed.

2.4.1 Sub-scores of the WCST. According to Heaton et al. (1993) and Sbordone and Saul (2000), a person's performance on the WCST is indicated by the following scores that allow analysis of the individual nature of the person's executive abilities:

- The number of categories sorted before the depletion of the stack of cards.
- The total numbers of errors made by the person being tested.
- The percentage of correct responses attained.
- The number of perseverations.
- The number of perseverative errors.
- The number of trials required to complete the first category.
- The percentage of conceptual level responses.
- The number of failures to maintain set errors.
- The ability to reduce errors over trials (also called learning to learn).

Each of the scores an individual may obtain after completion of the WCST will now be discussed individually.

2.4.1.1 Number of trials administered. The *Number of trials administered* score is merely calculated by counting the amount of cards that the person being tested has sorted (Heaton et al., 1993). This provides an indication of whether the test was fully administered or whether the person terminated the test prematurely. It is also used to calculate the percentage of scores of some of the other WCST scores (Heaton et al., 1993).

2.4.1.2 Number of categories completed. The *Number of categories completed* score refers to the number of sets (ten consecutive correct responses) that have been completed during the administration (Heaton et al., 1993).

2.4.1.3 Trials to complete first category. The *Trials to complete first category* score is calculated by adding the number of cards the client sorted before completing the first set of ten consecutive correct scores (Heaton et al., 1993). This score provides a measure of initial conceptualisation before the shifting of set is also required (Heaton et al., 1993).

2.4.1.4 Failure to maintain set. When a participant completes five or more consecutive correct matches, followed by an error before reaching a complete set of ten correct responses, the error is termed a *Failure to maintain set* (Heaton et al., 1993). Lezak et al. (2004) propose that the inability to maintain the set is as a result of an impaired capacity to pay the required attention to the task at hand. In this case, persons with frontal lobe pathology usually show an elevated number of random errors that imply continual shifts or variations in the principle to which they sort (Barcelo & Knight, 2001).

2.4.1.5 Learning to learn. "*Learning to learn* reflects the client's average change in conceptual efficiency across consecutive categories of the WCST" (Heaton et al., 1993, p. 13). This score consequently provides an indication of whether the person being tested improves as the test progresses. A *Learning to learn* score, however, can only be calculated when a minimum of three categories have been attempted, of which at least two were fully completed (Heaton et al., 1993). The *Learning to learn* score is calculated by adding the differences between the *Percent errors* scores of each set and then calculating the average of these scores. A positive score shows that the person has improved in efficacy across different categories, with the assumption that this may be ascribed to learning (Heaton et al., 1993). Heaton et al. (1993) note, however, that a negative *Learning to learn* score occurs more commonly. Yet, they also note that this score can be compared to the norm group to identify a relatively good performance even though the score is negative.

2.4.1.6 Total number correct. The *Total number correct* score is calculated by adding the number of trials that the person being tested has sorted correctly (Heaton et al., 1993).

2.4.1.7 Percent perseverative errors. The *Percent perseverative errors* score provides an indication of the concentration of perseverative errors in relation to the total number of responses (Heaton et al., 1993). A perseverative error occurs when a person persists in sorting the cards to an incorrect principle, despite feedback that the sorting is incorrect (Heaton et al., 1993). Persons with frontal lobe pathology usually show an abnormally elevated number of perseverative errors (Barcelo & Knight, 2001; Kolb & Wishaw, 2008). It can therefore be said that the number of perseverative errors is associated with the person's ability to switch between cognitive strategies (Heaton et al., 1993; Kolb & Wishaw, 2008).

2.4.1.8 Percent conceptual level responses. The *Percent conceptual level responses* score is presumed to be an indicator of the tested person's insight into the correct sorting principle. This is due to the fact that some insight awareness of the correct sorting strategy is needed to achieve three or more correct matches following each other (Heaton et al., 1993). Consecutive correct responses that occur in sets of three or more are referred to as *Conceptual level responses* (Heaton et al., 1993).

The final three scores to be discussed, *Percent errors*, *Percent perseverative responses*, and *Percent non-perseverative errors* are primarily intended to provide information for research purposes and are not recommended for interpretation for clinical purposes (Heaton et al., 1993).

2.4.1.9 Percent errors. The *Percent errors* score is the number of incorrect responses calculated as a percentage of the *Total number of trials* administered (Heaton et al., 1993).

2.4.1.10 Percent perseverative responses. The *Percent perseverative responses* score is an indication of the frequency with which the person being tested perseverated in a sorting principle, despite the sorting being either correct or incorrect (Heaton et al., 1993).

2.4.1.11 Percent non-perseverative errors. The *Percent non-perseverative errors* score refers to the percentage of errors the person being tested made that were not a function of perseveration (Heaton et al., 1993).

In view of the above, the WCST provides information on a variety of dimensions. The confidence with which these scores can be interpreted, however, depends on the accuracy with which the WCST can measure these scores. The accuracy of a test, moreover, can be determined by examining the stringency with which it was standardised. For this reason, the concept of standardisation will be discussed, followed by a discussion of the standardisation of the WCST.

2.5 Standardisation

The standardisation of tests is associated with the representation of psychological data. "Every psychological observation can be expressed either numerically as quantitative data or descriptively as qualitative data" (Lezak et al., 2004, p. 134).

As such, a qualitative interpretation requires close observation of both the processes the person being tested employ to solve tasks and the various behavioural factors that may influence the outcome of the test (Lezak et al., 2004; Orsini et al., 1988). This approach focuses on documenting rich clinical observation without objective standardisation (Lezak et al., 2004). The qualitative approach focuses on understanding *why* a person failed at a given task, allowing for a greater degree of insight than merely considering the fact that he or she failed (Zillmer et al., 2004).

In contrast to this, the quantitative interpretation focuses on constructing numerical data aimed at comparing scores where these scores are seen as a summary of the observed behaviour (Lezak et al., 2004). The quantitative approach is thus concerned with the generation and analysis of numerical scores. Interpretation of the WCST largely depends on a quantitative interpretation approach and is thus reliant on the generation of these scores.

Scores allow for a variety of behavioural observations to be condensed to a single, objective score (Lezak et al, 2004). As a result, scores can be seen as a summary of behavioural observations that are organised according to certain categories.

To obtain these scores, a test like the WCST is administered after which the results are tallied to obtain a raw score. This score is calculated by adding all the correct scores and subtracting the number of incorrect scores that were completed (Lezak et al., 2004). The raw scores are then converted to standard scores before interpretation (Foxcroft & Roodt, 2005; Orsini et al., 1988). Standard scores facilitate consistent interpretation and meaningful comparisons that allows the clinician to determine how much a person's total score actually differs from the average performance (Foxcroft & Roodt, 2005; Lezak et al., 2004; Orsini et al., 1988).

Once the scores have been converted to standard scores, a decision needs to be made whether the score is consistent with impaired functions or not. In the case of the WCST and many other neuropsychological tests, this is done by means of cut-off scores (Groth-Marnat, 2009; Kolb & Wishaw, 2008; Zillmer et al., 2004). When a person scores less than the cut-off score, their abilities are labelled as *impaired*; if the person scores above the cut-off score, their abilities are labelled as *within the normal range* (Zillmer et al., 2004). The cut-off score can thus be seen as a standard comparative score (Foxcroft & Roodt, 2005).

The use of standard scores and cut-off scores as comparative methods rely on the consistency between test administrations and the specific comparison to a population or standard (Groth-Marnat, 2009). This is exactly what the process of standardisation is aimed at.

Standardisation entails the systematic and scientific development of tests, as well as constructing adequate norms for the comparison of individual scores (Foxcroft & Roodt, 2005; Groth-Marnat, 2009). The term *standardised test* refers to a set of tasks that are designed to assess some aspect of psychological functioning that is completed under standard conditions (Foxcroft & Roodt, 2005; Zillmer et al., 2004). A standardised test is thus developed in such a way that the test items, administration, scoring, and interpretation procedures remain consistent every time the test is used to allow comparison between scores (Foxcroft & Roodt, 2005; Groth-Marnat, 2009).

In light of this, an assessment can be seen as a scientific experiment that constantly demands the same rigorous control (Foxcroft & Roodt, 2005). Standardised testing can thus elicit samples of behaviour in a standardised, replicable manner (Lezak et al., 2004). These samples need to be elicited in similar testing situations for each person being tested (Lezak et al., 2004). Lezak et al. (2004) thus note that this allows the clinician to compare behavioural samples across individuals, over time, or with expected levels of achievement. In turn, this then allows the clinician to extrapolate how the person being tested may function in a real life situation based on the behavioural sample compared to the expected test results.

The aim of standardisation can therefore be seen as the structuring of tests in order to compare different persons' scores. According to Groth-Marnat (2009), this is done by considering the construction of

- uniform administration and scoring instructions,
- adequate norms for the sample population,
- the psychometric properties for the population.

Each of these considerations will be discussed, firstly looking at the importance of standardised instructions.

2.5.1 Standardised instructions. As previously discussed, a standardised test is a task or set of tasks that are administered under controlled conditions and that remain consistent over every assessment situation. This allows for consistency between scores, which in turn, allows for systematic comparisons to be made (Zillmer et al., 2004).

Following this, it can be said that one aspect of standardisation comprises the procedures with which tests are administered (Groth-Marnat, 2009). A well-standardised test has instructions that allow clinicians to administer a test in a structured, consistent manner on various different occasions (Foxcroft & Roodt, 2005; Groth-Marnat, 2009). As a result, all administrations of tests should be conducted in the same manner (as close as possible); every person being tested should be given the same instructions and format of the test (Groth-Marnat, 2009; Foxcroft & Roodt, 2005). This often involves the clinician reading the test instructions verbatim and ensuring that time limits are kept (Foxcroft & Roodt, 2005). This necessitates clear administration procedures in the testing manual.

In addition, clear administration procedures further prescribe standard testing conditions that ensure that all administrations are conducted in the same manner as to allow the comparisons of scores between different administrations (Lezak et al., 2004).

Standardised instruction should thus not only specify the test instructions, but also ensure a consistent testing environment that is characterised by good lighting, no interruptions, a quiet space and adequate rapport in the assessment relationship (Groth-Marnat, 2009). Lezak et al. (2004) add that instructions should specify the presentation of the test items specifically focused on word usage and the handling of the material.

If any deviations are made from the standardised procedure, it needs to be noted and taken into consideration when interpreting the results (Foxcroft & Roodt, 2005). Deviating from the standardised procedures in any way, may alter the results obtained from the test and render the results invalid (Groth-Marnat, 2009). Therefore, if the tests are administered in different ways by different clinicians, valid comparisons cannot be made between the scores of the persons that were tested.

2.5.2 Standardised scoring. In addition to standardised instruction and administration, the consistency between scores further depends on the standardised manner in which test protocols are scored (Groth-Marnat, 2009). As with standardised administration, consistent, clear scoring methods are required to maintain consistency of scores.

Here computer-assisted scoring may aid in standardising the scoring of a test. "During the past 30 years, computer-assisted assessment has grown exponentially" (Groth-Marnat, 2009, p. 62). According to Groth-Marnat (2009), computer-assisted scoring presents a number of advantages, among which is the enhancement of clerical efficiency in test scoring.

Computerised scoring also shortens the administration and scoring time (Foxcroft & Roodt, 2005; Groth-Marnat, 2009). It has even shown to have increased the test-retest reliability of an instrument, minimise the possibility of examiner bias, and reduce the cost of assessment by enhancing the speed and efficacy thereof (Groth-Marnat, 2009).

Again, according to Groth-Marnat (2009), research on the reliability of computer-assisted assessment indicated that it generally displayed excellent reliability at least equal to the paper-and-pencil version of tests. Groth-Marnat further adds that research shows insignificant differences between scores obtained from computerised versions in comparison to scores obtained from the paper-and-pencil versions. “This finding supports the view that if a paper-and-pencil version of the test is valid, a computerized version will also have equal validity resulting from the comparability in scores” (Groth-Marnat, 2009, p. 63).

Within the South African context, computerised testing may, however, present difficulty, as a great deal of the population is not computer literate (Foxcroft & Roodt, 2005). Persons not confident with the use of a computer may thus present with increased levels of test anxiety and require extra practice time in order to gain confidence with the use of a computer (Foxcroft & Roodt, 2005). When making use of computerised assessment, care needs to be taken to consider a person’s level of computer literacy.

When taking computer literacy into consideration, computerised scoring may aid in standardising the scoring procedures of a test, and eliminate errors due to human factors.

In summary, the comparisons of test scores are reliant on the detail with which tests are standardised in terms of administered and scoring procedures, and the precision with which these instructions are executed. The greater the consistency with which the test is administered and scored, the greater the confidence with which scores can be compared. And highly standardised tests are essential to the comparison of scores using normative data (Lezak et al., 2004). Standardising the procedures of a test allows comparison of scores; this is done by means of comparison to a normative standard.

2.5.3 Normative data. "A fundamental tenet in determining patients' cognitive status is the comparison of individual test scores obtained from a normative population; when patients' performances are significantly worse than those from the matched normative cohort, they may be considered to reflect brain pathology" (Brickman, Cabo & Manly, 2006, p. 91*)*.

As previously discussed, the quantitative interpretation of neuropsychological assessment and the conversion of raw scores to standard scores is largely dependent on reference to normative data (Grieve, 2001), because score conversions are made on the basis of reference to a standardisation population (Lezak et al., 2004).

The standardisation population refers to the group of persons assessed in order to construct normative data for the instrument (Lezak et al., 2004). Norms provide an indication of the distribution of scores within the standardisation sample (Groth-Marnat, 2009). Normative standards are thus derived by calculating the average score of a population deemed suitable for comparison. In this manner, norms allow comparison of a single score to all the scores of the standardised population (Lezak et al., 2004).

Constructing norms for a test requires administering the test to a large sample by means of stratified sampling (Orsini et al., 1998). Stratified sampling entails the structuring of the sample according to selected variables (Whitley, 2002). This is done in order to obtain a sample that is reflective of the variables in the population being studied. Some of the variables to be stratified for are age, sex, race, geographical region, occupation, education, and urban–rural residential status (Orsini et al., 1988). The sample would then be constructed by setting quotas to be met for each of the above variables (Whitley, 2002).

Normative samples are usually described in terms of gender, race, age and/or educational level, and according to Lezak et al. (2004) a large body of evidence indicates that demographic variables, most notably age and education, and in some other tests gender and race, influence performance on neuropsychological tests. Most tests, however, do not provide separate norms for age, sex and education, despite the fact that these have been shown to impact on test performance – and only a few tests take geographical distribution into consideration (Lezak et al., 2004). As test development is growing in sophistication, more variables are being taken into account when constructing norms (Lezak et al., 2004).

Normative data may thus differ for the same instrument depending on different researchers who collect data on small samples (Orsini et al., 1988). Norms obtained from different samples, different areas, and for different reasons can provide very different results (Lezak et al., 2004). "The normative information used, is limited to the characteristics of the normative sample" (Foxcroft & Roodt, 2005, p. 5). If there is a similarity between the person being tested and the standardisation sample, accurate comparisons can be made (Crawford, 2004; Goldstein & McNeil, 2004; Groth-Marnat, 2009). The greater the difference between the person and the standardisation group, the less useful the test will be for the evaluation of that person's score (Crawford, 2004; Groth-Marnat, 2009). As a result, a person's test scores only have meaning in comparison to the norms of the standardisation sample (Groth-Marnat, 2009; Orsini et al., 1988; Lezak et al., 2004). When interpreting test results, it should thus be considered that the variations in tests scores may be a reflection of the difference between the person being tested and the normative sample (Crawford, 2004; Orsini et al., 1988).

Norms can consequently be seen as the yardstick that is used to measure the person against. If different yardsticks are used, the result for each person being tested will vary with each measurement. This can have significant effects on the diagnoses and inferences made on the basis of a score (Lezak et al., 2004).

When evaluating the adequacy of norms Groth-Marnat (2009) suggest answering the following three questions:

- *Is the normative sample representative of the population to which the examiner wishes to compare the results?*

  "The test manual should include sufficient information to determine the representativeness of the standardization sample. If this information is insufficient or in any way incomplete, it greatly reduces the degree of confidence with which clinicians can use the test" (Groth-Marnat, 2009, p. 10).

- *Is the normative sample large enough?*

  A small sample impairs the generalisations that can be made on basis of the data.

- *Are broad national norms as well as subgroup norms provided?*

  Subgroup norms allow clinicians greater flexibility and confidence in allowing tests to be interpreted according to similar subgroups, as these subgroups may perform differently from the overall norms whilst overall norms are more widely generalisable

A number of potentially useful tests are limited by relatively small normative samples containing less than 20 participants (Lezak et al., 2004). In this case, standard score conversions cannot provide an accurate representation of a person's functioning in comparison to the sample (Lezak et al., 2004).

One challenge currently faced by clinicians, is sourcing appropriate norms for each person being tested (Lezak et al., 2004). This can have a significant effect on the diagnoses and inferences made from the score. The confidence with which neuropsychological tests are interpreted can thus be strengthened by the production of population-specific normative data (Anderson, 2001). Access to demographically sensitive normative data may then serve to reduce interpretative bias (Anderson, 2001). In line with this, Smith-Seemiller et al. (1997) thus recommend the development of adequate norms as an important area for future research. In addition to the normative data, the psychometric properties of a test is a key consideration when evaluating a test.

2.5.4 Psychometric properties. "The utility of neuropsychological assessment for the identification and classification of brain dysfunction is contingent on the psychometric properties of the individual tests included in the battery" (Brickman, Cabo & Manly, 2006, p. 91). According to Grieve (2001), the interpretation of neuropsychological tests, such as the WCST, requires a sound knowledge of psychometric theory and the psychometric properties of the test. Psychometric properties refer to the technical measurement standards with which a psychological assessment measure must comply to provide useful results (Foxcroft & Roodt, 2005). "For any psychological test to be useful, it must be reliable and valid" (Zillmer et al., 2004, p. 67). For this reason, the concepts of reliability (and the resultant standard error of measurement) and validity will now be explored.

2.5.4.1 Reliability. Reliability refers to the degree to which results of a test are replicable (Belli, 2009; Whitley, 2002). The term is often used synonymously with the level of stability, consistency, predictability and/or accuracy of a test (Groth-Marnat, 2009). Reliability provides an indication of whether repeated administration of the test provides more or less similar results (Zillmer et al., 2004). Foxcroft and Roodt (2005) states that, for this reason, reliability is concerned with an instrument's ability to consistently measure the same results; it aims to answer the question: *If a person is tested with the same instrument on two different occasions, will their scores be more or less the same?* (Groth-Marnat, 2009).

The reliability of a test can be expressed by means of a reliability coefficient where a value between 0.80 and 0.90 is deemed appropriate to standardised measures (the closer the value is to 1.00, the higher the reliability of the test) (Foxcroft & Roodt, 2005).

According to Trochim (2006), reliability coefficients may be grouped into four classes:

- *Inter-rater reliability* investigates the degree to which different scorers provide similar scores when scoring a test.
- *Test-retest reliability* assesses the consistency between measurements from one assessment to another.
- *Parallel-forms* reliability determines the consistency of measurement between two similar tests constructed to assess the same construct in a similar manner.
- *Internal consistency* reliability investigates the consistency of results between different items within a test. *Split-half reliability,* which entails splitting the test into two equal halves and calculating the correlation between these two halves, provides one indication of internal consistency (Foxcroft & Roodt, 2005). The Cronbach coefficient en Kuder-Richardson formulas can also be used to calculate the internal consistency of a test (Foxcroft & Roodt, 2005).

Heaton et al. (1993) add the *generalisability coefficient* as a measure of reliability. Generalisability coefficients provide an indication of the fidelity of measurement by calculating the multiple sources of test score variance simultaneously and quantifying the impact of these sources of variance (Heaton et al., 1993). This is done on the basis of a factorial ANOVA and requires normally distributed data (Heaton et al., 1993).

As previously stated, the *Cronbach Alpha Reliability Coefficient* is an indication of the internal consistency of a test (Groth-Marnat, 2009). Cronbach's alpha is a measure of the intercorrelation of items, measuring the extent to which item responses obtained at the same time correlate with each other (Groth-Marnat, 2009). Mathematically, Cronbach's alpha can be seen as equivalent to the average of all possible split-half estimates (Trochim, 2006). This provides an indication of "the internal consistency of the test items rather than the temporal stability of different administrations of the same test" (Groth-Marnat, 2009, p. 14). In effect the reliability of the instrument is estimated by calculating whether items that purport to measure the same construct yield similar results (Trochim, 2006).

The Cronbach Alpha is deemed suitable for calculating the reliability of traits with a high degree of fluctuation where test-retest reliability cannot be used (Groth-Marnat, 2009) and only requires one administration of the test (Brown, 2002; Groth-Marnat, 2009). The Cronbach alpha does not require normally distributed data.

2.5.4.2 Standard error of measurement. The reliability coefficient is usually further used to calculate the *standard error of measurement (SEM).* The SEM provides a measure of reliability that may be used when interpreting individual scores (Foxcroft & Roodt, 2005; Heaton et al., 1993). The SEM provides an estimate of how much variability can be expected around the score a person obtains on the test due to unreliable variance (Brown, 2002). The greater the reliability, the smaller the SEM is expected to be (Tighe, McManus, Dewhurst, Chis & Mucklow, 2010). "In practical terms, there is approximately a 68% chance that an individual's "true" score on a test will fall within plus or minus 1 SEM from this or her obtained score" (Heaton et al., 1993, p. 41). In other words there is a 95% chance that a person's true score is within 1.96 SEM of the obtained score (Brown, 2002).

2.5.4.3 Validity. Validity refers to the degree of accuracy with which a test measures the construct it purports to measure (Belli, 2009; Whitley, 2002). In essence, validity attempts to answer the question of whether the test measures what it says it does (Zillmer et al., 2004). Foxcroft and Roodt (2005) therefore states that validity is thus concerned with what an instrument measures and how well it does at that task. “A test that is valid for clinical assessment should measure what it is intended to measure and should also produce information useful to a clinician” (Groth-Marnat, 2009, p. 16). As such, validity cannot be seen as a generalised concept, but is rather calculated for a specific test’s context, purpose, and population (Groth-Marnat, 2009).

Validity is initially established by theoretically defining the construct to be measured, and then determining whether the items on the test actually measure the theoretically defined construct in practice (Groth-Marnat, 2009). According to Groth-Marnat (2009), a validity correlation is thus concerned with the relationship between the test and the external construct it aims to measure; where a coefficient closest to 1.00 indicates a higher validity.

Groth-Marnat (2009) outlines the following validity categories:

- *Content validity* is concerned with the relevance to and the representation of the test items to the construct being measured.
- *Criterion validity* (also known as empirical or predictive validity) is calculated by comparing the test scores of the test under investigation with an external measure relating to the construct being measured. *Concurrent validity* (as a subset of *criterion validity*), is calculated by correlating the scores of the administered test and the scores on another outside measure administered at more or less the same time. *Physiological validity* refers to calculations of correlation between an external measure, when the external measure is a physiological measurement tool rather than a test.
- *Construct validity* is concerned with the extent to which the test measures the theoretical construct it purports to measure and requires theoretical knowledge of the construct being measured. It is thus concerned with the definition of a construct.

According to Foxcroft and Roodt (2005), the suitability of a test for a person of a particular context, culture, or society cannot be assumed without exploring test bias and re-standardisation the test. "Unless a test was standardized on the population of people from which your patient comes, you are on very shaky neuropsychological ground when you interpret those test results" (Pepping, 2003, p. 93). Therefore, it is important for clinicians to understand the standardisation of a test (the standardised procedures, normative standards and psychometric properties) before using the test and/or making inferences from the test results. For this reason, it is necessary to explore the current standardisation of the WCST.

2.5.5 Current standardisation of the WCST. The following section sets out to discuss the current standardisation of the WCST, as outlined by Heaton et al. (1993) in the WCST testing manual. This section will begin with an analysis of the standardised instructions and scoring of the WCST and will be followed by examining the current normative data available for the WCST.

2.5.5.1 Standardised instructions of the WCST. The WCST manual by Heaton et al. (1993) provides a standardised patter to use when introducing the test, as well as instructions on the standardised manner in which the cards of the test are to be presented. Also provided in the testing manual, is information that dictates the ideal environment in which testing should occur. Then, information is also provided on how certain responses and comments should be answered.

The instruction patter provided in the testing manual has purposefully been formulated in a vague style in order to provide the person being tested with as little information as possible. This is done so that the person is presented with an unstructured testing situation wherein he or she has to experiment with different strategies to deduce the correct problem-solving strategy (Heaton et al., 1993). (cf. par. 2.5.1).

2.5.5.2 Standardised scoring of the WCST. The WCST testing manual provides detailed scoring instructions that stipulate scoring rules and how to calculate each sub-test.

One weakness identified by Ormond-Brown (2010) in Heaton's standardisation, however, is the intricate scoring procedure that might result in the clinician making scoring errors. This is, consequently, confirmed by Heaton et al. (1993) who state that "scoring the WCST has been a source of difficulty for many individuals" (p. 7). Studies cited in Heaton et al. (1993) point out that the incorrect application of the scoring rules may impair the inter-scorer reliability of the WCST.

Ormond-Brown addressed this by constructing a computerised scoring system for the WCST (Ormond-Brown, 2010; Ormond-Brown, n.d.). As previously stated, computerised scoring significantly adds to the reliability of a test by minimising clinician-scoring errors (see par. 2.5.2 and par. 2.5.3.2). The computerised scoring system will be discussed in more detail in Chapter 3 (see par. 3.2.4.3).

2.5.5.3 Normative data for the WCST. Currently, the WCST testing manual provides normative data for individuals between 6½ and 89 years-of-age, and additional correction for the education of persons older than 20 years of age is also provided (Heaton et al., 1993). Where suitable, normalised scores and percentile scores are provided for the major WCST scores to allow for comparison of various persons and/or patient groups (Heaton et al., 1993).

For the US investigation into the WCST, the normative sample consisted of 899 participants who were all screened for a history of neurological difficulties, learning difficulties, emotional difficulties and attention difficulties (Heaton et al., 1993). One group of participants in Heaton's standardisation were tested on a computerised version of the

WCST (Heaton et al., 1993). Normative data is provided for each sub-scores provided in the WCST testing manual (Heaton et al., 1993).

Heaton et al. (1993) elected to use hierarchical polynomial regression analysis to determine the potential effects of age, gender and education on WCST performance. Regression analysis refers to a set of statistical tools aimed at estimating the impact of variables on a data set (Montgomery, Runger & Hubele, 2004). The impact of these demographic variables on WCST performance will now be discussed.

2.5.5.3.1 Age. Age may significantly impact a person's performance on a neuropsychological test (Lezak et al., 2004). According to Heaton et al. (1993), it has been shown that age has the greatest relationship with performance on the WCST. Heaton et al. (1993) report increasing proficiency from 6½ to 19 years of age with performance stabilising through 20-, 30-, 40-, and 50-year decades. After 60 years of age a hastened decrease in performance was noted. This is consistent with reports that the executive function matures in late adolescence and then declines with normal aging (Hughes & Graham, 2008).

2.5.5.4.2 Gender. "An obvious source of individual variation in human behaviour is sex: men and women behave differently" (Kolb & Wishaw, 2008, p. 316). Women usually perform better on tests of verbal skills whereas men perform better on arithmetic or visual spatial tasks. Heaton et al. (1993), however, report that gender does not appear to have an influence on WCST performance.

2.5.5.4.3 Level of education. Level of education tends to influence the performance on tasks of verbal skills, stored knowledge, and scholastic abilities, but may influence performance on any area of neuropsychological assessment (Lezak et al., 2004). "Successive generations tend to be better educated because of better opportunities for formal education"

(Louw et al., 2004, p. 496). When interpreting the psychological test scores of two persons of the same age, the quality and number of years or level of education may thus be responsible for the discrepancy between the two respective scores that are achieved (Louw et al., 2004). Within the South African context, Cave (2008) found that the quality of education had a significant impact on neuropsychological test performance on measures of the executive function, especially with the WCST.

Within the WCST testing manual, the impact of education on WCST performance was only calculated for age groups 20 years and older. This was done in order to prevent the confounding of the age and education variables in younger age groups (Heaton et al., 1993). Level of education was therefore found to be significantly pertinent to performance on the WCST (Heaton et al., 1993).

From the above it is evident that certain demographic variables may influence a person's performance on the WCST. It should be noted, however, that these variables may differ between populations that are to be tested. Therefore, it cannot not be assumed that a test standardised on one population, is suitable for use with another population.

2.5.5.4 Psychometric properties of the WCST. The psychometric properties of the WCST were constructed on populations other than the South African population, but are currently in use in South Africa (cf. par. 2.5.3).

2.5.5.4.1 The reliability of the WCST. As reported in the WCST testing manual, various different studies have proved that the WCST has an excellent reliability rating (Heaton et al., 1993). From reviewing the literature, it became evident that the following measures of reliability are most often used in calculating reliability coefficients for the

WCST: inter-rater reliability; generalisability coefficients; internal consistency; test-retest reliability and standard error of measurement.

According to Anderson, Demasio, Jones and Tranel (1991), the WCST has a high *inter-rater reliability* with a coefficient of 0.94. This is confirmed by Mitrushina, Boone, Razini and D'Elia (2005) with a report of a high inter-rater reliability of between 0.88 and 0.93. This inter-rater reliability is established by calculating the correlation between the scores of test protocols that are independently scored by two examiners (Groth-Marnat, 2009). This means that when two different examiners score the WCST protocols, they are likely to arrive at the same score.

Furthermore, Heaton et al. (1993) report an average *generalisability coefficient* of 0.57. When interpreting generalisability coefficients, a coefficient of 0.60 is considered an indication of very good reliability. The WCST is thus considered to have moderate to good reliability on the basis of the generalisability coefficients (Heaton et al., 1993).

Loeber, Duka, Welzel, Nakovicz, Heinz, Flor and Mann (2009) report a Cronbach Alpha of 0.96 for the WCST. Whilst Heaton et al. (1993) report no Cronbach coefficients for the WCST. On the basis of Loeber et al.'s findings it can thus be concluded that the WCST displays high levels of *internal consistency*.

Bowden et al. (1998), however, report low test-retest reliability for the WCST. Similarly, Heaton et al. (1993) also report a moderate test-retest reliability of between 0.39 and 0.72 in the WCST manual. According to Groth-Marnat (2009), *test-retest reliability* is calculated by administering the test and then repeating the administration at a later testing occasion, after which the correlation between the two scores is then determined. The process of learning, practice and memory, however, significantly impacts on test-retest reliability as these skills usually enhance a person's performance on the second administration (Groth-Marnat, 2009). As the WCST is highly influenced by the processes of learning, practice and memory (Basso, Lowery, Ghormley & Bornstein, 2001), test-retest reliability is not considered a suitable indication of reliability.

In conclusion, although the reliability of the WCST cannot entirely be established by test-retest reliability, it has been sufficiently proven to be a reliable measure in terms of internal consistency, inter-rater reliability and generalisability theory. No measures of *split-half reliability* are available for the WCST.

Currently research into the reliability of the WCST is hampered by the impact of learning on performance as test-retest reliability and alternative form reliability will inevitably provide skewed indications of reliability as persons will do better on second administrations of the WCST due to the learning that occurred during the first administration. At this stage, generalisability coefficients, internal consistency and inter-scorer and inter-rater reliability are most suitable options at the disposal of researchers investigating the reliability of the WCST. Further research is thus required to determine these forms of reliability of the WCST within the South African context.

2.5.5.4.2 Standard error of measurement of the WCST. Heaton et al. (1993) provides SEM's for each sub-score calculated on the basis of the generalisability coefficients presented above. The SEM's range between 7.94 and 11.91. Heaton et al. however note that these will differ between populations and recommends calculations of SEM's specific to various cultures.

2.5.5.4.3 The validity of the WCST. According to Heaton et al. (1993), evidence from various studies on tests that included children, adolescents, and adults suggests that the

WCST is a valid measure of executive function. Various studies have confirmed the *physiological validity* of the WCST, among which are Fallgatter and Strik (1998), who found increased blood flow and activity in the frontal lobes of persons being tested on the WCST. It can therefore be deduced that the WCST provides a valid indication of frontal lobe activity. Heaton et al. (1993), who report studies of a physiological correlation between the frontal lobe, executive function, and WCST performance, confirm the findings of Fallgatter and Strik.

In terms of the WCST, Mitrushina et al. (2005) report adequate *concurrent validity* between the test and other measures of the executive function. Heaton et al. (1993) also confirm this by reporting various studies that support the validity of the WCST as a measure of executive function.

The WCST is thus a well-validated, neuropsychological test that can be widely used to assess executive function within the realm of neuropsychological assessment (Purdon & Waldie, 2001; Smith-Seemiller et al., 1997).

However, despite the fact that the WCST displays adequate psychometric properties – in respect to its reliability and validity – these were calculated for the US population and it can therefore not be assumed that these are valid to the South African context. As previously discussed, the standardisation of a test is relevant to a certain population and context. For this reason, the standardisation of the WCST within the South African context will now be discussed.

2.5.6 Standardisation of the WCST in South Africa.

> Psychological assessment in South Africa faces many challenges at present. Among these challenges, is the fact that assessment practitioners need access to high quality tests so as to ensure that the assessments that they perform yield valid and reliable results. (Foxcroft, Paterson, Le Roux & Herbst, 2004, p. ii).

It is precisely this need for high quality tests that inspired a project by the Human Science Research Council to explore the needs of psychological test users in South Africa. "A major concern noted was that the majority of the tests being used frequently were in need of adapting for our multicultural context or required updating in view of the rapidly changing world of work" (Foxcroft et al., 2004, p. iv). Practitioners indicated a need for updated test norms and for testing instructions to be translated to all eleven official South African languages (Foxcroft et al., 2004). This is confirmed by Grieve (2001) who states that due to the fact that most neuropsychological tests were developed in countries other than South Africa, there is a great need for normative data to encapsulate the extremely heterogeneous South African population.

The cultural bias in verbal-based assessment measures within South Africa has been well documented (Lezak et al., 2004). Previously in the history of cross-cultural neuropsychology within the South African context, it was assumed that non-verbal assessment tasks, such as the WCST, were more suited to cross-cultural assessments (Lezak et al., 2004). A study conducted at the University of the Free State, however, disconfirmed this by documenting a significant discrepancy between the performance of Sesotho-speaking children and the Koppitz norms for the corresponding age groups (Makhele, 2005). Makhele (2005) concludes that the performance of the Sesotho sample appears not to correspond to the pattern of performance of the Koppitz sample. It can thus not be assumed that non-verbal measures can be confidently used cross-culturally without adequate normative data. There thus appears to be a need for normative data for neuropsychological assessment within the South African context.

In 2010, the SAJP also confirmed the need for greater involvement of psychologists in the area of standardisation of existing instruments for use in South Africa as well as in the development of new tests (Maree, 2010). "Local academics and practitioners have repeatedly expressed their concern that most assessment instruments were developed some time ago or were not standardized on a representative sample of the South African population" (Maree, 2010, p. 229).

Even so, very few psychological instruments have been standardised for use in South Africa thus far. Some examples of standardisations that have already been conducted include publications on a national level, for example, Claassen, Krynauw, Paterson & Mathe, 2001, as well as studies conducted at the University of Limpopo (e.g. Burns, 1996; Endres, 1996; Phipps, 1997). At his point in time, very few neuropsychological instruments have been standardised for use in South Africa.

According to Grieve (2001), normative data exists for the Grooved Peg Board, the Trail Making Test as well as the Rey Auditory Verbal List Learning Test for Zulu-speaking factory workers. Grieve also reports norms established for the Bender Gestalt for Zulu-speaking children. Grieve however reports no norms for the WCST or any other measure of executive function in the South African context.

From the literature review it is evident that the WCST has not been standardised for use in South Africa. As a result, within the South African context, confident valid inferences cannot be made on the basis of WCST scores. Accordingly, the suitability of the WCST for the South African population cannot be assumed without investigation of the WCST performance of the South African population.

2.6 Conclusion

This chapter began with a discussion on the fields of psychological assessment and neuropsychological assessment. The discussion focused on executive function as one measurement structure within the realm of neuropsychology. The measurement of executive function was then explored with particular focus on the WCST and led to a discussion on the concept of standardisation with a specific focus on the current standardisation of the WCST as set out by Heaton et al. (1993). This was followed by a discussion of the standardisation of the WCST in the South African context. And from the literature reviewed, the conclusion was made that the WCST has not been standardised for use with the South African population and a need exists for a standardised measure of executive function, such as the WCST, within South Africa.

Since the theoretical context for the research has now been established, Chapter 3 will embark on an exploration of the planned experimental design and methodology employed in conducting the current research.

Chapter 3: The Investigation

This investigation chapter outlines the proposed research design and reports on the actual methodology that has been employed. The chapter will start by stating the research objectives and then outlining the research design aimed at achieving the specified objectives. This includes a description of the variables and formal experimental design. This will then be followed by a discussion on sampling that includes the target population, type of sampling, sample size and how the sampling will be carried out. The data collection process will then be discussed, with a focus on the measurement strategies that are planned and the process of translating the test instructions to Setswana. The section on the research design will conclude with a discussion on the statistical procedures to be employed in analysing the data and to construct the preliminary norm tables.

The final section of this chapter will discuss the actual methodology that has been employed. The section will focus on the documentation of areas where the field research deviated from the proposed research design.

The proposed research design will now be discussed in terms of the research objective of the study.

3.1 Research Objective

The aim of this study is to construct a preliminary standardisation of the WCST for Setswana-speaking university students between the ages of 18 and 29 years-of-age that attend the University of Limpopo (Medunsa Campus). This standardisation will entail preliminary standardising the administration and scoring procedures for the study population as well as constructing preliminary normative data and psychometric properties for the study population.

Now that the research objective has been defined, the proposed research plan to investigate it will be discussed.

3.2 Research Design

"A research design is a strategic framework for action that serves as a bridge between research questions and the execution or implementation of the research" (Durrheim, 1999a, p. 29). The research design therefore refers to the strategic plan that is devised to scientifically investigate a phenomenon as set out in the research question (Babbie, 2005).

The research design of this study is largely based on a previous project run at the University of Limpopo (Medunsa Campus) that aimed to establish a preliminary standardisation of the Wechsler Intelligence Scale for Children – Revised for the Ga-Rankuwa population (Burns, 1996; Endres, 1996; Phipps, 1997) in combination with the current standardisation of the WCST by Heaton, Chelune, Talley, Kay and Curtiss (1993). Throughout this chapter, reference will be made to specific strategies in order to maintain uniformity between this study and the initial parent studies.

This research design is embedded within a theoretical research paradigm, which has certain implications for the research plan (Durrheim, 1999a; Whitley, 2002). For this reason, the discussion of the research design will begin with an outline of the research paradigm before describing the research plan in further detail.

3.2.1 Research paradigm. According to Durheim (1999a) research designs are guided by the research paradigms within which they function, and he describes it as follows:

> Paradigms act as perspectives that provide a rationale for the research and commit the researcher to particular methods of data collection, observation and interpretation. Paradigms are thus central to research design because they impact both on the nature of the research

> question – i.e. what is to be studied – and on the manner in which the question is to be studied (Durrheim, 1999a, p. 36).

This study will function within the quantitative research paradigm. A quantitative research approach makes use of numerical information and statistical analysis (Durrheim, 1999a; Whitley, 2002) and involves the creation of knowledge by counting and assessing numbers (Walliman, 2005; Babbie, Mouton, Payze, Vorster, Boshoff & Prozesky, 2001). It therefore focuses on the production of empirical, numeric and quantifiable data (Belli, 2009). Quantitative research generates data by means of standardised quantitative measures and aims to employ the data in such a way that it is possible to make broad, generalisable comparisons (Durrheim, 1999a).

Maree and Pietersen (2007a) summarise the quantitative paradigm as follows:

> Quantitative research is a process that is systematic and objective in its ways of using numerical data from only a selected subgroup of a universe (or population) to generalize the findings to the universe that is being studied (p. 145).

According to their summary, the core principles within the quantitative paradigm focus on objectivity, the use of numerical data, and generalisability of conclusions (Maree & Pietersen, 2007a).

The quantitative research approach, as appropriate to the use of numerical data, will consequently be used in this study. This allows for the calculation of norms, means, standard error of measurement, and other relevant statistical measures. This approach has been selected not only to reduce possible bias, but also to enhance the generalisability, validity and replicability of the proposed study.

Nonetheless, when investigating a research question by means of a quantitative approach, some careful thought should be put into identifying the variables that are to be explored; this will dictate a suitable research design for the investigation of the specific variables (Belli, 2009; Babbie et al., 2001). The following section will now outline the variables that will be investigated in this study.

3.2.2 Variables. A quantitative investigation of the research question requires the examination of certain variables (Belli, 2009; Maree & Pietersen, 2007a). The term *variable* is used within the study to refer to a characteristic or trait that differs across persons and that can be quantified in terms of differing numerical values (Aron & Aron, 1999; Belli, 2009). These variables can either be measured or categorised in order to assign them numbers and generate numerical information that are then called scores (Aron & Aron, 1999).

The variables that will be studied in this investigation will correlate with those explored by Heaton et al. (1993) in the WCST testing manual. These variables will be discussed on two levels.

On the first level, the study will explore the similarities between the current standardisation by Heaton et al. (1993) conducted in the United States (US) on English-speaking participants in English, in comparison to the current study that aims to study Setswana-speaking students assessed in Setswana. On the second level, the study will explore the impact of certain demographic variables on the WCST score within the Setswana-speaking student population in order to determine whether these variables influence test performance in a similar manner as was found in the current normative information by Heaton et al.

The first level of the study will make use of the variables “Setswana-speaking student assessed with Setswana instructions” as the *independent variable* or *predictor* and “WCST score” as the *dependent variable* or *criterion*. This will allow for a comparison of the study population with the normative data found by Heaton et al. (1993). The WCST score as dependent variable can then be divided further into each of the sub-scores of the WCST as discussed in Chapter 2. These scores are:

- Number of categories completed
- Total number of correct responses
- Total number of errors
- Perseverative Responses
- Perseverative Errors
- Non-perseverative errors
- Trials to complete first category
- Failure to maintain set
- Learning to learn
- Percent conceptual level responses

This allows one to explore the impact of assessing the students with the Setswana instructions on each of the sub-scores. The WCST sub-score variables can be described as quantitative variables because they are measured on a scale with assigned numerical meanings that allow for arithmetic calculations (Belli, 2009).

The second level of investigation will explore the impact of demographic variables on the WCST score. This will involve the use of demographic variables as independent variables and the WCST sub-score as discussed above as dependent variables. The independent variables that will be explored are gender, age, and level of education as employed by Heaton et al (1993). These variables will be quantified as follows:

- Gender will be used to classify participants into *male* or *female*.
- Age will be used to classify the sample into two age groups that corresponds to Heaton et al. (1993), namely *18-to19-year-old* or *20- to 29-years-old*.
- Level of education will be used to group participants according to the number of years of formal tertiary education that they have received. They will thus be grouped into a *first year*, *second year*, *third year*, *fourth year* or *fifth year* group.

Accordingly, gender will be examined as a categorical variable (where numerical assignments of categories are arbitrary) and age and level of education will be examined as quantitative variables (where a higher numerical value indicates a higher age or level of education).

If it then becomes evident that any of these variables have a significant impact on participants' performance, the variables' effects will require to be controlled in the norm tables. Further attention will be given to this matter when the sampling approach is discussed.

The nature of the variables that are investigated influences the research design that is appropriate to the investigation thereof (Belli, 2009). As discussed above, categorical and quantitative variables are to be investigated in this study. A research design that is appropriate to this study is therefore required. The following section will discuss the formal experimental design that will be proposed in order to examine the defined variables.

3.2.3 Formal experimental design. As stated above, this study functions within the quantitative paradigm. Within this paradigm, experimental or non-experimental research can be used (Belli, 2009; Maree & Pietersen, 2007a).

This study will make use of a non-experimental approach that implies a quantitative research study without the manipulation of variables or random assignments (Belli, 2009; Johnson, 2001; Maree & Pietersen, 2007a). This approach is recommended when the investigated variables cannot be manipulated because they are described as naturally occurring attributes (Belli, 2009). Belli (2009) further indicates that a non-experimental approach is useful when a study wishes to examine the relationship between categorical and quantitative variables. This approach can therefore be considered suitable for this research study.

Non-experimental research design can also be classified according to the timeframe of the data collection and the purpose of the study (Belli, 2009; Johnson, 2001). When considering the purpose of the research, the study may be defined as a descriptive non experimental design. This involves the study of variables in such a manner that it documents the variable characteristics, or describes a phenomenon, in order to record the status quo (Belli, 2009; Johnson, 2001).

According to Belli (2009) and Johnson (2001), when considering the timeframe of the data collection the research may be described as a cross-sectional research design where the data is collected at a single point in time – thus focusing on the comparison of different participant groups at a single point in time rather than over a certain time span.

The overall research design can therefore be described as a descriptive, cross-sectional, non-experimental research design that is aimed at documenting data to describe a current state of the variables as they occur in the study population, but without the manipulation of the variables by the researcher (Belli, 2009).

One limitation inherent in the use of non-experimental research, however, is arriving at assumptions regarding the variables when another underlying inherent variable may account for the conclusion (Belli, 2009). Thus, in non-experimental research one can neither assume that the variable investigated is responsible for the outcome difference, nor make inferences about causal relationships. Belli (2009) states that this necessitates the consideration of alternative possibilities for conclusions, and to analyse several variables to enhance the scope of the variables that are addressed. Similarly, conclusions need to be presented without implying any causal statements. Reporting the conclusion properly and investigating a variety of variables rather than a single one, will account for these limitations.

The aim of many non-experimental research studies is to explore potential relationships between variables, and within this context the terms *predictor* and *criterion* variables are used to describe them (Belli, 2009). In this sense, the investigation will centre around the examination of whether the predictor variable has an influence on the criterion variable. The term *predictor variable* can therefore be used interchangeably with *independent variable*, and the term *criterion variable* with *dependent variable*, despite the absence of an experimental research design when investigating the possibility of a relationship by means of a non-experimental research design (Belli, 2009).

In order to investigate a relationship by means of a non-experimental design, a study population and sample is required. The sampling approach used to select the participants for this research design will now be discussed.

3.2.4 Sampling. "Sampling involves decisions about which people, settings, events, behaviours and/or social processes to observe" (Durrheim, 1999a, p. 44). Sampling therefore refers to the process by which participants are selected for the study (Van Vuuren & Maree, 1999). As a result, it is concerned with how the target population selected for the study, is reduced to a representative sample (Whitley, 2002).

The target population of a study refers to the entire group of people to which the results of the study may apply (Maree & Pietersen, 2007a; Whitley, 2002). Within behavioural research, target populations are usually defined in terms of hypothetical constructs, where these hypothetical constructs are then operationally defined (Whitley, 2002). People who then match the operational definition of the target population, comprise the study population (Whitley, 2002). The research sample is then selected from the study population – thus subset of the selected population – to participate in the study (Whitley, 2002; Maree & Pietersen, 2007a).

Considering the definitions of the target population, the study population and research sample is of great importance as it may influence the validity of the study. Whitley (2002) maintains that errors in the definition of the study population may result in it being an incorrect representation of the target population. Similarly, Whitley states that errors in the selection of the sample may result in an inaccurate representation of the sample population – reducing the generalisability of the results to the target population.

The main concern in sampling is thus the *representativeness* of the sample to the population being studied. This is especially important in studies, such as this study, aiming to describe a set of behaviour in a certain population (Durrheim, 1999a).

Sampling subsequently plays an important role, because the conclusions drawn from the research are only generalisable to the persons on whom the research was conducted (Van Vuuren & Maree, 1999). The inferences made on the basis of the research are thus dependent on the accuracy of the sampling process (Onwuegbuzie & Collins, 2007). The following section will outline the target population, the study population, and the study sample and how these are representative of the population under investigation. The section will also identify the population to which the conclusions of this study can be generalised.

3.2.4.1 Target population. The target population of this preliminary standardisation will be Setswana-speaking university students of South Africa.

3.2.4.2 Study population. The target population will be operationally defined as Setswana-speaking students between the ages of 18 and 29 years-of-age that attend the University of Limpopo (Medunsa Campus). This corresponds with two of the age groups identified by Heaton et al. (1993), namely the 18- to 19-year-old and 20- to 29-year-old age groups.

As stated in the literature review, executive functions mature in late adolescence and declines again with normal aging (Hughes & Graham, 2008). Heaton et al. (1993) report that performance on the WCST stabilises by early adulthood. The 18- to 29-year-old age group was thus selected, seeing that executive functions and WCST performance are said to stabilise during this life phase. This eliminates the impact of developmental changes in executive function as a nuisance variable.

As the level of education has also been shown to impact on WCST performance (Heaton et al., 1993), the selection of university students as study population allows for the investigation into the impact of education on participants. As the research is conducted under supervision of the University of Limpopo (Medunsa Campus), the same university will be selected as the study population.

The University of Limpopo was founded in January 2005 as a result of a merger between the former Medical University of Southern Africa (Medunsa) and the University of the North (University of Limpopo, n.d.). The former Medical University of South Africa is now known as the Medunsa Campus, and is situated north-west of Pretoria adjoining the townships of Ga-Rankuwa and Mabopane (University of Limpopo, n.d.).

Medunsa was founded in 1976 in a move to provide tertiary education to the educationally disadvantaged within the fields of health care sciences (University of Limpopo, n.d.) and to address the underrepresentation of black professionals in South Africa's health professions as well as the lack of adequate healthcare in disadvantaged areas (Haynes, Lee & Drew, 1995; Tshabalala-Msimang, 2004; Medunsa, n.d.). The majority of students at Medunsa are trained as black physicians, dentists, veterinarians or allied health care professionals (Haynes et al., 1995). Many students studying at Medunsa, moreover, are from disadvantaged backgrounds (Tshabalala-Msimang, 2004).

The University of Limpopo (Medunsa Campus) is situated in Ga-Rankuwa. Currently, Ga-Rankuwa is regarded as a "largely uniform urban residence" (Phipps, 1997, p. 57). According to Burns (1996), the predominant language of discourse in the Ga-Rankuwa area is Setswana. For this reason the study will thus focus on the Setswana-speaking university population.

3.2.4.3 Research sample. The research sample will be constructed by selecting Setswana-speaking students who attend the University of Limpopo (Medunsa Campus). The type of sampling used to construct the sample, the sample size, and the manner in which the sample will be obtained will now be discussed.

3.2.4.3.1 Type of sampling. When the aim of a study is to generalise the findings to the population being studied, a random sampling method is required (Onwuegbuzie & Collins, 2007). Random sampling implies that each member of the study population has an equal chance of being selected for the sample seeing that participants are selected by chance (Babbie, 2005; Whitley, 2002). Random sampling is one technique that aims to increase the representativeness of samples, and therefore tries to ensure that all relevant subgroups are represented in the sample (Durrheim, 1999a; Salkind, Dougherty & Frey, 2010). According to Onwuegbuzie and Collins (2007), random sampling is usually associated with the quantitative paradigm, and thus fits to the paradigm within which this research will operate.

The random sampling method most often used to select standardisation samples is *stratified random sampling* (Groth-Marnat, 2009; Phipps, 1997). Stratified sampling is employed in situations where the population that is under investigation consists of different subgroups that may be of interest when interpreting the research conclusions (Van Vuuren & Maree, 1999). Stratified random sampling will thus also be used to select the sample in this study.

Stratified random sampling entails structuring the sample according to selected variables (Maree & Pietersen, 2007b; Onwuegbuzie & Collins, 2007; Whitley, 2002). "The ultimate function of stratification is to organize the population into homogenous subsets (with heterogeneity between subsets) and to select the appropriate number of elements from each" (Babbie et al., 2001, p. 191). The population is thus divided into subgroups called *strata* and random samples are drawn from each of the strata (Onwuegbuzie & Collins, 2007; Van Vuuren & Maree, 1999).

When making use of stratified random sampling, a researcher may either make use of *proportional* or *constant (equal)* random sampling (Maree & Pietersen, 2007b; Salkind et al., 2010). When using proportional stratified random sampling, the strata are selected in a manner that is representative of these strata as they occur in the population (Maree & Pietersen, 2007b; Salkind et al., 2010). This strategy is employed when the researcher aims to estimate values in a population where different subgroups within the population may differ on the variable of interest (Salkind et al., 2010).

Constant stratified random sampling selects a constant number of participants from each stratum despite the distribution of thereof in the population (Maree & Pietersen, 2007b; Salkind et al., 2010). This technique is employed when the research aims to make comparisons between the strata within the population when these are not equally represented within the population (Salkind et al., 2010). As such, this sampling technique is optimal for making comparisons between strata, but not for making general estimates regarding the overall population (Salkind et al., 2010). Constant stratified random sampling may thus be considered to allow for strata with sufficient sizes to calculate of the impact of demographic variables.

The strategy most often used in standardisation studies, however, is *proportionate stratified random sampling* (Phipps, 1997). Heaton et al. (1993) constructed a proportional stratified random sample on the basis of the US census conducted in 1984. In order to maintain consistency with this strategy, proportionate stratified random sampling may be considered on the basis of census information obtained from the University of Limpopo (Medunsa Campus).

This study will however rather employ constant stratified random sampling in order to ensure adequate cell sizes for statistical analysis. Samples will be stratified according to age, gender and level of education to emulate the normative data as reported in Heaton et al. (1993). This will be elaborated upon in the section on statistical analysis (cf. 3.2.6.3). Other variables, such as ethnicity and rural/urban residence, will be eliminated by the sampling procedure to enhance the specificity of the preliminary study.

3.2.4.3.2 Sample size. An additional consideration relevant to the representativeness of the sample is the sample size (Durrheim, 1999a). “The choice of sample size is as important as is the choice of sampling scheme because it also determines the extent to which the researcher can make statistical and/or analytic generalizations” (Onwuegbuzie & Collins, 2007, p. 287). Seeing that a small random sample may be just as unrepresentative as a larger non-random sample, an adequate sample size is required in order to make inferences about a target population (Durrheim, 1999a; Van Vuuren & Maree, 1999).

Maree and Pietersen (2007b) recommend at least 15 participants per group when aiming to compare subgroups with each other, 30 participants when making use of correlational research, and 100 participants with 20 to 50 participants per subgroup when conducting other major research.

Heaton et al. (1993) included 123 participants in the two age groups that were under investigation. Heaton’s sample consisted of 56 participants in the 18-to 19-year-old age group and 67 participants in the 20- to 29-year-old age group.

The sample for this study will consist of approximately 100 Setswana-speaking students. The sample will be equally divided between the two age groups studied. 50 participants will fall in the age group 18- to 19-years-old and 50 participants in the age group 20- to 29-years-old. The strata sizes for the level of education will be calculated on the basis of census information obtained from the University of Limpopo (Medunsa Campus).

3.2.4.3.3 Conducting the sampling. The sample will be constructed by selecting and visiting different year groups at the University and administering a screening questionnaire. Students who are willing to participate and match the selection criteria will be divided into a sampling frame, taking into consideration the stratification variables including age, gender, and level of education. Participants in the strata will be assigned chronological numbers. The appropriate number of participants will then be selected accordingly from each age group to ensure sufficient cell sizes for statistical analysis.

However, certain exclusion criteria will be considered when compiling the sample. The screening questionnaire will focus on eliminating confounding variables. This is done to keep the sample as similar as possible to the original normative population. According to Heaton et al. (1993), "clients should have normal or corrected vision and hearing sufficient to adequately comprehend the test instructions and to visually discriminate the stimulus parameters of colour, form and number" (p.3).

Adhering to these exclusion criteria will also ensure the specificity of the sample. Persons with hearing or visual impairments will therefore be excluded. Persons with a history of psychiatric or neurological difficulty will be excluded. Persons with previous exposure to the WCST will also be excluded. These exclusions will ensure that the participant's scores on the WCST is an accurate reflection of their executive abilities and not one of the variables mentioned above.

Bias in the sample will be reduced by screening all willing participants and then randomly selecting 100 participants. Setting selection criteria for participation in the study will further reduce sampling bias. Participants with a history of brain damage or psychiatric problems will be excluded to enhance the specificity of the sample. Participants with visual or auditory difficulties will also be excluded. Only selecting Setswana-speaking participants will again ensure sample specificity.

Another factor that might influence bias is the participants' test-wiseness. To eliminate this factor, university students are specifically selected since they are often exposed to testing situations which allows for the assumption that participants are test-wise.

Some bias can, however, not be eliminated. Students are asked to volunteer for the study and this may result in selection bias as the students who volunteer for the study may share some unknown characteristics that can influence the data (Whitley, 2002). More importantly, this limitation cannot be overcome in this study since the Health Professions Council of South Africa (HPCSA) dictates that data for such a study may not be obtained from a research participant without their written and informed consent (Professional Board for Psychology, 2004). Data can therefore not be collected from participants who do not volunteer for the research.

Now that the sampling procedures have been defined, it becomes necessary to discuss the measurement strategies that are employed in quantifying the constructs that are under revision in this sample.

3.2.5 Data collection. As previously discussed, many variables are involved in this study. This section aims to discuss the measurement and operational definition of these variables that are used to generate data.

Data refers to the core material produced and analysed by the research, and therefore the quality of the research is directly related to the quality of the data that is generated (Durrheim, 1999a).

Within the quantitative paradigm, the quality of the data obtained depends on the stringency of the two-fold deductive process by which constructs are conceptualised and operationalised (Durrheim, 1999a; Durrheim, 1999b). This is achieved by the process of measurement.

Measurement refers to limiting the data of a phenomenon in order to interpret and compare the data (Leedy & Ormrod, 2005). Measurement can thus be seen as a set of rules that guide the assignments of numbers to concepts in order to represent differences in the quantities of these concepts (Durrheim, 1999b). It thus aims to allow differentiation between concepts in terms of their relative standing towards each other (Durrheim, 1999b). Measurement consequently allows the quantification of operationally defined concepts (Walliman, 2005), turning abstract concepts into quantitative variables (Durrheim, 1999b).

The process of measurement begins with conceptualisation (Durrheim, 1999b). Conceptualisation refers to the abstract definition of the construct to be observed (Durrheim, 1999a). This requires the linguistic clarification of the concept that is under investigation and the construct of a theoretical definition (Durrheim, 1999b).

Operationalisation refers to the translation of the theoretical construct to observable measures (Durrheim, 1999a). There are usually many ways in which a construct can be measured – the most accurate, however, will be the measurement strategy that most closely fits the operational definition (Durrheim, 1999b). Operationalisation is thus concerned with how to measure the construct empirically, ensuring agreement between the conceptualised and operationalised definitions (Durrheim, 1999b).

Within this research study, the concept investigated is that of executive functions as described in Chapter 2. The construct of executive functions is thus conceptualised as the ability to develop and maintain an appropriate problem-solving strategy across changing stimulus conditions in order to achieve a future goal (Heaton et al., 1993) (c. 2.3.1). Accordingly, the operationalisation of this conceptualisation is thus the person's performance on the WCST.

Now that the constructs under investigation have been operationalised, these need to be quantified. Different types of quantifications are referred to as *levels of measurement* (Aron & Aron, 1999; Walliman, 2005). The following section will discuss the various levels of measurement used to quantify the variables relevant to this study.

3.2.5.1 Levels of measurement. Levels of measurement should be closely considered as they dictate the statistical and mathematical measures that may be employed to analyse the data that is obtained through the measurement strategy (Durrheim, 1999b).

The *nominal level* of measurement allows for the naming or labelling of characteristics that enable the categorisation of data as well as the comparison of these categories (Aron & Aron, 1999; Delport, 2005; Durrheim, 1999b; Maree & Pietersen, 2007a; Walliman, 2005). Nominal measurement consequently refers to the assignment of numbers to different items to indicate their placement in distinct mutually-exclusive and exhaustive categories that can be compared with each other (Delport, 2005; Walliman, 2005). Numerical assignment on this level thus indicates a mere difference in category between constructs which allows the numbers to be used as labels to assign constructs to different categories (Durrheim, 1999b).

This measurement strategy will be used to measure *gender of participant*, *level of education*, and *age of participant*. Gender of participant can thus be classified as "male" or "female". And with regards to age of participant and level of education, subjects will be assigned to categories in terms of age groups and number of years of education.

Interval levels of measurement involves the assignment of items to numbered categories where is it assumed that there are equal distances between scale values and an arbitrarily assigned zero point (Delport, 2005; Leedy & Ormrod, 2005; Walliman, 2005). Categories can thus be ordered in numerical order where a higher number would be equivalent to a higher degree of the item measured (Leedy & Ormrod, 2005). Interval levels demarcate differences in rank between constructs but also allow meaningful interpretation of the differences between two numbers (Durrheim, 1999b; Maree & Pietersen, 2007a).

Interval levels of measurement will be used to measure executive functions as indicated by sub-test scores on the WCST. A higher number will thus indicate a higher degree of exhibited executive functions.

It may, however, be argued that in measuring executive functions, interval levels of measurement might not be appropriate, since it cannot with certainty be said that a score difference between, for example, 3 and 5 relates to the same difference in executive function as indicated by a score difference between 9 and 11.

In this study, interval levels of measurement will however be assumed as it allows for the most useful statistical summary measures and tests, such as the calculation of the mean and standard deviation (Leedy & Ormrod, 2005; Walliman, 2005).

Now that the levels of measurement involved in this study have been described, the instruments used to measure will be discussed. Measurement requires a standardised measuring instrument to ensure accuracy. As described in Chapter 2, the WCST, which is to be used as measuring instrument, has been adequately standardised in English for a US population (cf. 2.5.5). There is, however, a lack of a measuring instrument to collect data in Setswana. For this reason, the current WCST testing instructions need to be translated to Setswana to provide an appropriate measuring instrument. This will be the focus of the following section.

3.2.5.2 Translating the WCST. As described earlier in this chapter and in Chapter 2, executive functions will be measured by making use of the WCST. Chapter 2 reports that the WCST shows sufficient reliability as a measurement of executive functions when administered in English to English-speaking participants.

However, the WCST has not yet been translated to Setswana. For the purpose of this study, it is therefore necessary to translate the test instructions. Test items, on the other hand, are non-verbal in nature and do not require translation.

This will involve the translation of the English instructions to Setswana by an independent professional translator. The Setswana translation will then be translated back to English by another independent translator. This approach in translation aims to eliminate bias in the instructions.

As described in Chapter 2, the instruction patter provided by Heaton et al. (1993) is necessarily vague in order to provide the person being tested with as little information as possible, as they are required to create their own structure within the unstructured environment. Therefore, care will be taken to maintain the vagueness of the test instructions during translation, as this is a necessary characteristic of the WCST.

The translations will be printed and handed to participants to read through before the test starts in a similar manner to the computerised instructions presented by Ormond-Brown (2010) and Ormond-Brown (n.d). This will serve the function of standardising the test instructions in Setswana. It will also enable clinicians who are not proficient in Setswana to administer the WCST to Setswana-speaking persons.

The following section will discuss the manner in which the translated test instructions will be used to collect data.

3.2.5.3 Administration and scoring procedures. Data will be collected by administering a computerised version of the WCST to the sample participants. The program was developed by Ormond Software Enterprises by a qualified clinical psychologist with extensive experience in neuropsychology (Ormond-Brown, 2010; Ormond-Brown, n.d.). It provides computerised scoring of performance according to the scoring rules as set out by Heaton (Ormond-Brown, 2010; Ormond-Brown, n.d.). Administrations are automatically scored and visually recorded allowing them to be replayed at a later stage (Ormond-Brown, 2010; Ormond-Brown, n.d.). This, in turn, allows for scoring to be checked and replayed at a later stage to more fully assess a person's performance. According to the test developer, the program is suitable for clinical use and the computerised scoring eliminates examiner scoring errors (Ormond-Brown, 2010). This is consistent with reports of the minimisation of scoring errors in computerised scoring as discussed in Chapter 2 (see par. 2.5.2 and par. 2.5.3.2).

The standardised instructions as set out by Ormond-Brown (2010) and Ormond-Brown (n.d.) will be followed seeing that they maintain consistency with the instructions by Heaton et al. (1993) with the addition of instructions clarifying the use of the computer program. Administration will make use of the translated test instructions and the computerised administration and scoring program for the WCST.

Data will be collected from the participants on several testing days by the researcher who is a trained clinical psychologist with experience in administering the WCST and establishing rapport with participants. The computerised test program will eliminate the possibility of bias and tampering with the data. As mentioned previously, informed consent will be obtained from the participants before any data is collected.

3.2.5.4 Testing. Participants will be tested on consecutive days over a time span of approximately two weeks. Testing will be done at the Department of Clinical Psychology at the University of Limpopo (Medunsa Campus). In this manner data will be constructed for statistical analysis.

3.2.6 Statistical analysis. Data analysis aims to convert the data that is generated through the research, to answer the research question (Durrheim, 1999a). Within the quantitative paradigm, various statistical techniques are at the researcher's disposal to achieve just this (Durrheim, 1999a; Durrheim, 1999c).

Within this research, descriptive statistics will be used in order to describe the data and summarise its distribution in an attempt to make the collected data understandable (Aron & Aron, 1999; Belli, 2009; Durrheim, 1999c; Pietersen & Maree, 2007a). Inferential statistics will further be employed to draw conclusions from the population on the basis of the sample data, by allowing conclusions that are beyond the collected data and the analysis of hypotheses (Aron & Aron, 1999; Belli, 2009; Durrheim, 1999c; Pietersen & Maree, 2007a).

Data will be analysed with the assistance of a professional statistician. This will aid in reduced researcher bias. The statistician will also have no knowledge of the aims of the study in order to eliminate the possibility of further bias.

The statistical analysis will involve five aspects: (a) comparing the data from this study with the data provided by Heaton et al. (1993) to determine the need for a re-standardisation; (b) exploring the distribution of the data to determine which statistical measures may be employed; (c) categorising significant variables to determine which of the variables require consideration in constructing normative data; (d) constructing preliminary normative data; and (e) constructing preliminary psychometric properties for use of the WCST in the Setswana-speaking university population.

3.2.6.1 Hypothesis testing. Hypothesis testing refers to an inferential statistical procedure that is used to determine whether two hypotheses are significantly different from each other (Tredoux & Smith, 1999; Montgomery, Runger & Hubele, 2004; Pietersen & Maree, 2007b). Hypothesis testing will be used to determine if there is a significant discrepancy between the means obtained in this study and the means provided by Heaton et al. (1993).

If the hypothesis test shows a significant difference between the means, it can be said that there is a significant difference between the South African data and the current norms, and that the test may need to be re-standardised for the South African context.

If the hypothesis test shows no significant difference, however, there will not be a difference between the data sets, and the test may not need to be re-standardised at all.

3.2.6.2 Examining the distribution of the data. Frequency distribution will be used as a graphical summarisation of the frequency with which participants obtained a specific score on a variable, allowing for the distribution of the scores to become visible (Aron & Aron, 1999; Durrheim, 1999c). This will be displayed by means of histograms, where the height of each bar corresponds to the frequency with which each score is obtained (Aron &

Aron, 1999). This will allow for a visual representation to determine whether the data is distributed normally.

Further tests to examine the normal distribution of the data, such as the Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von Mises, and Anderson-Darling tests will determine whether the data is normally distributed. Further analysis to determine a person's relative standing in relation to the norms will depend on the distribution of the data.

3.2.6.3 Identification of significant variables. The following level of statistical analysis will investigate the co-variation between variables, examining whether the variables vary in relation to each other or whether they function independently (Durrheim, 1999c).
When one wishes to compare the means of different sets of the same variable, ANOVA techniques are commonly used (Tredoux & Smith, 1999). These techniques investigate whether the variation between the different means is greater than what would be expected from random variation. As such, the factorial ANOVA is the technique used in studies that contain more than one independent variable (Tredoux & Smith, 1999). ANOVA techniques can be employed when the data is normally distributed and the variance of the variables that are compared is the same in both data sets (Pietersen & Maree, 2007c).

Factorial analysis of variance (or factorial ANOVA) for a balanced design will be done to determine the significance of the previously identified stratification variables. This technique emulates the methodology employed by Phipps (1997). This technique is also chosen for its robust nature (Williams, Sweeney & Anderson, 2009).

The robustness of the technique is, however, improved by a balanced sampling design. A balanced design is characterised by equal numbers of participants in each cell or variable group where each cell should also at least contain 10 to 15 subjects (Williams et al., 2009). As previously discussed, this will result in the use of constant stratified random sampling (CSRS) rather than proportional stratified random sampling (PSRS) in an attempt to ensure an equal number of participants per cell.

The factorial ANOVA will further be limited to three-way interactions to reduce the complexity that is presented by four-way interactions. This will provide an indication of which stratification variables influence performance on the WCST more significantly.

3.2.6.4 Formulation of norms. Norm tables will be constructed on the basis of the factorial ANOVA. If a certain demographic variable influences WCST performance separate norm tables will be required considering these variables. If they do not influence WCST performance, one norm table will be constructed.

The norms tables will be constructed by considering the following statistical techniques:

- **Measures of central tendency** will be calculated to provide a finer description of the distribution of scores by making the centremost score of the distribution visible (Durrheim, 1999c). In this respect, the mean, mode, and median will be calculated. The mode indicates the most commonly occurring score in the distribution, the median the middlemost score in the distribution and the mean the arithmetic average of the set of scores (Aron & Aron, 1999; Durrheim, 1999c; Pietersen & Maree, 2007a).
- **Measures of variability** and **spread of the data** will provide an indication of the degree to which scores vary around the measures of central tendency (Aron & Aron, 1999; Durrheim, 1999c). The range will provide the difference between the highest and lowest score; the variance will provide the average distance between a score and the mean (Aron & Aron, 1999; Durrheim, 1999c; Pietersen & Maree, 2007a). The variance will then be converted to a standard deviation to provide a more user-friendly indication of the

average difference between the mean and an obtained score (Aron & Aron, 1999; Durrheim, 1999c; Pietersen & Maree, 2007a).

3.2.6.5 Psychometric properties. Preliminary psychometric properties, including reliability coefficients and standard error of measurements (SEM) will be calculated to determine the properties of the WCST in the Setswana-speaking university population.

Test-retest reliability and alternate form reliability are not deemed appropriate to use with the WCST as these are highly influenced by the impact of learning on WCST performance. Inter-rater and inter-scorer reliability are deemed redundant, as the computerised scoring system standardises the scoring to the extent that, in principle, a reliability coefficient of one can be expected as all scoring is done in the exact same manner and the possibility of examiner-scoring error has been eliminated. Split-half reliability can also be excluded from this study, as the WCST cannot be split into equal halves in a meaningful manner. Test-retest, alternate form, split-half, inter-rater and inter-scorer reliability are thus beyond the scope of the research.

The measures of reliability available to this research are thus the generalisability coefficient and internal consistency.

The generalisability coefficient may be calculated to provide an indication of the amount of error in the scores due to external sources of measurement error (Heaton et al., 1993). This requires only one test administration (Heaton et al., 1993). It is, however, calculated on the basis of a factorial ANOVA and thus requires normally distributed data (Heaton et al., 1993).

A Cronbach Alpha Coefficient may be calculated as an indication of the internal consistency of the WCST. As discussed in the literature review, The Cronbach Alpha is deemed suitable for calculating the reliability of traits with a high degree of fluctuation where test-retest reliability cannot be used (Groth-Marnat, 2009) and only requires one administration of the test (Brown, 2002; Groth-Marnat, 2009). It also does not require normally distributed data.

Furthermore, the reliability coefficient will be used to calculate the standard error of measurement (SEM). The SEM provides a measure of reliability that may be used when interpreting individual scores (Foxcroft & Roodt, 2005; Heaton et al., 1993). The greater the reliability, the smaller the SEM is expected to be (Tighe, McManus, Dewhurst, Chis & Mucklow, 2010). The SEM thus provides an estimate of how much variability can be expected around the score a person obtains on the test due to unreliable variance (Brown, 2002). As discussed in the literature review, calculating the validity of the WCST is however beyond the scope of this research (cf. par. 2.5.3).

Now that the proposed research design has been discussed, it becomes necessary to reflect on the ethical principles that will be considered in order to protect the participants and ensure the research is conducted in an ethical manner.

2.3.7 Ethical considerations. In order to protect participants and the quality of the research, the following principles will be implemented:

- Ethical clearance will be sought from the Medunsa Campus Research Ethics Committee (MCREC) in order to ensure the ethical compliance of the proposed study.
- The privacy of each participant will be respected and all data will be kept confidential.
- The anonymity of participants will be protected as far as possible and no identifying information of participants will be published.

- Identifying information will only be kept on record if participants need to be contacted and referred to for further neurological screening.
- All data will be kept in locked cupboards and encoded files to prevent the participants' anonymity being impaired by negligence.
- Participants will be informed about the purpose of the study and a letter of informed consent will be obtained from participants before any data is collected from them.
- Special care will be taken to protect participants from harm – if any difficulties are identified by the research, participants will be referred to the relevant medical or psychological professionals at the Dr George Mukhari Hospital.
- Participants will not be treated by the researcher in order to prevent dual roles that can result in ethical dilemmas.
- Effort will be taken to ensure honesty with the participants in explaining the purpose, benefits, and risks of the study.
- Care will also be taken to ensure honesty with professional colleagues – the research will be conducted in an open and transparent manner as to allow input from supervisors and colleagues – whilst keeping the anonymity of participants in mind.
- Findings will be released and published in such a manner as to protect the anonymity of participants.
- Findings will be released in such a manner as to protect the study population from being stigmatised.

As such, the proposed study aims to be the researcher's own and original work. Care will be taken to acknowledge and quote all sources that are not the researcher's own words.

The proposed research design to be conducted under optimal conditions has now been discussed. The final section will discuss the actual research methodology that is to be employed, focusing on areas where the methodology has to deviate from the proposed research design.

3.3 Research Methodology

The following section on research methodology aims to record the instances where deviations occurred from the research design.

3.3.1 Translating the WCST instructions. Financial limitations required a change in the translation strategy. Although the first translation – from English to Setswana – was performed by an independent professional translator, the second translation or *back-translation* – from Setswana to English – could not be done due to financial constraints. As an alternative, the back-translation was voluntarily undertaken by an independent clinical psychologist who is fluent in both English and Setswana, and who is trained in the nuances of verbal and non-verbal communication. This adaption, it may be argued, actually served to strengthen the translation procedure, because the language competency of the clinical psychologist concerned, more closely corresponds with that of the target population, thereby ensuring the overall validity of the translation.

The translation was deemed suitable as the back-translation maintained the vagueness of the original testing instructions whilst providing sufficient structure to successfully complete the task. The core principles present in the original instructions were also present in the back-translated version. The instructions are thus deemed appropriate to use in introducing the WCST.

3.3.2 Sampling. Attempts to obtain a random stratified sample were hampered by difficulties in gaining access to the students. Classes were visited and as many volunteers as possible were found. The numbers were, however, insufficient to make up the sample size of

100. These participants were then asked to provide contact details for Setswana-speaking friends and colleagues that may be willing to participate. These participants were then phoned, screened telephonically, and appointments were booked if they matched the participation criteria. If, during testing, it was found that some of the criteria were not met, the data was excluded from the analysis.

Whitley (2002) refers to this manner of sampling as *snowball sampling*. "Snowball sampling – also known as chain referral sampling – is a method whereby participants with whom contact has already been made are used to penetrate their social networks to refer the researcher to other participants who could potentially take part or contribute to the study" (Nieuwenhuis, 2007, p. 80). This strategy aids in the facilitation of a greater sample size (Whitley, 2002), and allows access to what seems like inaccessible populations (Maree & Pietersen, 2007b; Nieuwenhuis, 2007). This method is primarily used in explorative research as it impairs the representativeness of the sample (Babbie, Mouton, Payze et al., 2001). This method has thus impaired the representativeness of the sample, but ensured adequate sample sizes for analysis. Even so, the sample sizes are not equal to each other as intended by the original constant stratified random sample.

In total, 93 participants were included in the sample and were tested. Of this number of participants, 43 were male and 50 were female. Thirty-one participants were in the 18- to 19- year-old age group and 62 participants in the 20- to 29-year-old age group. Thirty-two participants were in their first year of tertiary study, 27 participants were completing their second year of tertiary study. Twenty-eight participants were third year tertiary students; one participant was in his/her fourth year; one in his/her fifth year, and four in their sixth year of study. The different gender, age, and level of education groups were therefore not balanced.

Inevitably, this influenced the statistical procedures that were utilised. The changes to the statistical procedures will now be discussed.

3.3.3 Statistical analysis. Due to the changes in the sampling and on advice of the statistician on the suitably of the data obtained for analysis, the statistical strategies employed were adapted as follows.

An analysis of the WCST scores was done to show whether or not the data is normally distributed. As a result, it was found that the data was not normally distributed, and therefore the factorial ANOVA could not be completed as a prerequisite because the factorial ANOVA consists of normally distributed data. Furthermore, balanced cell sizes were not obtained and the factorial ANOVA was thus substituted by regression analysis to determine the impact of the demographic variables.

Regression analysis refers to the process by which the data is plotted on a graph and attempts to draw a straight line through the data – this line is called a regression line (Durrheim, 1999c). The regression line summarises the distribution of the two data sets in relation to each other (Aron & Aron, 1999; Durrheim, 1999c) and provides an indication of relationships between two variables (Belli, 2009; Pietersen & Maree, 2007c).

A high correlation coefficient within the regression analysis thus indicates a strong relationship between the two variables (Durrheim, 1999c). A correlation coefficient is the calculation of the degree to which the data is scattered around the regression line (Aron & Aron, 1999; Durrheim, 1999c).

Regression analysis was thus conducted to determine whether there was a linear, quadratic, or polynomial relationship between the WCST scores and the demographic variables. This replaced the planned factorial ANOVA which required normally distributed data and equal cell sizes, which could not be provided by this study.

Regarding the construction of psychometric properties, a generalisability coefficient could not be calculated as this calculation is dependent on the use of ANOVA and as has previously been discussed, the ANOVA requires normally distributed data which was not obtained in this study. The Cronbach Alpha was thus calculated as a measure of internal consistency and can be used a measure of the reliability of the WCST as it does not require normally distributed data.

Furthermore, considering the construction of psychometric properties, the SEM could not be calculated. This is due to the fact that the Cronbach coefficient provides an overall indication of the internal consistency, but the WCST does not provide an overall score, but rather sub-scores. Calculating the SEM on the basis of an overall score provides information on an overall score level and not on a sub-score level. To calculate an SEM for the WCST on the basis of a Cronbach coefficient thus results in meaningless data that cannot be used to provide an indication of the expected variability around a sub-score.

Chapter 4: Results and Discussion

This chapter will document the findings of the investigation as discussed in Chapter 3: The Investigation. As such, each statistical measure that has been employed will be presented and then interpreted. The discussion will start with an outline of the characteristics of the specific sample that has been investigated. This will be followed by a comparison of the investigation into the necessity of normative data for the Setswana population (which focused on comparing the current normative data presented in the testing manual by Heaton, Chelune, Talley, Kay & Curtiss (1993)) with the corresponding data that has been obtained. Thereafter, the preliminary normative data that has been obtained from this study will be discussed by exploring the distribution of the data, investigating the impact of the demographic variables on the data and constructing a normative table based on the findings. Preliminary psychometric properties will also be presented for the Setswana standardisation. The findings of the investigation will then be discussed and followed by a presentation of issues for consideration when interpreting the findings together with recommendations for further research.

Each statistical test was conducted using $\alpha = 0.05$ and 1 degree of freedom, ensuring a 95% confidence interval. Inferences can thus be made with a certainty of 95%.

4.1 Describing the Sample

This section will lay the foundation for statistical analysis by documenting the characteristics of the sample. Tables 1(a), 1(b) and 1(c) will present the cell sizes for the age groups, gender and levels of education of the persons tested. The mean age and levels of education will also be provided in Tables 1(a) and 1(b).

Table 1(a) Sample distribution – Age

	Cell size (n)	Mean age	Mean years of study
Overall sample size (N)	93	21.36	2.18
18 to 19 years old	31	19.13	1.29
20 to 29 years old	62	22.48	2.63

From Table 1(a) it can be seen that 93 persons tested were included in the sample. The entire sample shows an average age of 21.36 years and an average level of education of 2.18 years of study.

31 persons tested are in the 18- to 19-year-old age group and 62 in the 20- to 29-year-old age group. The 18- to 19-year-old group shows an average age of 19.13 and a mean level of education of 1.29 years of study. The 20- to 29-year-old age group has an average age of 22.48 and an average level of education of 2.63 years.

Table 1(b) Sample distribution – Year of study

	Cell size (n)	Mean age
1st year	32	19.82
2nd year	27	21.05
3rd year	28	22.69
4th year	1	22.83
5th year	1	27.58
6th year	4	24.58

From Table 1(b) it can be seen that 32 persons tested are in their first year of study, 27 are in their second year of study, 28 are in their third year of study, 1 participant is in his or her fourth year of study, 1 participant is in his/her fifth year of study, and 4 are in their sixth year of study. The average age in the first-year group was 19.82 years; the average age in the second-year group was 21.05 years; 22.69 years in the third-year group; 22, 83 years in the fourth-year group; 27.58 years in the fifth-year group, and 24.58 years in the sixth year group.

Table 1(c) Sample distribution – Gender

	Cell size (n)	Mean age	Mean years of study
Female	50	21.44	2.2
Male	43	21.27	2.16

Table 1(c) shows that 50 persons tested are female and 43 persons tested are male. The female group has a mean age of 21.44 years and a mean level of education of 2.2 years, whilst the male group has a mean age of 21.27 years and a mean level of education of 2.16 years.

In Chapter 3, Maree & Pietersen (2007b) recommends cell sizes of at least 15 persons tested per group when aiming to compare subgroups with each other. In this study, 15 participants will thus be considered an adequate cell size.

In comparison to these standards, and from the tables above, it is evident that the sample size and cell sizes are adequate, with the exception of the fourth-, fifth- and sixth-year groups. These three groups can therefore be eliminated from the interpretive process.

The following section will present the investigation into the need for a standardisation of the WCST for the Setswana-speaking university population based on the data obtained from the sample.

4.2 Examining the Need for a Setswana Standardisation

The need to construct a standardisation for the Setswana-speaking university population is examined by compiling a table that depicts the mean and standard deviation for the 18- to 19-year-old and 20- to 29-year-old age groups, and comparing the data that has been obtained from these tables with the corresponding tables that can be obtained from Heaton et al. (1993). If a significant difference exists between the data, new normative data is required for the Setswana-speaking university population. The following section will present the mean and standard deviation for the each of the WCST sub-scores for both age groups. Thereafter, the corresponding data for the Setswana-speaking university population as obtained in this study will be presented.

4.2.1 Heaton et al.'s (1993) mean and standard deviation for each age group. This section will present the mean and standard deviation for the 18- to 19-year-old and the 20- to 29-year-old age groups in a similar manner to the normative tables as presented in Heaton et al. (1993). Table 2(a) displays the mean (M) and standard deviation (SD) for the 18- to19-year-old and 20- to 29-year-old age groups separately for each of the WCST sub-scores.

Table 2(a) Cell sizes, means and standard deviations of WCST scores for the separate age groups according to Heaton et al. (1993).

Group	Ages 18 -19 n = 56		Ages 20 - 29 n = 67	
	M	SD	M	SD
Number of categories completed	5.29	1.29	5.75	0.77
Total number of correct responses	71.05	10.37	70.10	8.75
Total number of errors	25.91	19.34	18.52	14.06
Perseverative responses	13.16	9.05	9.70	7.83
Perseverative errors	12.05	7.86	8.93	6.70
Non-perseverative errors	13.86	12.50	9.60	8.36
Trials to complete first category	14.38	9.09	11.67	2.44
Failure to maintain set	0.68	1.19	0.52	1.02
Learning to learn	-0.83[1]	5.43	-2.41[1]	5.26
Percent conceptual level responses	69.92	17.80	76.94	13.84

Table 2(a) shows that in the normative data presented by Heaton et al. (1993), the 18- to 19-year-old age group (n = 56), for example, the variable *Number of categories completed*

[1] Sample size for *Learning to learn* scores not available for the data presented in Heaton et al. (1993).

shows a mean of 5.29 and a standard deviation of 1.29 whilst the variable *Trials to complete first category* shows a mean of 14.38 and a standard deviation of 9.09.

Similarly, Table 2(a) shows that within the 20- to 29-year-old age group (n = 67) presented by Heaton et al (1993), for example, the variable *Total number of correct responses* shows a mean of 70.10 and a standard deviation of 8.75, and the variable *Perseverative responses* shows a mean of 9.70 and a standard deviation of 7.83.

The means and standard deviations by Heaton et al. (1993) presented above will be compared with the Setswana mean and standard deviation that follows.

4.2.2 Setswana mean and standard deviation for each age group. This section will present the mean and standard deviation for the 18- to 19-year-old and 20- to 29-year-old age groups obtained from the Setswana-speaking population in this study in a similar manner to the normative tables as presented in Heaton et al. (1993) (cf. 4.2.1). Table 2(b) will display the mean and standard deviation for the 18- to 19-year-old and 20- to 29-year-old age groups respectively for each of the WCST sub-scores.

Table 2(b) Cell sizes, means and standard deviations of WCST scores for the Setswana-speaking university population for the separate age groups

Group	Ages 18 -19 n = 31		Ages 20 - 29 n = 62	
	M	SD	M	SD
Number of categories completed	5.52	2.23	4.98	2.42
Total number of correct responses	85.13	16.37	79.94	18.42
Total number of errors	42.87	16.37	48.08	18.39
Perseverative responses	26.19	12.80	27.47	13.54
Perseverative errors	16.58	10.38	17.48	11.33
Non-perseverative errors	26.29	8.44	30.58	11.44
Trials to complete first category	12.00	2.38	16.34	15.40
Failure to maintain set	0.32	0.75	0.59	1.00
Learning to learn	-2.68[2]	4.67	-3.95[3]	5.07
Percent conceptual level responses	72.23	22.13	64.93	23.46

Table 2(b) shows that in the 18- to 19-year-old age group (n = 31), for example, the variable *Number of categories completed* shows a mean of 5.52 and a standard deviation of 2.23; the variable *Trials to complete first category* shows a mean of 12.00 and a standard deviation of 2.38.

Similarly, Table 2(b) shows that within the 20- to 29-year-old age group (n = 62), for example, the variable *Total number of correct responses* shows a mean of 79.94 and a standard deviation of 18.42; the variable *Perseverative responses* shows a mean of 27.47 and a standard deviation of 13.54.

The data constructed in table 2(a) and 2(b) above will now be compared with each other in order to determine whether new normative data is required for the Setswana population under investigation.

---

[2] Mean and standard deviation calculated using the 28 subjects for whom a *Learning to learn* score could be calculated.
[3] Mean and standard deviation calculated using the 48 subjects for whom a *Learning to learn* score could be calculated.

4.2.3 Comparing the two mean sets. Hypothesis tests are conducted to investigate whether the mean and standard deviation obtained for each age group in this study differs significantly from the means provided by Heaton et al. (1993). The hypotheses are set out as follows for each data set:

H_o: meanUS - meanSA = 0 if $p > \alpha$

H_a: meanUS - meanSA $\neq$ 0 if $p < \alpha$

If $p > \alpha$, the difference in the means does not differ significantly from 0 and it can be assumed that the means for the US and the current study is equal and come from the same sampling distribution. However, if, $p < \alpha$, the difference between the means does differ significantly from 0 and it can be assumed that the means for the US study and the current study are not equal and therefore come from different sampling distributions.

The hypothesis tests are conducted using $\alpha = 0.05$ to ensure a 95% confidence interval. The first investigation was between the two norm sets for the 18- to 19-year-old age group followed by the 20- to 29-year-old age group. This will be displayed in table format for each age group followed by the discussion of an example from the table presented in order to demonstrate the interpretation of the table. A table summarising the results of the hypothesis tests, and the discussion thereof, will then follow before the next age group is discussed.

4.2.3.1 Hypothesis tests between the two data sets for the 18- to 19-year-old age groups. Table 3(a) displays the comparison between the US means and the means for the current study for the 18- to 19-year-old age group for each of the WCST sub-scores.

Table 3(a) Hypothesis tests between the US and Setswana data sets for the 18- to 19-year-old age groups

Variable	Mean US	Mean Setswana	T value	Probability	Significance
Number of categories completed	5.29	5.52	-0.527473	0.600683	No significant difference
Total number of correct responses	71.05	85.13	-4.332565	0.0000853	Significant difference
Total number of errors	25.91	42.87	-4.332565	0.0000475	Significant difference
Perseverative responses	13.16	26.19	-5.016102	7.9871^{-6}	Significant difference
Perseverative errors	12.05	16.58	-2.117	0.0393238	Significant difference
Non-perseverative errors	13.86	26.29	-5.510564	4.0747^{-7}	Significant difference
Trials to complete first category	14.38	12.00	1.848227	0.0689463	No significant difference
Failure to maintain set	0.68	0.32	1.7274054	0.0877938	No significant difference
Learning to learn	-0.83	-2.68 [4]	N/A	N/A	N/A
Percent conceptual level responses	69.92	72.23	-0.282813	0.7784494	No significant difference

The hypothesis test in Table 3(a) shows that, for example, the p-value is greater than 0.05 ($p = 0.600693 > \alpha$) for the variable *Number of categories completed*. The hypothesis that the means are equal to each other can thus not be rejected. Therefore, the mean obtained from this study and the US mean for the variable *Number of categories completed* for the 18- to 19-year-old age group appears to come from the same sampling distribution. The 18- to

[4] Mean calculated for the 28 subjects for whom a *Learning to learn* score could be calculated.

19-year-old persons tested in this study and corresponding participants in the US study completed approximately the same number of category sorts.

As a further example, the hypothesis test in Table 3(a) shows that the p-value is smaller than 0.05 ($p = 7.9871^{-6} < \alpha$). The hypothesis that the means are equal to each other is thus rejected. Therefore the mean obtained from this study and the US mean for the variable *Perseverative responses* for the 18- to 19-year-old age group do not come from the same sampling distribution and significantly differ from each other. In this regard, the mean obtained in the current study (M = 26.19) is significantly higher than the mean presented in the US normative data (M = 13.16) for the 18- to 19–year-old age group. The 18- to nineteen-year-old persons tested in the current study thus made more perseverative sorts and performed poorer than their US counterparts.

Conducting a hypothesis test requires the mean, standard deviation and sample size for the two data sets being compared. Heaton et al. (1993) unfortunately do not provide a sample size for the *Learning to learn* score. The hypothesis test could thus not be conducted for this variable and the performance of the two groups could not be compared.

Table 3(b) summarises the findings of the hypothesis tests between the US and Setswana data sets for the 18- to 19-year-old age group.

Table 3(b) Summary of hypothesis tests for the 18- to 19-year-old age groups

	Is hypothesis that means are equal rejected or not?	**Is the US or Setswana mean higher?**	**Comparison of performance**
Number of categories completed	Do not reject hypothesis	N/A	Similar performance between groups
Total number of correct responses	Reject hypothesis	Study mean higher than US mean	Participants in Setswana study performed better than US norms
Total number of errors	Reject hypothesis	Study mean higher than US mean	Participants in Setswana study performed poorer than US norms
Perseverative responses	Reject hypothesis	Study mean higher than US mean	Participants in Setswana study performed poorer than US norms
Perseverative errors	Reject hypothesis	Study mean higher than US mean	Participants in Setswana study performed poorer than US norms
Non-perseverative errors	Reject hypothesis	Study mean higher than US mean	Participants in Setswana study performed poorer than US norms
Trials to complete first category	Do not reject hypothesis	N/A	Similar performance between groups
Failure to maintain set	Do not reject hypothesis	N/A	Similar performance between groups
Learning to learn[3]	Sample size not available	N/A	Similar performance between groups
Percent conceptual level responses	Do not reject hypothesis	N/A	Similar performance between groups

From Table 3(b) it can be seen that the hypothesis tests shows that the means for the variables *Number of categories completed*, *Trials to complete first category*, *Failure to maintain set* and *Percent conceptual level responses* do not differ significantly from the US means. It can therefore be deduced that the US and Setswana samples performed similarly on these sub-scores.

The means for the variables *Total number of correct responses*, *Total number of errors*, *Perseverative responses*, *Perseverative errors* and *Non-perseverative errors*, however, differ significantly from each other.

For each of these variables, the mean obtained in the current study is higher than that of the US study obtained from Heaton et al. (1993). This indicates that on the variable *Total number of correct responses,* 18-to-19-year-old persons tested in the current study performed better than their US counterparts, but on the variables *Total number of errors, Perseverative responses, Perseverative errors and Non-perseverative errors*, persons tested in the current study performed poorer than their US counterparts.

It can thus be assumed that the data obtained in this study for the 18- to 19-year-old age group, does not follow the same distribution as the data provided by Heaton et al. (1993).

4.2.3.2 Hypothesis tests between the two data sets for the 20-to-29-year-old age groups. Table 4(a) will display the comparison between the US means and the means for the current study for the 20-to-29 year-old age group for each WCST sub-score.

Table 4(a) Hypothesis tests between the US and Setswana data sets for the 20-to-29-year-old age groups

Variable	Mean US	Mean Setswana	T value	Probability	Significance
Number of categories completed	5.75	4.98	2.39566	0.0191779	Significant difference
Total number of correct Responses	70.10	79.94	-3.825798	0.0002469	Significant difference
Total number of errors	28.52	48.08	-10.19601	0.00	Significant difference
Perseverative responses	9.70	27.47	-9.030643	1.776^{-14}	Significant difference
Perseverative errors	8.93	17.48	-5.161362	1.2952^{-6}	Significant difference
Non-perseverative errors	9.60	30.58	-11.68054	0.00	Significant difference
Trials to complete first category	11.67	16.34	-2.360507	0.0213167	Significant difference
Failure to maintain set	0.52	0.59	-0.449787	0.653633	No significant difference
Learning to learn	-2.41	-3.95 [5]	N/A	N/A	N/A
Percent conceptual level responses	76.94	64.93	3.5028695	0.000697	Significant difference

The hypothesis test in Table 4(a) shows, for example, that the p-value is smaller than 0.05 ($p = 0.0002469 < \alpha$) for the variable *Total number of correct responses*. The hypothesis that the means are equal to each other is thus rejected. Therefore, the mean obtained from this study and the US mean for the variable *Total number of correct responses* for the 20- to

[5] The mean was calculated using the 48 subjects for whom a *Learning to learn* score could be calculated.

29-year-old age groups do not to come from the same sampling distribution and differ significantly from each other. In this regard, the mean obtained in the current study (M = 79.94) is significantly higher than the mean presented in the US normative data (M = 70.10) for the 20- to 29-year-old age groups. The 20- to 29–year-old persons tested in the current study achieved more correct sorts and performed better than their US counterparts.

As a further example, the hypothesis test in Table 4(a) shows that the p-value is smaller than 0.05 ($p = 1.776^{-14} < \alpha$) for the variable *Perseverative responses*. The hypothesis that the means are equal to each other is thus rejected. Therefore, the mean obtained from this study and the US mean for the variable *Perseverative responses* for the 20- to 29-year-old age groups do not to come from the same sampling distribution and differ significantly from each other. In this regard, the mean obtained in the current study (M = 27.47) is significantly higher than the mean presented in the US normative data (M = 9.70) for the 20- to 29-year-old age groups. The 20- to 29-year-old persons tested in the current study sorted more cards

As previously stated, Heaton et al. (1993) do not provide a sample size for the *Learning to learn* score as the hypothesis test for this variable could not be conducted whereby performances between the groups could be compared (cf. par. 4.2.3.1).

Table 4(b) will now display the summary of the hypothesis tests between the US and Setswana means for the 20- to 29–year-old age groups.

Table 4(b) Summary of the hypothesis tests for the 20- to 29-year-old age groups

	Is hypothesis that means are equal rejected or not?	**Is the US or Setswana mean higher?**	**Comparison of performance**
Number of categories completed	Reject hypothesis	Setswana mean lower than US mean	Participants in current study performed poorer than US norms
Total number of correct responses	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed better than US norms
Total number of errors	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed poorer than US norms
Perseverative responses	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed poorer than US norms
Perseverative errors	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed poorer than US norms
Non-perseverative errors	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed poorer than US norms
Trials to complete first category	Reject hypothesis	Setswana mean higher than US mean	Participants in current study performed poorer than US norms
Failure to maintain set	Do not reject hypothesis	N/A	Similar performance between groups
Learning to learn	No sample size available	N/A	Similar performance between groups
Percent conceptual level responses	Reject hypothesis	Setswana mean lower than US mean	Similar performance between groups

From Table 4(b) it can thus be seen that the hypothesis tests show that only the means for the variable *Failure to maintain set* did not differ significantly from the US mean – it can thus be assumed that the US and Setswana samples performed similarly on this variable.

All the other variables (*Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Trials to complete first category* and *Percent conceptual level responses*) differ significantly from the US means. The Setswana means for the variables *Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors* and *Trials to complete first category* are higher than the US means, whilst the Setswana means for the variables *Number of categories completed* and *Percent conceptual level responses* are lower than the US means.

This indicates that on the variable *Total number of correct responses,* participants in the Setswana study performed better than their US counterparts, but on the variables *Number of categories completed, Total number of errors Perseverative responses, Perseverative errors, Non-perseverative errors, Trials to complete first category* and *Percent conceptual level responses*, persons tested in the current study performed poorer than their US counterparts.

It can thus be assumed that the data obtained in this study for the 20- to 29-year-old age groups, does not follow the same distribution as the data provided by Heaton et al. (1993).

4.2.3.3 The need for a new standardisation. From the hypothesis tests discussed above (cf. 4.2.3), it is evident that the data obtained for the Setswana-speaking university population obtained in this study does not follow the same distribution as the data presented in Heaton et al. (1993) for the US population of the same age groups.

When comparing Setswana-speaking university students between the ages of 19 and 29 to the US norms currently in use, a skewed indication of their WCST performance may thus be obtained. In order to compare Setswana-speaking university students to a standardisation sample consisting of their peers, the construction of a new standardisation is required. The following section will discuss the construction of a preliminary standardisation for the Setswana-speaking university population.

4.3 Conducting the Preliminary Standardisation

This section aims to construct a preliminary standardisation of the WCST for a Setswana-speaking university population between the ages of 18 and 29. This will be done by examining the impact of the demographic variables *gender*, *age* and *level of education* on WCST performance. This is done to determine whether separate norm tables are required for the separate levels of these variables. This procedure, however, necessitates examining the distribution of the overall data set to ascertain which techniques are suitable to investigate the impact of the demographic variables.

Once the impact of the demographic variables has been determined, norm tables will be constructed for each separate variable that influences WCST performance. This will then be followed by a calculation of some of the psychometric properties of the WCST for the study population. This process will now commence by documenting the investigations into the distribution of the data.

4.3.1 Examining the distribution of the data. This section investigates the distribution of the data as required for further statistical procedures that will depend on the distribution of the overall data set. Firstly, histograms depicting the frequency distributions of each WCST score will be discussed (cf. Appendix D) to visually display the distribution of the data. Thereafter, investigations examining the normal distribution of the data will be displayed in table format and discussed for each sub-score.

The histograms provided in Appendix D visually represent the distribution of the various WCST scores. These depictions provide an indication that the data is not normally distributed over WCST scores. On this basis, it is assumed that the factorial ANOVA may not be suitable to determine the impact of the demographic variables. Regression analysis may be required as the data may not be normally distributed. This, however, requires further investigation.

To further investigate whether the data is normally distributed the Shapiro–Wilk, Kolmogorov–Smirnov, Cramer–Von Mises and Anderson–Darling tests were completed. These were conducted using $\alpha = 0.05$ ensuring a confidence level of 95%.

These statistical tests work on the principle of hypothesis testing, with the null hypothesis being normally distributed data and the alternative hypothesis being non-normally distributed data. If the p-value is less than 0.05 ($p < \alpha$) the null hypothesis is rejected at the 5% level of significance and the data differs significantly from a normal distribution. If the p-value is greater than 0.05 ($p > \alpha$) the null hypothesis cannot be rejected and the data does not differ significantly from the normal distribution.

The results of the investigations will now be displayed in table format. After each table a short explanation of the findings will be given. Once all the tables have been presented, a table summarising the tests for normality, and a discussion thereof, will be provided.

Table 5(a) Test for normality of distribution of the *Number of categories completed* variable

Test for normality	**Statistic**	**P-value (p)**	**Significance**
Shapiro–Wilk	W 0.950235	Pr<W 0.0014	Significant
Kolmogorov–Smirnov	D 0.140249	Pr>D <0.0100	Significant
Cramer–Von Mises	W-Sq 0.267725	Pr>W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 1.638549	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(a), the p-value ranges between 0.0014 and 0.0100, this is smaller than 0.05. ($p < \alpha$; p = <0.0014 to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Number of categories completed* therefore differs significantly from a normal distribution.

Table 5(b) Test for normality of distribution of the *Total number of correct responses* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.948862	Pr<W 0.0011	Significant
Kolmogorov–Smirnov	D 0.105991	Pr>D 0.0109	Significant
Cramer–Von Mises	W-Sq 0.19126	Pr>W-Sq 0.0069	Significant
Anderson–Darling	A-Sq 1.326297	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(b), the p-value ranges between 0.0011 and 0.0109, this is smaller than 0.05. ($p < \alpha$; p = <0.0011 to 0.0109; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Total number of correct response* therefore differs significantly from a normal distribution.

Table 5(c) Test for normality of distribution of the *Total number of errors* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.948277	Pr < W 0.0011	Significant
Kolmogorov–Smirnov	D 0.106367	Pr > D 0.0102	Significant
Cramer–Von Mises	W-Sq 0.193367	Pr > W-Sq 0.0064	Significant
Anderson–Darling	A-Sq 1.343877	Pr > A-Sq < 0.0050	Significant

In each of the tests as mentioned in Table 5(c), the p-value ranges between 0.0011 and 0.0102, this is smaller than 0.05. ($p < \alpha$; p = <0.0011 to 0.0102; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Total number of error* therefore differs significantly from a normal distribution.

Table 5(d) Test for normality of distribution of the *Perseverative responses* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.912726	Pr < W <0.0001	Significant
Kolmogorov–Smirnov	D 0.12528	Pr > D <0.0100	Significant
Cramer–Von Mises	W-Sq 0.323999	Pr > W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 1.883777	Pr > A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(d), the p-value ranges between less than 0.0001 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = <0.0001$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Perseverative responses* therefore differs significantly from a normal distribution.

Table 5(e) Test for normality of distribution of the *Perseverative errors* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.917052	Pr < W < 0.0001	Significant
Kolmogorov–Smirnov	D 0.143051	Pr > D < 0.0100	Significant
Cramer–Von Mises	W-Sq 0.405368	Pr > W-Sq < 0.0050	Significant
Anderson–Darling	A-Sq 2.335528	Pr > A-Sq < 0.0050	Significant

In each of the tests as mentioned in Table 5(e), the p-value ranges between less than 0.0001 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = <0.0001$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Perseverative errors* therefore differs significantly from a normal distribution.

Table 5(f) Test for normality of distribution of the *Non-perseverative errors* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.935676	Pr < W 0.0002	Significant
Kolmogorov–Smirnov	D 0.086263	Pr > D 0.0871	Not significant
Cramer–Von Mises	W-Sq 0.195557	Pr > W-Sq 0.0090	Significant
Anderson–Darling	A-Sq 1.486439	Pr > A-Sq < 0.0050	Significant

In the Shapiro–Wilk, Cramer–Von Mises and Anderson–Darling tests, as mentioned in Table 5(f), the p-value ranges between 0.0002 and 0.0090 ($p < \alpha$; $p = 0.0002$ to 0.0090; $\alpha = 0.05$) and the hypothesis that the data is normally distributed, is rejected. The Kolmogorov–Smirnov test, however, shows a p-value of 0.0871 which is greater than 0.05 ($p > \alpha$; $p = 0.0871$; $\alpha = 0.05$). The hypothesis that the data is normally distributed can, in this case, not be rejected. It can thus be said that the data for the variable *Non-perseverative errors* is approximates a normal distribution.

Table 5(g) Test for normality of distribution of the *Trials to complete first category* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.369057	Pr<W <0.0001	Significant
Kolmogorov–Smirnov	D 0.355788	Pr>D <0.0100	Significant
Cramer–Von Mises	W-Sq 4.258075	Pr>W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 20.93425	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(g), the p-value ranges between less than 0.0001 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = <0.0001$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Trials to complete first category* therefore differs significantly from a normal distribution.

Table 5(h) Test for normality of distribution of the *Failure to maintain set* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.603958	Pr<W <0.0001	Significant
Kolmogorov–Smirnov	D 0.405884	Pr>D <0.0100	Significant
Cramer–Von Mises	W-Sq 3.049514	Pr>W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 15.92537	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(h), the p-value ranges between less than 0.0001 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = <0.0001$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Failure to maintain set* therefore differs significantly from a normal distribution.

Table 5(i) Test for normality of distribution of the *Learning to learn*[6] variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.877758	Pr<W <0.0001	Significant
Kolmogorov–Smirnov	D 0.182164	Pr>D <0.0100	Significant
Cramer–Von Mises	W-Sq 0.695382	Pr>W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 3.853924	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(i), the p-value ranges between less than 0.0001 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = <0.0001$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Learning to learn* therefore differs significantly from a normal distribution.

[6] These investigations of normality were conducted for the 76 participants for whom a *Learning to learn* score could be calculated.

Table 5(j) Test for normality of distribution of the *Percent conceptual level responses* variable

Test for normality	Statistic	P-value (p)	Significance
Shapiro–Wilk	W 0.95149	Pr<W 0.0017	Significant
Kolmogorov–Smirnov	D 0.112773	Pr>D <0.0100	Significant
Cramer–Von Mises	W-Sq 0.215894	Pr>W-Sq <0.0050	Significant
Anderson–Darling	A-Sq 1.403599	Pr>A-Sq <0.0050	Significant

In each of the tests as mentioned in Table 5(j), the p-value ranges between 0.0017 and 0.0100, this is smaller than 0.05. ($p < \alpha$; $p = 0.0017$ to 0.0100; $\alpha = 0.05$). The null hypothesis that the data is normally distributed is thus rejected. The distribution of the data for the variable *Percent conceptual level responses* therefore differs significantly from a normal distribution.

Table 5(k) summarises all the results reported in the tables above.

Table 5(k) Summary of normality investigations

	Comment on distribution
Number of categories completed	Normal distribution rejected
Total number of correct responses	Normal distribution rejected
Total number of errors	Normal distribution rejected
Perseverative responses	Normal distribution rejected
Perseverative errors	Normal distribution rejected
Non-perseverative errors	Approximately normally distributed
Trials to complete first category	Normal distribution rejected
Failure to maintain set	Normal distribution rejected
Learning to learn[6]	Normal distribution rejected
Percent conceptual level responses	Normal distribution rejected

From Tables 5(a) to 5(k) it is possible to see that at a 95% confidence level the hypothesis that the data is normally distributed is rejected for most variables except for the variable *Non-perseverative errors*. The data is thus not normally distributed for the majority of the variables are thus not distributed normally, with *Non-perseverative errors* approximating normality. This eliminates the possibility of making use of the factorial ANOVA to determine the impact of demographic variables, as the factorial ANOVA requires normally distributed data.

The following section will discuss the use of regression analysis in order to determine the impact of demographic variables on the data set.

4.3.2 Examining the impact of demographic variables. Regression analysis has been conducted to investigate whether a significant relationship exists between the variables *gender, age* and *level of education* and either of the WCST sub-scores. Regression analysis works on the principle of hypothesis testing where:

H_0: $\beta = 0$ if $p > \alpha$ and H_A: $\beta \neq 0$ if $p < \alpha$

The estimate value (β) provides and indication of the impact that the variable has on the data set. If the β is close to zero the variable does not influence the data set. Hypothesis testing is used with the null hypothesis being $\beta = 0$.

If $p > \alpha$, the null hypothesis cannot be rejected, indicating that the estimate does not have a significant influence on the data. In this case, the estimate does not significantly differ from zero and the demographic variable does not influence the WCST variable under investigation.

However, if $p<\alpha$, the null hypothesis is rejected, necessitating the alternative hypothesis of $\beta \neq 0$. The estimate is therefore not significantly close to zero and the demographic variable has a significant influence on the WCST score.

All regressions were conducted allowing one degree of freedom and using a 95% confidence interval (α = 0.05). The regression is conducted on the entire data set as to determine the impact of the demographic variables thereon. This information will be used in evaluating the necessity of each separate norm table as discussed above. If the demographic variable influences the data, separate norm tables are needed. If the variable does not influence the data, one norm table can be used for all participants.

Tables 6(a) to 6(j) document the findings of the regression analysis. Each table will be followed by a discussion of the findings for each variable. Thereafter, Table 6(k) summarises the findings from the regression analysis and is followed by a discussion thereof.

Table 6(a) Regression analysis for the WCST variable *Number of categories completed*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	7.46692	0.0082	
Age	-0.00784	0.5042	Not significant
Gender	-0.02600	0.9582	Not significant
Years of study	-0.11725	0.6591	Not significant

From Table 6(a) it can be seen that for the variables *age* ($p = 0.5042 > \alpha$), *gender* ($p = 0.9582 > \alpha$) and *years of study* ($p = 0.6591 > \alpha$), the p-value is greater than 0.05 and the hypothesis that $\beta = 0$ cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Number of categories completed*.

Table 6(b) Regression analysis for the WCST variable *Total number of correct responses*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	116.31170	< 0.0001	
Age	-0.13703	0.1214	Not significant
Gender	-0.54126	0.8843	Not significant
Years of study	0.60254	0.7622	Not significant

From Table 6(b) it can be seen that for the variables *age* (p = 0.1214 > α), *gender* (p = 0.8843 > α) and *years of study* (p = 0.7622 > α), the p-value is greater than 0.05 and the hypothesis that β = 0 cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Total number of correct responses*.

Table 6(c) Regression analysis for the WCST variable *Total number of errors*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	11.72955	0.5718	
Age	0.13670	0.1220	Not significant
Gender	0.56152	0.8799	Not significant
Years of study	-0.59181	0.7661	Not significant

From Table 6(c) it can be seen that for the variables *age* (p = 0.1220 > α), *gender* (p = 0.8799 > α) and *years of study* (p = 0.7661 > α), the p-value is greater than 0.05 and the hypothesis that β = 0 cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Total number of errors*.

Table 6(d) Regression analysis for the WCST variable *Perseverative responses*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	4.06604	0.7929	
Age	0.09550	0.1478	Not significant
Gender	0.66741	0.8100	Not significant
Years of study	-1.15976	0.4360	Not significant

From Table 6(d) it can be seen that for the variables *age* (p = 0.1478 > α), *gender* (p = 0.8100 > α) and *years of study* (p = 0.4360 > α), the p-value is greater than 0.05 and the hypothesis that β = 0 cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Perseverative responses*.

Table 6(e) Regression analysis for the WCST variable *Perseverative errors*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	-3.66778	0.7740	
Age	0.09501	0.0816	Not significant
Gender	-0.09564	0.9667	Not significant
Years of study	-1.53905	0.2112	Not significant

From Table 6(e) it can be seen that for the variables *age* ($p = 0.0816 > \alpha$), *gender* ($p = 0.9667 > \alpha$) and *years of study* ($p = 0.2112 > \alpha$), the p-value is greater than 0.05 and the hypothesis that $\beta = 0$ cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Perseverative errors.*

Table 6(f) Regression analysis for the WCST variable *Non-perseverative errors*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	15.35607	0.2179	
Age	0.04202	0.4249	Not significant
Gender	0.63690	0.7747	Not significant
Years of study	0.93651	0.4324	Not significant

From Table 6(f) it can be seen that for the variables *age* ($p = 0.4249 > \alpha$), *gender* ($p = 0.7747 > \alpha$) and *years of study* ($p = 0.4324 > \alpha$), the p-value is greater than 0.05 and the hypothesis that $\beta = 0$ cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Non-perseverative errors*.

Table 6(g) Regression analysis for the WCST variable *Trials to complete first category*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	-23.47190	0.1084	
Age	0.16999	0.0068	Significant
Gender	-0.33017	0.8990	Not significant
Years of study	-2.15527	0.1243	Not significant

From Table 6(g) it can be seen that for the variables *gender* (p = 0.8990 > α) and *years of stud*y (p = 0.1243 > α), the p-value is greater than 0.05 and the hypothesis that β = 0 cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Trials to complete first category*.

However, for the variable *age* (p = 0.0068 < α), the p-value is less than 0.05. The hypothesis that β = 0 has to be rejected and it is assumed that the estimate value differs significantly from zero and therefore significantly impacts the variable *Trials to complete first category*. The 20- to 29-year-old age group (M = 16.34) performed better on the WCST than the 18- to 19-year-old age group (M= 12.00).

Table 6(h) Regression analysis for the WCST variable *Failure to maintain set*

	Estimate (β)	**P-value (p)**	**Significance**
Intercept	-1.15913	0.2861	
Age	0.00677	0.1422	Not significant
Gender	0.10866	0.5760	Not significant
Years of study	-0.10951	0.2936	Not significant

From Table 6(h) it can be seen that for the variables *age* (p = 0.1422 > α), *gender* (p = 0.5760 > α) and *years of study* (p = 0.2936 > α), the p-value is greater than 0.05 and the hypothesis that β = 0 cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable Failure to maintain set.

Table 6(i) Regression analysis for the WCST variable *Learning to learn*[7]

	Estimate (β)	P-value (p)	Significance
Intercept	-2.06404	0.7604	
Age	-0.01261	0.6723	Not significant
Gender	1.85166	0.1059	Not significant
Years of study	-0.48470	0.5323	Not significant

From Table 6(i) it can be seen that for the variables *age* ($p = 0.6723 > \alpha$), *gender* ($p = 0.1059 > \alpha$) and *years of study* ($p = 0.5323 > \alpha$), the p-value is greater than 0.05 and the hypothesis that $\beta = 0$ cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Learning to learn*.

Table 6(j) Regression analysis for the WCST variable *Percent conceptual level responses*

	Estimate (β)	P-value (p)	Significance
Intercept	99.55596	0.0004	
Age	-0.12505	0.2755	Not significant
Gender	0.28231	0.9535	Not significant
Years of study	-0.41222	0.8735	Not significant

From Table 6(j) it can be seen that for the variables *age* ($p = 0.2755 > \alpha$), *gender* ($p = 0.9535 > \alpha$) and *years of study* ($p = 0.8735 > \alpha$), the p-value is greater than 0.05 and the hypothesis that $\beta = 0$ cannot be rejected. The estimate value is thus close to zero and the variables age, gender and level of education do not influence the variable *Percent conceptual level responses*.

[7] This regression analysis was conducted using the 76 participants for whom a *Learning to learn* score could be calculated.

Table 6(k) Summary of regression analysis

	Age	Gender	Year of Study
Number of categories completed	No impact	No impact	No impact
Total number of correct responses	No impact	No impact	No impact
Total number of errors	No impact	No impact	No impact
Perseverative responses	No impact	No impact	No impact
Perseverative errors	No impact	No impact	No impact
Non-perseverative errors	No impact	No impact	No impact
Trials to complete first category	Significant impact	No impact	No impact
Failure to maintain set	No impact	No impact	No impact
Learning to learn[9]	No impact	No impact	No impact
Percent conceptual level responses	No impact	No impact	No impact

From Tables 6(a) to 6(k) it can be seen that the regression analysis shows that overall the polynomial regression analysis revealed no significant quadratic curvilinear relationship between the WCST scores and age, gender or level of education, with one exception being the variable *Trials to complete first category* which is significantly impacted by the variable age.

From a statistical perspective, no evidence was found to motivate for separate norm tables for the separate age-, gender- and level of education groups with the exception of the variable *Trials to complete first category* where age may be considered to construct separate norm tables. The following section will thus present preliminary normative data for the Setswana-speaking university population on the basis of the regression analysis that has been discussed.

4.3.3 Presenting preliminary normative data. The regression analysis above indicated that separate norm tables are not required for age, gender or level of education for most of the WCST sub-scores. The only exception to this is the variable *Trials to complete first* category which showed a significant difference between the age groups and thus requires separate norm tables for the two age groups under investigation.

Table 7(a) will now present the preliminary normative data for the entire 18- to 29-year-old age group. Tables 7(b) and 7(c) will then present the preliminary normative data for the separate age groups for the variable *Trials to complete first category*. Expanded normative data including the median, mode, range and variance may be seen in Appendix E.

Table 7(a) Preliminary normative data for Setswana-speaking university students between the ages of 18 and 29 years

N=93	Mean	Standard Deviation
Number of categories completed	5.16	2.36
Total number of correct responses	81.67	17.84
Total number of errors	46.34	17.83
Perseverative responses	27.04	13.26
Perseverative errors	17.18	10.98
Non-perseverative errors	29.15	47.00
Failure to maintain set	0.51	0.92
Learning to learn[8]	-3.48	4.93
Percent conceptual level responses	67.03	23.10

Table 7(a) shows that the overall data set (N = 93) displays, for example, *Number of categories completed* with a mean of 5.16 and a standard deviation of 2.36 variable whilst *Total number of correct responses* has a mean of 81.67 and a standard deviation of 17.84.

Table 7(b) Preliminary normative data for Setswana-speaking university students between the ages of 18 and 19 years for the variable *Trials to complete first category*

N=31	Mean	Standard Deviation
Trials to complete first category	12.00	2.38

Table 7(b) shows that the 18- to 19-year-old age group (N = 31) displays a mean of 12.00 and a standard deviation of 2.38.

---

[8] Note the all calculations for the *Learning to learn* score are based on the 76 subjects for whom a learning score could be calculated.

Table 7(c) Preliminary normative data for Setswana-speaking university students between the ages of 20 and 29 years for the variable *Trials to complete first category*

N=62	Mean	Standard Deviation
Trials to complete first category	16.34	15.40

Table 7(c) shows that the 20- to 29-year-old age group (N = 62) displays a mean of 16.34 a standard deviation of 15.40.

4.3.4 Psychometric properties. This section will discuss the results of the investigations into the psychometric properties of the WCST for use with the Setswana-speaking university population.

4.3.4.1 Reliability. An instrument with a high Cronbach alpha coefficient (close to 1.00) will consist of parts that are so interrelated that they form a whole implying that the items are similar in content and the construct they measure (Gatewood, Feild & Barrick, 2008). The test can then be seen as a measure of a unified construct. The widely accepted social science cut-off is that alpha (α) should be at least 0.70 (Gatewood, Feild & Barrick, 2008).

Table 8 Reliability (internal consistency) of the WCST for Setswana-speaking university students.

Cronbach Alpha Coefficient	0.85

This study found a Cronbach Alpha Coefficient of 0.85 (as seen in Table 8) for use with the WCST in the Setswana-speaking university population. This is lower than the alpha of 0.96 reported by Loeber, Duka, Welzel, Nakovicz, Heinz, Flor and Mann (2009), but still indicates a high level of internal consistency between the responses. (cf. par. 2.5.5.3.1). There is thus a high degree of correlation between items within the test. The items within the WCST are thus consistent enough to include them in one scale that measures one construct.

This becomes especially significant when one takes into consideration that the Cronbach Alpha procedure provides a conservative underestimation of reliability (Brown, 2002). Furthermore, the results must be considered within the context that normally distributed data tends to produce higher Cronbach Alpha coefficients (Brown, 2002). It can thus be assumed that a coefficient of 0.85 is an under estimation of the internal consistency of the WCST, as the data obtained in this study is not normally distributed.

According to Dornyei and Taguchi (2010) although internal consistency only covers one aspect of reliability, it is often surprisingly close to the actual reliability of a test. Brown (2002), however, outlines the importance of considering that reliability, regardless of the type of reliability, is not an inherent test property but rather an estimation of the consistency of an item set administered to a certain population under certain conditions. The WCST thus appears to have a high degree of internal consistency when used in the Setswana-speaking university population of South Africa.

4.3.4.2 Standard Error of Measurement (SEM). The Cronbach Alpha calculated above, however, provides an overall indication of internal consistency reliability and cannot be calculated for individual sub-scores.

As discussed in Chapter 3 (cf. par. 3.3.3) the WCST only provides sub-scores and no overall score, calculating an SEM does not provide meaningful information regarding the interval within sub-scores are to be interpreted. On the basis of the Cronbach reliability coefficient, the SEM could not be calculated meaningfully for the Setswana-speaking university population. Further investigation is thus required to determine alternative means of calculating the SEM for the WCST.

4.4 Discussion

This chapter started with an investigation on the need for re-standardising the WCST for the Setswana-speaking university population by comparing the mean scores of the Setswana-speaking university population with the US mean scores presented in Heaton et al. (1993). These investigations indicated that the test needs to be re-standardised, as normative data by Heaton et al. did not follow the same distribution as that of the Setswana-speaking university population.

This research project thus aimed to construct a preliminary standardisation of the WCST for Setswana-speaking university students between the ages of 18 and 29 years. In order to do this, the testing instructions for the WCST were translated to Setswana. These instructions were used to assess participants who were matching the sampling criteria on a computerised version of the WCST. This served the purpose of standardising the testing instructions and scoring procedures for the Setswana-speaking university population. Furthermore, normative tables were constructed on the basis of investigations on the impact of age, gender and level of education on WCST sub-score.

The following section will aim to discuss the findings of the study in more detail and aim to interpret these findings by taking the factors limiting the interpretation of the results into consideration. On the basis of this discussion, recommendations for further research will then be made.

4.4.1 Findings. Ninety-three Setswana-speaking university students that attend the University of Limpopo (Medunsa Campus) were included in the study (N = 93). Adequate cell sizes were obtained for the strata male (n = 50), female (n = 43), 18- to 19-year-old age group (n = 31), 20- to 29-year-old age group (n =62), first-year level of education (n = 32), second-year level of education (n = 27) and third-year level of education (n = 28). Fourth-year level of education (n = 1), fifth-year level of education (n = 1) and sixth-year level of education (n = 4) had to be excluded due to inadequate sample sizes.

Firstly, hypothesis tests were conducted to compare the means obtained from this study with the mean currently in use that has been constructed by Heaton et al (1993). The two age groups under investigation were respectively compared to each other in order to determine whether Heaton et al.'s norms may be used with the Setswana-speaking university population or whether the WCST requires a new standardisation for the Setswana-speaking university population.

For the 18- to 19-year-old age group the variables *Number of categories completed, Trials to complete first category, Failure to maintain set* and *Percent conceptual level responses* showed no significant difference between the means, whilst the variables *Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors* and *Non-perseverative errors* showed a significant difference with the mean obtained in the current study being higher than the US mean.

This indicates that on the variable *Total number of correct responses*, 18- to 19-year-old persons tested in the current study performed better than the US norm, whereas on the variables *Total number of errors, Perseverative responses, Perseverative errors* and *Non-Perseverative errors*, persons tested in the current study performed poorer than the US norms.

It can thus be concluded that the data obtained from this study, for the 18- to 19-year-old age group does not follow the same distribution as the distribution of the normative information constructed by Heaton et al. (1993) for the corresponding age group (cf. par. 4.2.3.1).

Heaton et al.'s (1993) normative data may thus be appropriate for interpreting the variables *Number of categories completed, Trials to complete first category, Failure to maintain set* and *Percent conceptual level responses* in the 18- to 19-year-old age group. When considering the variables *Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors* and *Non-perseverative errors*, Heaton et al.'s norms may provide a skewed indication of WCST performance in the 18- to 19-year-old age group. Normative data specific to the Setswana university population is thus required when interpreting these scores.

For the 20- to 29-year-old age group the variable *Failure to maintain set* showed no significant difference between the two means, whilst the variables *Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors Trials to complete first category* and *Percent conceptual level responses* showed a significant difference. The mean obtained in the current study was higher than the US means for the variables *Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors* and *Trials to complete first category*. However, for the variables *Number of categories completed* and *Percent conceptual level responses* the mean obtained from the current study was lower than the US mean.

On the variable *Failure to maintain set*, the two groups performed similarly, while on the variable *Total number of correct responses* the Setswana population performed better than the US norms, but on the *Number of categories completed, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Trials to complete first category* and *Percent conceptual level responses* variables, the Setswana population performed poorer than the US population.

It can therefore be concluded that the data obtained from this study, for the 20- to 29-year-old age group, does not follow the same distribution as the distribution of the normative data as constructed by Heaton et al. (1993) for the corresponding age group (cf. par. 4.2.3.2).

The normative data constructed by Heaton et al. (1993) may thus be appropriate for interpreting the variable *Failure to maintain set* in the 20- to 29-year-old age group. When considering the variables *Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Trials to complete first category* and *Percent conceptual level responses*, the norms by Heaton et al. may, however, provide a skewed indication of WCST performance in the 20- to 29-year-old age group. Normative data specific to the Setswana university population is thus required when interpreting the majority of scores for the 20- to 29-year-old age group.

On the whole, it can be concluded that the Setswana norms attained in this study do not follow the same distribution as the US norms currently in use, with most of the variables showing lowered performance by Setswana-speaking university students in comparison to their US counterparts of the same age. This is especially significant when one takes into consideration that the sample for the current study consists of university students whereas the US sample consisted of university students as well as lower levels of education. Setswana students may thus be at a disadvantage when assessed on the WCST using the norms provided by Heaton et al. (1993). On this basis, it is thus assumed that the WCST needs to be re-standardised for the Setswana-speaking university population.

This study thus endeavoured to construct a preliminary standardisation of the WCST for the Setswana-speaking university population. This required translating the testing instructions to Setswana and administering the WCST to 93 participants using a computerised version of the WCST making use of the Setswana instructions. Care was taken to ensure the translations maintained the vague nature and gist of the original WCST instructions. This, in conjunction with the computerised scoring system, served as standardising the administration and scoring procedures of the WCST for the Setswana-speaking university population.

The data obtained from the persons tested were then used to construct preliminary normative data for the use of the WCST with the population under investigation. This required investigating the distribution of the data in order to determine which further statistical procedures might be employed. Investigations into the normality of the distribution of the overall data set indicated that the data is not normally distributed for most of the variables (*Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Trials to complete first category, Failure to maintain set, Learning to learn, Percent conceptual level responses*), except the variable *Non-perseverative errors* which approximates a normal distribution (cf. par. 4.3.1).

Once the distribution of the data was established, investigations into the impact of age, gender and level of education was required to determine which variables requires consideration in constructing normative tables. As the data was not normally distributed, regression analysis was employed to investigate whether the demographic variables influence the WCST score variables. The regression analysis showed that neither age, gender, nor level of education influences the WCST sub-scores (*Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Failure to maintain set, Learning to learn, Percent conceptual level responses*) with the exception of *Trials to complete first category,* which is influenced by age but not by level of education or gender, with the 20- to 29-year-old age group performing better than the 18- to 19-year-old age group (cf. par. 4.3.2). These findings are in line with reports that executive functions mature and plateau after the teenage years, but differ from the normative data obtained from Heaton et al. (1993) as they found that age and level of education impact WCST performance on all sub-scores. These finding may, however, be accounted for by the fact that only two age groups were included in the study and requires confirmation by a further study investigating the impact of age, gender and level of education on WCST performance on the entire Setswana population (including a wider variety of age groups).

However, on the basis of the regression analysis conducted in thus study, one norm table was constructed for the entire 18- to 29-year-old age group for all the variables, excluding *Trials to complete first category,* for which two separate norm tables were constructed under the 18- to 19-year-old and 20- to 29-year-old age groups respectively (cf. par. 4.3.3).

These norms may provide a fairer indication of WCST performance when used with the Setswana-speaking university students. However, it needs to be noted that these norms may place Setswana-speaking participants between the ages of 18 and 29 at a disadvantage if they have not been educated at university level.

Subsequently, preliminary psychometric properties were presented for use of the WCST with Setswana-speaking university students. A Cronbach Alpha Coefficient of 0.85 was found in the Setswana-speaking university population (cf. par. 4.3.4). This indicates that the WCST displays adequate internal consistency in the Setswana-speaking university population. On the basis of the Cronbach Alpha Coefficient, a meaningful SEM could, however, not be calculated.

4.4.2 Issues for consideration. When interpreting the findings of this study, the following issues require consideration, as they may have influenced the findings of the study:

4.4.2.1 The normal distribution. One limitation of this study is the fact that the data was distributed irregularly. The data obtained from this study will thus not follow the predictable traits and probabilities expected from a normal distribution. Assumptions characteristic to the normal distribution, such as that 68% of the sample lies between 1 standard deviation above and below the mean (Aron & Aron, 1999), can thus not be made on the basis of the means and standard deviations provided in the study.

This raises the question of whether WCST performance is not normally distributed in the population or whether the research design resulted in a sample that does not display the characteristic of normality in WCST performance.

4.4.2.2 Standardisation sample. The sampling approach most often used in normative studies is stratified random sampling (Phipps, 1997). This could, however, not be achieved in this study, and snowball sampling was consequently used.

Snowball sampling is considered appropriate to developmental studies (Magnania, Sabinb, Saidela & Heckathorn, 2005; Whitley, 2002). However, it presents a limitation in that there is not any way of knowing whether the sample is representative of the population under investigation (Black, 1999).

In addition, it may result in sampling bias as it accesses participants who are more willing to participate in research studies and participants with larger social networks, than could be achieved by random sampling (Magnania et al., 2005). As a result, it can generate a skewed sample and hampered generalisability of the study.

Nevertheless, according to Atkinson and Flint (2001), the representativeness of snowball samples increase as the sample size increases. The large sample size of this study can thus be considered to have increased the representativeness of the sample. Atkinson and Flint further recommend replication of studies to confirm the data that is obtained from snowball samples. For this research, the sample will thus be considered representative, but recommendations will be made to replicate the study in confirming the findings.

4.4.2.3 Home language versus language of education. One explanation often proposed in explaining cultural differences in test performance, is the impact of language on test performance. In this investigation participants were assessed in Setswana, their home language, it is thus proposed that home language did not influence performance. The persons tested do, however, receive their tertiary education in English – this may have resulted in them feeling more comfortable completing assessments in English. In this respect, assessing university students in their home language may leave them at a disadvantage when completing the WCST.

When consulting the literature, the investigation into conducting assessments in a participant's home language versus his or her language of education, has not yet been concluded. Certain studies report similar performance between groups assessed in their home language and groups assessed in the language in which they are educated (Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman & Radloff, 2004), whilst other researchers found that persons tested were at a disadvantage when assessed in their home language rather than their language of education (Bethlehem, de Picciotto & Watt, 2003). Even as other researchers such as Gasquoinea, Croyle, Cavazos-Gonzalez and Sandoval (2007) report that visual-perceptual tests such as the WCST are not significantly impacted by the language in which administration is conducted, the debate regarding the most suitable language to use for assessment when the participant's home language and language of education differs, is still an on-going issue and requires further research.

4.4.2.4 Discrepancies between the US and Setswana means. In conducting the hypothesis tests between the Setswana and US means, it was found that many of the means differ significantly from each other and that the two data sets do not follow the same distribution. Various explanations may however be presented for the difference between the US and Setswana means found by the hypothesis tests (cf. par. 4.2.3).

One explanation that may be presented for the discrepancy between the means, is that the test items, or the test as a whole, are biased against the Setswana university population. This explanation is feasible as participants were assessed in their home language and were still at a disadvantage. This hypothesis is, however, unlikely as the test items are non-verbal in nature, making use of designs that are not bound to a specific culture.

Another possibility is that the construct of executive functions as measured by the WCST differs between the US population and the Setswana population investigated. This question can, however, not be answered by this research. The possibility thus remains that the construct differs between populations. These differences may be ascribed to cultural differences in test-taking behaviour. The population under investigation may not place as much value on attaining a correct answer and determining the underlying principle on the basis of the responses. Furthermore, they may not function in a context that places value on the abstract logical reasoning required for successful completion of the WCST. If this is the case, in addition to re-standardising the test for separate populations, extensive research into the construct of executive functions and cultural differences in test taking-behaviour is required.

A further hypothesis stems from the fact that executive functions and WCST performance is strongly influenced by the quality of education received (cf. par. 2.5.5.4.3). In this respect, the discrepancy between the mean may not allude to the WCST being biased, as is expected. The WCST may merely accurately measure the inequalities in education as a function of past inequalities in South Africa.

Irrespective of the reason for the difference, when attempting to diagnose difficulties, using the norms derived from Heaton et al. (1993), difficulties in executive functions may be over diagnosed in the Setswana population. The norms provided by this study may thus provide a fairer comparison for executive functioning within the Setswana speaking university population.

4.4.3 Recommendations. As evident from the discussion above on the findings of the research (cf. par. 4.4.1) and the issues taken into consideration when interpreting these findings (cf. par. 4.4.2) various questions regarding the measurement of executive functions and the use of the WCST in South Africa on a Setswana speaking population still remain unanswered. These questions each require further investigation within the South African context.

Firstly, further research is required into the impact of assessment in home language versus language of education on WCST performance of the Setswana university population. This requires comparative research aimed at answering the question of whether Setswana university students perform better when assessed in their home language or their language of education. Replicating this study may achieve this, with the addition of a comparative element comparing the scores of Setswana speaking university students assessed in Setswana versus those assessed in English.

Secondly, research investigating the Setswana-speaking university population's performance on other measures of executive functions may allow consideration of the construct of executive functions and how it differs between the US and Setswana population. This will, however, necessitate investigations on the differences in test-taking behaviour between the US and Setswana population to ascertain whether the construct differs and whether the difference can be ascribed to testing behaviour.

Moreover, it is recommended that other neuropsychological instruments be preliminary standardised for this population to ascertain whether the results obtained in this study are specific to the WCST or is also present when investigating the performance of the Setswana-speaking university population on other neuropsychological tests.

All of the recommendations made above are, however, still grounded in the South African context and history. Therefore, it is proposed that the fields of assessment in South Africa may benefit greatly from further research investigating the long-term impact of past inequalities in education and resources on the South African population's test performance.

Then, a need exists for normative data for other neuropsychological instruments. It is thus also recommended that this study be repeated with all 10 other official South African languages to determine whether the normative information obtained in this study is consistent with the entire South African population or is specific to the Setswana population. This may hopefully lead to more standardised tests suitable for use in the South African context.

The final recommendations stemming from this research is focussed at investigating the impact of possible limiting factors within the research. These require investigating the impact of the sampling approach and the absence of a normal distribution. One recommendation entails replicating this study and making use of stratified random sampling to ascertain the impact the sampling approach had on the data. Replication is also required in order to determine whether the absence of normally distributed data is a function of the research design or whether WCST performance is, in fact, irregularly distributed within the Setswana population. Replication of this study can also confirm the findings of the regression analysis, which indicated that only age impacts one of the WCST variables and that gender and level of education does not impact WCST performance at all. Furthermore, replication of this research may confirm or disconfirm the preliminary norms that have been presented in this study.

Extensive investigation into alternative means of investigating the reliability of the WCST is also required, as the Cronbach Alpha coefficient can only provide an indication of internal consistency, and other measures are not suitable due to the unique nature of the WCST.

For the most part, is it assumed from this study that enough evidence exists to motivate for a full standardisation of the WCST in South Africa in order to allow the accurate measurement of executive functions in the South Africa population. The most prominent recommendation stemming from this research is thus the construction of a full standardisation of the WCST for Setswana speaking university students, and ideally for the whole South African population.

4.5 Conclusion

The chapter started with a comparison of the age-grouped Setswana means with the US means documented by Heaton et al. (1993), and indicated significant differences between certain variable means and similarity between the means for other variables. It was thus concluded that the distribution for the data obtained in this study for the Setswana speaking university students did not follow the same distribution, as the US normative data currently in use and the WCST need to be re-standardised for the Setswana-speaking university population. This study thus aimed to construct a preliminary standardisation of the WCST for the Setswana-speaking university population of South Africa. Investigations on the distribution of the data found that the normative data was not normally distributed with the exception of the variable *Non-perseverative errors*, which approximates a normal distribution. This was followed by investigations into the impact of age, gender and level of education on WCST performance. These investigations found that the WCST does not require separate norms for different age groups, gender groups or level of education groups (with the exception of the variable *Trials to complete first category*). Finally, preliminary norms and psychometric properties were presented for the use of the WCST with the Setswana-speaking university population. The findings, limitations and recommendations stemming from this research was then discussed. Of this, the foremost recommendation was that a full standardisation of the WCST is required for at least the Setswana-speaking university population of South Africa and ideally for the entire South African population.

Chapter 5 will reflect on the findings presented here within the context of neuropsychological assessment in South Africa and the need for standardised assessment measures.

Chapter 5: Conclusion

The present research aimed to construct a preliminary standardisation of the Wisconsin Card Sorting Test (WCST) for Setswana-speaking University students between the ages of 18 and 29 years that attend the University of Limpopo (Medunsa Campus). This required a review of the literature (Chapter 2) that describes psychological assessment and neuropsychological assessment with specific focus on executive functions and the measurement thereof by the WCST. The discussion further focussed on the concept of standardising assessment measures, which outlined the lack of tests standardised for the South African context. Despite an extensive review of the literature available, no standardisation of the WCST, or any other measure of executive function, was found for the South African population. Chapter 3 followed by outlining the research design planned to preliminary standardise the WCST for the study population, followed by a discussion of how the actual research deviated from the planned design. Essentially, the testing instructions by Heaton, Chelune, Talley, Kay and Curtiss (1993) were translated to Setswana and 93 participants were assessed using a computerised version of the WCST. Chapter 4 then discussed the results of the investigation. Hypothesis tests showed that some sub-scores obtained in this study significantly differed from the U.S. norms and others did not.

The distribution of WCST performance in the study population therefore did not follow the same distribution as the US distribution, and on this basis, a need was identified to standardise the WCST for the study population. The standardisation of the testing instructions and scoring procedures were provided by the translated testing instructions and computerised scoring system. The process of constructing normative data was initiated by conducting a regression analysis, which indicated that not gender, age, or level of education influenced participants' WCST scores for all sub-scores, with the exception of the *Trials to complete first category* score that was influenced by age. On this basis, preliminary normative data was constructed. As a result, one normative table was constructed for the entire 18- to 29-year-old age group for the variables *Number of categories completed, Total number of correct responses, Total number of errors, Perseverative responses, Perseverative errors, Non-perseverative errors, Failure to maintain set, Learning to learn* and *Percent conceptual level responses,* whilst separate norm tables for the 18- to 19-year-old and 20- to 29-year-old age groups were respectively constructed for the *Trials to complete first category* sub-score. Finally, preliminary psychometric properties were presented for the WCST in the study population. As a measure of reliability, the Cronbach Alpha Coefficient was calculated. In this regard, the WCST displayed adequate internal consistency in the study population. Having now completed the study, it seems appropriate to reflect on the status of neuropsychological assessment in South Africa and in particular the issue of standardisation of neuropsychological tests.

This research project found a discrepancy in the performance between a sub-section of the South African population – the Setswana-speaking university population—and the US normative population. Such a finding is in keeping with other findings that test performance frequently differs between different racial, cultural and/or ethnic groups. Among these reports are Baker, Denburg, Fonseca and De Mattos (2010), Makhele (2005), Razani, Murcia, Tabares and Wong (2006), as well as Roselli and Ardila (2003). This finding that a South African population performs differently to the US population is no less expected, especially when considering the extremely varied cultural, racial and/or ethnic groups that comprise South Africa's population. It is for this reason that many authors are unanimous in stating that tests need to be standardised for the population in which they are used, which in the case of this research, is the South African context (Foxcroft & Roodt, 2005; Foxcroft, Paterson, Le Roux & Herbst, 2004; Groth-Marnat, 2009; Lezak, Howieson, & Loring, 2004; Zillmer, Spiers, & Culbertson, 2004). In other words, there appears to be an agreement that it is not appropriate to use norms constructed for one population on a different population and, therefore, tests need to be standardised for the population within which it is used. Very few authors, however, indicate how this should be done. Therefore, by carefully considering this issue, the first aspect that requires attention has to do with how tests should be standardised for the South African context and who should be represented in the normative data.

5.1 Constructing Normative Data for the South African Context

As discussed in Chapter 2, norms allow for comparison of a single score to all the scores of the population for which a test has been standardised (Lezak et al., 2004) (cf. par. 2.5.4) Norms can thus be seen as the yardstick to which individuals' scores are compared in order to comment on their behavioural functioning relative to their peers. Currently, the few tests that have been normed in South Africa have only done so by taking a limited portion of the South African population into consideration (Foxcroft & Roodt, 2005), that is, only certain South Africans, mostly white Afrikaans- and English-speaking persons, have been included in normative studies. This raises many questions on the accuracy of assessments conducted in South Africa with a majority population of people who are not white, and do not speak English or Afrikaans as their home language. As all members of the South African population are not represented in the normative sample, valid comparisons cannot be made between the entire South African population and a normative standard consisting of only a certain sector of the population. The field of neuropsychology is thus currently in need of normative data that allows assessment of the entire South African population.

Recently, the issue of considering race when constructing normative data has come to the forefront of neuropsychology (Van de Vijver & Rothmann, 2004). This stems from studies that have found discrepancies in the performance of different racial and cultural groups on neuropsychological tests (Baker et al., 2010; Makhele, 2005; Razani et al., 2006; Roselli & Ardila, 2003).

When considering whether to include race in constructing normative standards, there are currently two options at the disposal of test developers. The one option is to develop normative data specific to each racial, cultural or ethnic group (namely, race-based normative data) and the other to construct an overall normative standard for the entire South African population (namely, national normative data). Each of these possibilities will now be discussed starting with the consideration of race-based normative data.

5.1.1 Race-based normative data. As previously stated, various studies have found significant differences between the performances of different racial or cultural groups on neuropsychological tests. These discrepancies are often used to motivate the construction of normative data sets specific to each racial group.

The reasoning behind this is that race and ethnicity are correlated or associated with factors such as levels of education, acculturation, literacy, and test-wiseness, among others, that may affect brain functioning and neuropsychological test performance (Brickman, Cabo & Manly, 2006). This means that racial groups are said to have differing degrees of access to education and healthcare, or have differing levels of literacy or test-wiseness. These factors may, in turn, influence how a neuropsychological construct develops and how a person performs on a neuropsychological test. In this manner, a racial group may be seen as being correlated or related to factors that may influence neuropsychological test performance and as such it is proposed that racial, cultural or ethnic group may impact on test performance.

According to Brickman, Cabo and Manly (2006), race and ethnicity contain a consortium of factors which may be associated with brain functioning and neuropsychological test performance. These associated factors may include education, acculturation, literacy, test-wiseness and other racial socialisation factors (Brickman et al., 2006). In this manner, racial groups may be seen as being correlated or related to factors that may influence neuropsychological test performance.

Brickman et al. (2006) provide the following example: “Factors such as socioeconomic status may interact with race to influence brain development or functioning because those of lower socioeconomic class may have poorer nutrition and access to health care” (p. 93). As a result, these factors may influence the construct of executive functions (or any other neuropsychological construct) as measured by a neuropsychological test. Taking into consideration that race and its correlates may account for significant discrepancies on neuropsychological test performance; the construction of race-based norms is likely to be associated with particular outcomes or consequences.

5.1.1.1 Consequences of race-based norms. According to Brickman et al. (2006), the accuracy of neuropsychological diagnosis is greatest when the person assessed is demographically similar to the persons on whom the test was standardised for. “The use of race-based norms may increase the accuracy of diagnosis of an individual patient by accounting for factors […] which impact test performance” (Brickman et al., 2006, p. 94). In consequence, race-based norms purport to provide a comparative tool that takes into consideration all the factors or variables associated with race that may have an effect on test performance. The underlying assumption being, that if a variable associated with race impacts the test performance of one member of the race, it affects all members of that race. The impact of these variables would therefore be controlled in the race-based normative data as these variables, although not individually named, are all included in the normative sample. Race-specific norms thus allow comparison to a normative standard that comprises elements that are as similar as possible to the person being tested; thus enhancing the specificity of neuropsychological assessments. Enhanced specificity in assessment, however, has certain implications for the generalisability of assessment.

5.1.1.1.1 Generalisability of race-based norms. The greatest consideration in evaluating the suitability of constructing race-based norms relates to the conclusions and generalisations that may be made on the basis of assessment results that are obtained using race-based norms. Findings can only be generalised to the population from which the sample was constructed (Whitley, 2002). When a person is compared to a normative sample consisting of his cultural peers, generalisations about his behaviour can only be made, regarding his functioning within that specific cultural context – inferences cannot be made on his functioning within the greater South Africa society. Thus, limiting the extent to which findings can be generalised.

If, on the basis of cultural norms, we can only make inferences on a person's functioning within their cultural context, how can the results of persons from different cultures be compared? If a test score provides information on a Sesotho-speaking person relative to Sesotho-speaking peers and on a Setswana-speaking person's performance relative to Setswana-speaking peers, how can Sesotho and Setswana speaking persons be compared when done for, for example, selection purposes?

In summary, race-based normative data may enhance the specificity of measurement but hampers the generalisability of results if considered in standardisations. Even if the specificity of measurement is considered desirable, however, there are certain challenges that emerge in this respect.

5.1.1.2 Challenges in constructing race-based norms. Central to deliberating whether race should be considered in standardisations of neuropsychological tests, and the challenges associated with this, is defining race or ethnicity (Brickman et al., 2006). Most agree that the definition of race and ethnicity is socially and politically determined (Brickman et al., 2006). Pedrasa and Mungas (2008) define race as "socially-constructed labels that serve as proxies for between-group variation in socio-economic status, education, language, acculturation, health care access, and geographic ancestry, among other factors" (p. 185). These may manifest as labels such as African, Caucasian, Asian, or to be more specific, Zulu, Sesotho, Venda, Afrikaans and English, among others.

These labels leave the impression that race, ethnicity or culture is a construct with discrete categories. However, from the definition provided above, it is clear that race, ethnicity and culture are actually abstract constructs, consisting of a multitude of socio-political factors that cannot be discretely defined – this results in a difficulty in the construction of normative data in relation to race. Gasquoine (1999) is in support of the above, maintaining that no clear criterion is available by which to separate cultural and ethnic groups. From this, a point of deliberation begs the question: Who can be classified as belonging to which race? This becomes especially significant when considering multi-racial or multi-cultural families as well as acculturation between groups, in addition to each discreetly identified racial label.

As a result of the difficulty associated with defining discreet racial groups, various questions remain unanswered when considering the construction of race-based norms. For example, when considering the Setswana-speaking population, do Setswana norms provide a fair yardstick or are separate norms needed for urban and rural Setswana-speakers? Are separate norms required for northern as well as southern Setswana-speakers, respectively? Which norm sets are then to be used when assessing a child with one Setswana-speaking parent and one Zulu-speaking parent?

From the above discussion and example, it is evident that conducting studies to answer these questions and then to standardise neuropsychological tests for each possible racial, cultural or ethnic group while ensuring adequate sample sizes would be virtually impossible, impractical as well as cost and resource intensive. Should resources allow for such extensive investigation, further difficulties may become apparent.

Even if constructing race-based normative data for each possible classification of race was possible, a high degree of heterogeneity within racial groups remains (Brickman et al., 2006). This is demonstrated by Touradji, Manly, Jacobs and Stern (2001), who found within-group differences in the neuropsychological test performance of non-Hispanic, white, US elderly persons. These authors demonstrate that "[although] non-Hispanic Whites are often treated as a homogenous group, performance differences exist even within this group" (p. 643). It can thus not be assumed that constructing racial norms provide an overall yardstick for the entire racial group, as even within the racial group significant differences may be present between groups within the same racial group.

Persons within the same race, ethnic or cultural group may differ according to level of acculturation; the degree to which they associate with the race group; their place of residence (urban or rural area); their geographic locality; among numerous other factors such as employment, level of education, marital status, home language, and language of education, to name a few. This demonstrates that every person being tested stems from a different context comprising a unique combination of factors, which would in turn have an influence on their neuropsychological test performance. To construct normative data that is as similar as possible to each person being tested appears to be an unattainable utopian ideal.

Gasquoine (1999) confirms the above and proposes that it is not feasible to use race, culture or ethnicity as an independent variable in research as this may encourage speculations around cultural, racial or ethnic superiority. Gasquoine recommends focusing on measureable psychological variables that may vary between cultural groups and impact on neuropsychological test performance, as these constructs can be satisfactorily defined. These variables may include factors such as age, socio-economic status, level of education, and level of language proficiency, among others. These variables are seen as measurable in contrast to racial, cultural or ethnic groups, which cannot be measured and divided into discrete racial, cultural or ethnic groups.

Despite the utopian nature of constructing race-based norms, these norms allow comparison to a normative standard that is as similar as possible to the person being tested, enhancing the specificity of assessment. However, this limits the extent to which information may be generalised, which in turn limits the conclusions that may be drawn on the basis of comparison to these norms. Race-based normative data may only be generalised to the specific population for which it was generated, and conclusions can only be drawn in this regard. An alternative option is to construct national normative data.

5.1.2 National normative data. The construction of national normative data entails the construction of one normative data set for the entire South African population that is stratified to include all relevant variables impacting on test performance within this context. Some theorists reason that, as the construct of race is politically and socially determined and has no biological impact on brain functioning, it should not be included in normative studies (Brickman et al., 2006). From this perspective race is seen as a mere surrogate for the impact of socio-economic status, access to health care and education and the impact thereof on test performance (Brickman et al., 2006). Hence, motivating for the construction of national-based norms stratifying for the inclusion of the variables for which race is a surrogate. More particularly, by means of random stratified sampling, accounting for all relevant stratification variables, norms representative of the entire South-African population may be constructed. Thus, as stated above, these stratification variables should be measurable variables such as socio-economic status, level of language proficiency and level of education among others, rather than racial, cultural or ethnic labels that group these factors together. Constructing such national normative data will, however, have certain consequences.

5.1.2.1 Consequences of national norms. As previously stated, Gasquoine (1999) proposes that psychological research should focus on measureable psychological constructs, such as socio-economic status and level of education, rather than on the racial labels used to group these constructs together. The construction of national normative standards do not depend as heavily on defining indefinable racial constructs, but can rather focus on including measurable constructs as stratification variables. Therefore, there is less vagueness in defining the constructs investigated and more focus on measurable, definable constructs.

In addition, national normative data is constructed on a much wider sample of persons. In principle, all possible persons are included in the normative sample allowing the assessment and comparison of a much wider range of person. This in turn provides normative data that can be used much more extensively than data constructed for specific racial, cultural or ethnic groups. Furthermore, as representatives of all persons are included in the normative sample, all people can be compared irrespective of racial label.

5.1.2.1.1 Generalisability of national norms. The most prominent advantage of normative data constructed for the entire South African population is that it allows comparison of a person's performance to the greater South African society. On the basis of test scores, inferences may thus be made on the person's level of functioning within the greater South African context rather than just his or her current context, such as with race-based normative data. Thus, using a national-normative sample increases the generalisability of resulting scores from a neuropsychological test.

If the aim of the assessment is then to determine whether the person will function effectively in their current context (despite the lack of access to education and healthcare), normative data constructed for their specific racial group – that is race-based norms – will provide an adequate comparative tool. If the purpose of the assessment is, however, to determine whether the person will function effectively in the greater South African context, national norms will provide a more useful comparative standard. Despite this, various challenges are also inherent in the construction of national normative data.

5.1.2.2 Challenges in constructing national norms. National normative data requires researchers to gain access to a vast array of people that would represent the dynamic characteristics of the South African population, which is resource intensive and demands much logistical planning.

Furthermore, researchers are tasked with determining what the appropriate variables are for constructing national-normative data and defining each variable in a measurable manner. A significant amount of research is thus required prior to the construction of national normative data.

In reviewing the possible consequences of race-based and national norms, it would appear that both strategies offer some advantageous and some challenging consequences. Race-based normative data allows comparison to a normative data that is as similar as possible to the person being tested, but is limited in the extent to which assessment findings may be generalised. Further, defining the construct of race and deciding who to include in which racial norm group presents a challenge. In contrast, national normative data allows the generalisation of findings to the greater South African context, but is limited in the extent to which persons being tested are dissimilar to the normative sample. National norms, however allows the investigation of measureable constructs rather than vague racial categories, but requires extensive research in determining which measurable variables to include in the norms. As both normative strategies are cost and resource intensive, care needs to be taken in deciding which strategy is most appropriate to the South Africa context.

5.1.3 Deciding between race-based and national-based normative data. In deciding the most appropriate normative strategy in standardising neuropsychological tests, the similarities and discrepancies between the strategies come to the fore. Similarly to race-based norms, the process of generating national-norms is time-consuming as well as labour and resource intensive, making the process of standardising tests relevant to the South African context a challenge for researchers as well as test developers. Taking cognisance of this, great care needs to be taken in order to apply these resources in the most effective manner when constructing normative data. From the above discussion, it is evident that the normative strategy to be employed depends on the purpose of the assessment and on the conclusions one wishes to draw on the basis of assessment results. Should clinicians wish to draw conclusions on an individual's functioning within a specific context, race-based norms will be the most appropriate – however, should one wish to draw conclusions, relevant to a wider query, national-norms would suffice better. While race-based norms have the advantage of specificity, national norms have the advantage of generalisability.

As clinicians, considerable thought thus needs to be put into the process of determining an acceptable balance between generalisability and specificity – deciding on the most effective manner in which to approach the field of neuropsychology and in determining the most appropriate normative strategy.

Having considered the issue around constructing normative data and whom to include in these normative standards, closer attention will now be paid to the normative data currently available in the South African context.

5.2 Evaluating the Normative Data Currently Available for the South African Context

The second challenge facing neuropsychologists in South Africa is the evaluation of the normative data that is currently available. Currently, most normative data constructed for neuropsychological tests are constructed by means of small-scale research studies. Most of these studies are limited by time and resource restraints. As a result, most focus on a specific racial, cultural or ethnic group. Examples of these are Anderson (2000), Burns (1996), Endres (1996), Makhele (2005) and Phipps (1997). Much of the normative data currently available is thus specific to a certain racial, cultural or ethnic group and, as discussed in depth above, cannot be generalised to the greater South African population as it subject to the limitations of race-based normative data.

In addition, to the limitations presented by race-based normative data, much of the normative data available is limited by small sample sizes. Considering the lack of resources available, a question that arises is: Should small sample sizes be accepted as best practice considering the limited resources at our disposal, or should normative data constructed on small samples be discarded? Are norms constructed on small South African samples more appropriate than international norms constructed on bigger sample sizes?

A difficulty that should be considered regarding smaller sample sizes is that data constructed from a few participants may not be representative of the functioning of the group to which the results are expected to be generalised (Whitley, 2002). When using norms as a comparison tool, a sample of adequate size randomly selected from the population and stratified according to the relevant demographic variables are required to assume that the normative sample provides a fair sample of the behaviour under investigation (Orsini, Van Gorp & Boone, 1988).

When using normative data constructed from small samples, more bias may result in the South African population than when using norms constructed in a different population. An example of this is the normative data established by Anderson (2000) for Zulu-speaking factory workers. When considering factory workers with no formal education, the normative data is based on a sample size of only three persons (Anderson, 2000).

As a demonstration of the impact this may have, the following example considering the Grooved Peg Board performance of the dominant hand of a Zulu-speaking factory worker with no formal education: Anderson (2000) reports a mean score of 101 and a standard deviation of 29,31. Whilst the US norms provide a mean of 94 with a high cut-off score of 46 and a low score cut off score of 139 (Lafayette Instrument, 2002). If a Zulu speaker obtains a score of 129, he or she will thus be labelled as impaired in comparison to the Zulu data as constructed in the study by Anderson, but within normal standards according to the US normative data. The question therefore arises: Which data would allow a fairer comparison of a Zulu-speaking individual in the South African context?

The example above illustrates that merely labelling normative data as standardised in South African is insufficient. The adequacy of the norms need to be evaluated and, it is not inconceivable, in some instances that non-South-African normative data may be more suitable than South African normative data constructed on inadequate sample sizes.

In addition to inadequate sample sizes, the accuracy of assessments conducted in the South African context depends on the construct equivalence of tests in South Africa.

5.3 Evaluating the Construct Equivalence of the Tests Available for Use in South Africa

Construct equivalence is concerned with whether members from different cultures attach the same meaning to a certain construct (Welkenhuysen-Gybels & Van de Vijver, 2001). This entails investigating whether psychological constructs function in a similar or different manner when investigated in different contexts. Specific to the South African context, it entails answering the question: When using this test in South Africa, does it measure the same construct as it was originally developed for?

Taking into consideration the extreme degrees of diversity in the South African context, it becomes especially significant to investigate whether the constructs differ between each of these contexts, in addition to it differing in the South African context as a whole. This entails considering what skills are required to function effectively in certain contexts and what the impact of functioning in these different contexts have on the development of a construct.

In demonstrating this, the concept of executive functions will be evaluated to determine whether the definition may differ between contexts, commenting on the implications this has for the measurement of executive functions by the WCST. In Chapter 2, the construct of executive function was defined as the ability to develop and maintain an appropriate problem-solving strategy across changing stimulus conditions in order to achieve a future goal (Heaton et al., 1993) (cf. par. 2.3.1). Commenting specifically on the definition of executive functions as required for the successful completion of the WCST, the construct heavily depends on the process of abstract logical reasoning in addition to problem solving across various contexts. A key consideration in this regard is the various contexts within which the person is required to solve problems.

Illustrating this point, within an urban business world, problem solving across different contexts requires strong abstract logical reasoning. Within a rural context however, a person is able to solve problems across different contexts without the same level of abstract logical reasoning required in an urban context. Thus, within an urban context the WCST may provide an adequate indication of a person's executive functions. Within a rural context however, the WCST may provide an underestimation of the person's executive functions, as a person may function and solve problems appropriately within a rural context without the same levels of abstract logical reasoning required in an urban area. In consequence, it is evident that the resultant scores obtained on the WCST may have different meanings in different contexts.

When using the WCST within the South African context clinicians thus need to use their clinical judgment in deciding whether the person being assessed functions in a context where abstract logical reasoning in addition to problem solving is required to function effectively in their context. In this case, the WCST may provide an appropriate measure of executive functions, such as in an urban area. When abstract logical reasoning is not required to the same extent, such as in a rural context, the construction of alternate measures of executive functioning that depend less on abstract logical reasoning and is more consistent with real world executive functions. In so doing, mimicking the daily tasks requiring executive functions within the rural South African context.

As evident from this discussion, the construct of executive functions may differ between different contexts, and the construct is not equivalent between contexts. This does, however, not imply that the test should be discarded within the South African context because it is low in construct equivalence. It highlights the extent to which psychologists need to be aware of the different South African contexts and consider these when planning an assessment, as the construct equivalence of tests cannot be taken for granted and clinical judgement is required in this regard.

A great amount of discourse and research is thus required around how neuropsychological constructs differ between contexts and which tests are appropriate to which contexts in order to assist clinicians in developing the clinical judgement required to plan appropriate assessment batteries in a variety of contexts.

From the above discussion, a number of prominent challenges have been identified in relation to neuropsychological tests in the South African context, including constructing normative data, evaluating the normative data currently available and evaluating the construct equivalence of tests currently available. Within the context of the challenges outlined, it is clear that neuropsychological assessment is a field where much academic discourse is needed in identifying and pursuing a way forward specifically within the South African context.

5.4 The Way Forward for Neuropsychological Assessment in South Africa

The challenges outlined in this chapter bring to light the extent to which clinicians need to critically evaluate the assessment measures at their disposal. Considering that assessment elicits a sample of behaviour within a certain context, clinicians need to consider which tests are applicable to which contexts and to which context the results of the assessment may be generalised. The context of every assessment scenario is different and will require the use of different tests and the results of the tests will be applicable to different contexts. Clinicians therefore need to expand their knowledge on the tests currently available and critically evaluate these in order to develop clinical judgement that allows for the planning of assessments suitable to a variety of contexts.

Furthermore, research to examine the contexts to which current tests and normative data are applicable as well as the development of new assessment instruments specific to the South African context, needs to be a priority in the area of neuropsychological assessment in order to make fair assessment available to every South African citizen. This research has brought forward one example of this by demonstrating the need to standardise the WCST for the South African population and difficulties encountered when doing so.

Throughout this research, it became evident that in the quest to provide fair assessment for all South Africans, many questions remain unanswered. And despite various attempts and intentions for change, psychological and neuropsychological assessment still faces many challenges in this regard. Currently, South Africa is thus faced with a great challenge when considering the field of psychological assessment. This challenge, however, has the potential to spur test developers to new heights in developing measures allowing fair assessment, potentially contributing to the benefit of all South Africans (Foxcroft & Roodt, 2005).

The onus thus lies with South African psychologists in deciding whether they wish to take up the challenge, engage in research and discourse regarding the issues facing assessment in South Africa and contribute to the development of fair assessment for all South Africans, or whether they will merely continue to echo the call for culture fair assessment.

Appendix A

English Consent Form

UNIVERSITY OF LIMPOPO (Medunsa Campus) CONSENT FORM

Statement concerning participation in a Clinical Research Study

Name of Study:

A Preliminary Standardisation of the Wisconsin Card Sorting Test for Setswana-Speaking University Students

I have heard the aims and objectives of the proposed study and was provided the opportunity to ask questions and given adequate time to rethink the issue. The aim and objectives of the study are sufficiently clear to me. I have not been pressurised to participate in any way. I understand that participation in this clinical study is completely voluntary and that I may withdraw from it at any time and without supplying reasons.

I know that this study has been approved by the Research, Ethics and Publications Committee of the Faculty of Medicine, University of Limpopo (Medunsa Campus). I am fully aware that the results of this study will be used for scientific purposes and may be published. I agree to this, provided my privacy is guaranteed.

I hereby give consent to participate in this study.

____
Name of Recipient

_________________________ ______________________
Place *Date*

____
Witness

Statement by the Researcher

I provided verbal information regarding this study.
I agree to answer any future questions concerning the study as best as I am able.
I will adhere to the approved protocol.

Christi Gadd
Researcher

________________ _______________ ____________________
Signature *Date* *Place*

Setswana Consent Form

UNIVERSITY OF LIMPOPO (Medunsa Campus) CONSENT FORM

Seteitemente se se ka ga go tsaya karolo mo Porojeke ya Patlisiso.

Leina la Patlisiso

A Preliminary Standardisation of the Wisconsin Card Sorting Test for Setswana-Speaking University Students

Ke utlwile maitlhomo le maikemisetso a patlisiso e e tshitshintsweng mme ke filwe tšhono ya go botsa dipotso le go fiwa nako e e lekaneng ya go akanya gape ka ntlha e. Maitlhomo le maikemisetso a patlisiso e a tlhaloganyega sentle. Ga ke a patelediwa ke ope ka tsela epe go tsaya karolo.

Ke tlhaloganya gore go tsaya karolo mo Tekopatlisisong e ke boithaopo le gore nka ikgogela morago mo go yona ka nako nngwe le nngwe kwa ntle ga go neela mabaka. Se ga se kitla se nna le seabe sepe mo kalafong ya me ya go le gale ya bolwetsi jo ke nang le jona e bile ga se kitla se nna le tlhotlheletso epe mo tlhokomelong e ke e amogelang mo ngakeng ya me ya go le gale.

Ke a itse gore Tekopatlisiso e e rebotswe ke Patlisiso le Molao wa Maitsholo tsa Khampase ya Medunsa (MCREC), Yunibesithi ya Limpopo (Khampase ya Medunsa). Ke itse ka botlalo gore dipholo tsa Tekelelo di tla dirisetswa mabaka a saentifiki e bile di ka nna tsa phasaladiwa. Ke dumelana le seno, fa fela go netefadiwa gore se e tla nna khupamarama.

Fano ke neela tumelelo ya go tsaya karolo mo Patlisiso.

____

Leina ka molwetse/moithaopi Tshaeno ya molwetse kgotsa motlamedi.

_________________________ ______________________

Lefelo *Letlha*

____

Paki

Seteitemente ka Mmatlisisi

Ke tlametse tshedimosetso ka molomo malebana le Patlisiso.
Ke dumela go araba dipotso dingwe le dingwe mo nakong e e tlang tse di amanang le Patlisiso ka moo nka kgonang ka teng.
Ke tla tshegetsa porotokolo e e rebotsweng.

Christi Gadd

________________ ________________ ______________ ________________

Lefelo *Leina la Mmatlisisi* *Tshaeno* *Letlha*

Appendix B
Screening Questionnaire

I hereby make myself available for 20 min to participate in the study entitled:

A Preliminary Standardisation of the Wisconsin Card Sorting Test for Setswana-Speaking University Students

Please mark the appropriate response:

Are you a student at the University of Limpopo?	YES	NO
Are you fluent in Setswana?	YES	NO
Have you ever completed the Wisconsin Card Sorting Test before?	YES	NO
Have you ever received psychiatric treatment?	YES	NO
Have you ever been diagnosed with a neurological difficulty?	YES	NO
Can you use a computer and mouse?	YES	NO
What is you gender?	MALE	FEMALE
What is your date of birth?		
How old are you?		

What are you studying?	
Which year of study are you currently completing?	

Name: ____

Cell number: ____

Email: ____

Signature: ____

The purpose of entering you details here is purely for contact purposes. All your information will be kept confidential.

Kind regards,
Christi Gadd
Researcher, MSc Clinical Psychology II

Appendix C

Computerised-WCST Testing Instructions

This test is a little unusual because I am not allowed to tell you very much about how to do it.

When the test starts you will see the screen shown below.

Please sort out the deck of cards by moving each card to the placeholder below the key card you think it matches.

Once you have moved a card you cannot move it again. After each move you will receive feedback if the sort was correct or not.

You have to use the feedback to work out the correct way of sorting the cards. Please try to sort all the cards correctly.

Good luck!

Translated Setswana WCST Instructions

Teko eno efapane go se nene, gonne ga ke yatshwanela go bolella gore o e dire jang.

Fa teko e semologa o tla bona ka setshwantsho se tlhagalelang ko tlase.

Ka kopo baya dikarata ka manane, ka go sutisa karata ngwe le ngwe mo tulong e o gopolang gore ee diretswe kwa tlase.

Fa o sutisitse karata eo gangwe ga wa tshwanelwa go e sutisa gape. Morago ga motsamao mongwe le mongwe o amogela molaetsa o go kaelang fa motsamao o siame kgotsa o sa siama. O tshwanelwa ke go dirisa molaetsa oo go bona fa tiro ya gago ele e siameng, go baya dikarata ka manane. Ka kop leka go baa dikarata tseo ka manane.

Mathlhokgonolo!

Back Translated Instructions

This test is a bit tricky, because I'm not allowed to say how it should be done.

When the test begins, you will see by the picture that appears below.

Please place the cards in order, by moving each one of them to a position that you think is suitable, below.

Once you have moved the card, you are not allowed to move it again. After every move, accept the message that informs you whether the move was correct or not. You must use the message in order to check if you have placed the cards in the correct order. Please try to put the cards in the correct order.

Good luck!

Appendix D

Histograms Depicting Distribution of the WCST Data Sets for Chapter 4

This section will display the histograms depicting the distribution of the data as discussed in Chapter 4. Each histogram, with the exception of *Learning to learn,* was constructed on the basis of the entire data set (N = 93). Each figure will be presented and thereafter briefly discussed.

Figure 1 Histogram depicting the distribution of the *Number of categories completed* scores for the entire data set

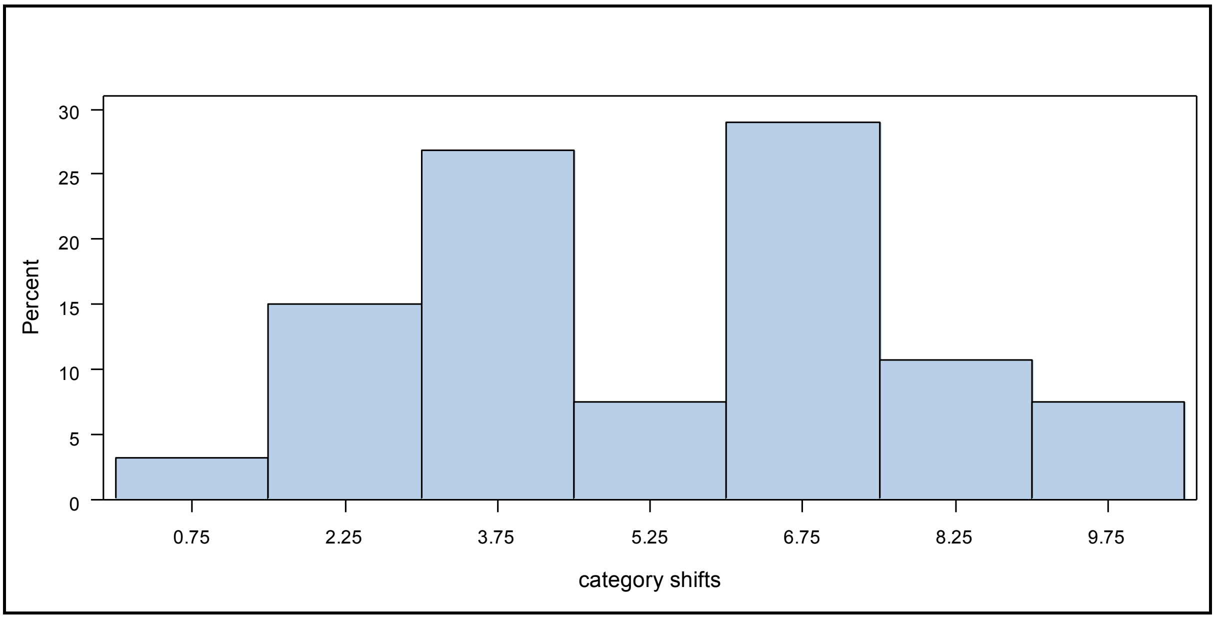

Figure 1 shows a histogram depicting the distribution of the *Number of categories completed* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 2 Histogram depicting the distribution of the *Total number of correct responses* scores for the entire data set

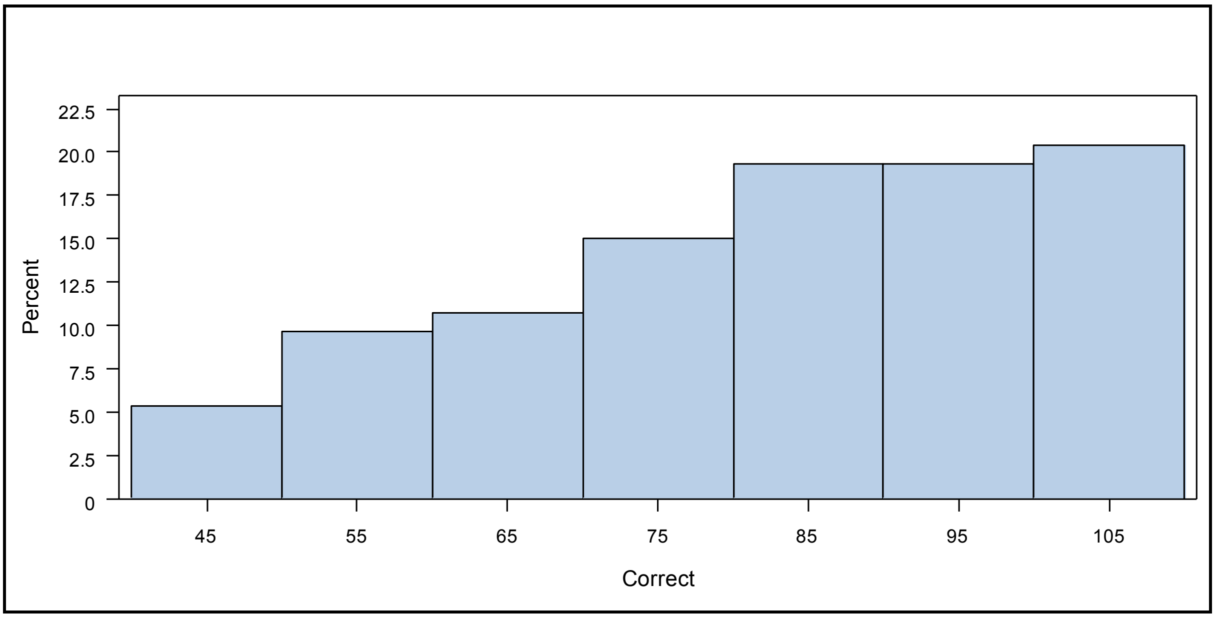

Figure 2 shows a histogram depicting the distribution of the *Total number of correct responses* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 3 Histogram depicting the distribution of the *Total number of errors scores* for the entire data set

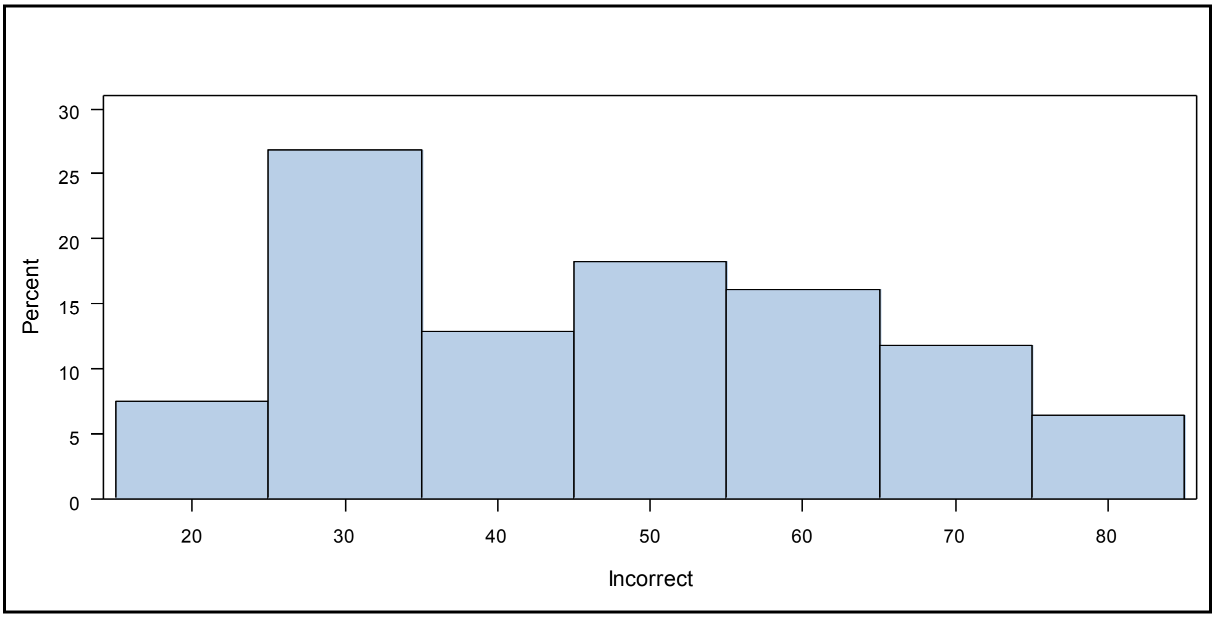

Figure 3 shows a histogram depicting the distribution of the *Total number of errors* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 4 Histogram depicting the distribution of the *Perseverative responses* scores for the entire data set

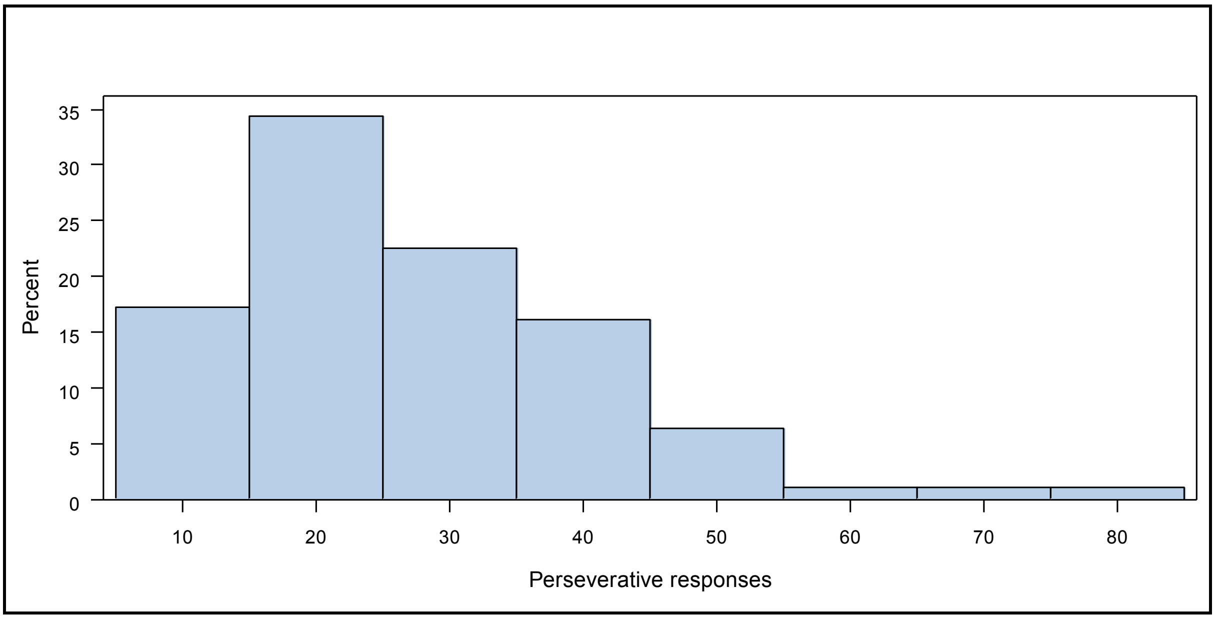

Figure 4 shows a histogram depicting the distribution of the *Perseverative responses* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 5 Histogram depicting the distribution of the *Perseverative errors* scores for the entire data set

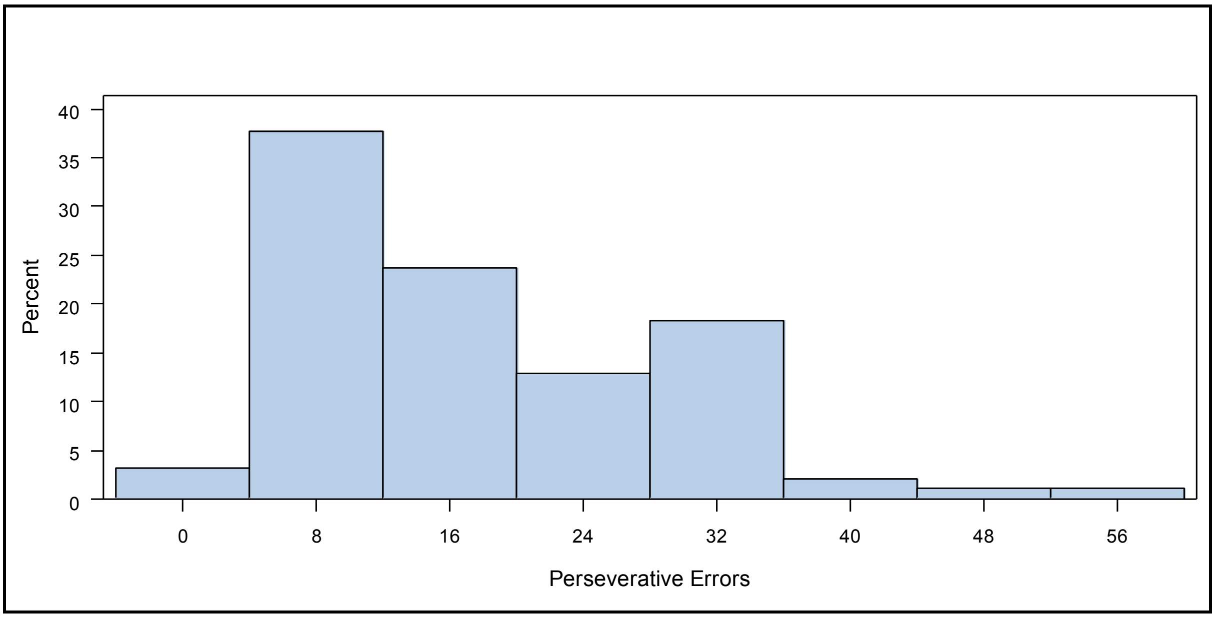

Figure 5 shows a histogram depicting the distribution of the *Perseverative errors* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 6 Histogram depicting the distribution of the *Non-perseverative errors* scores for the entire data set

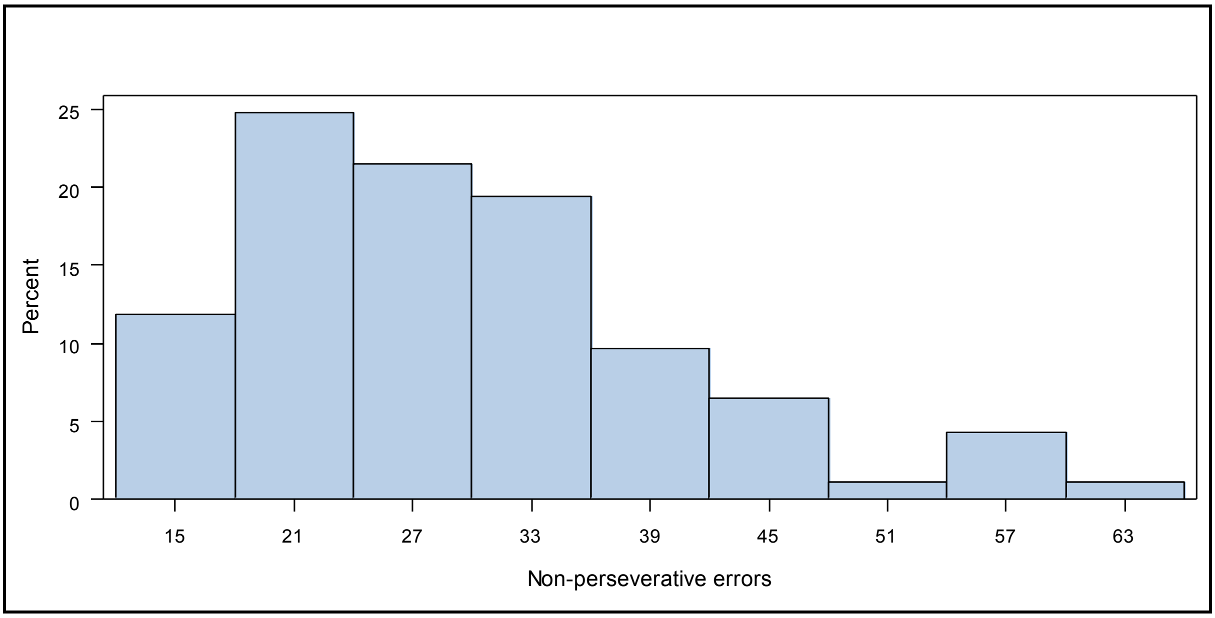

Figure 6 shows a histogram depicting the distribution of the *Non-perseverative errors* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 7 Histogram depicting the distribution of the *Trials to complete first category* scores for the entire data set

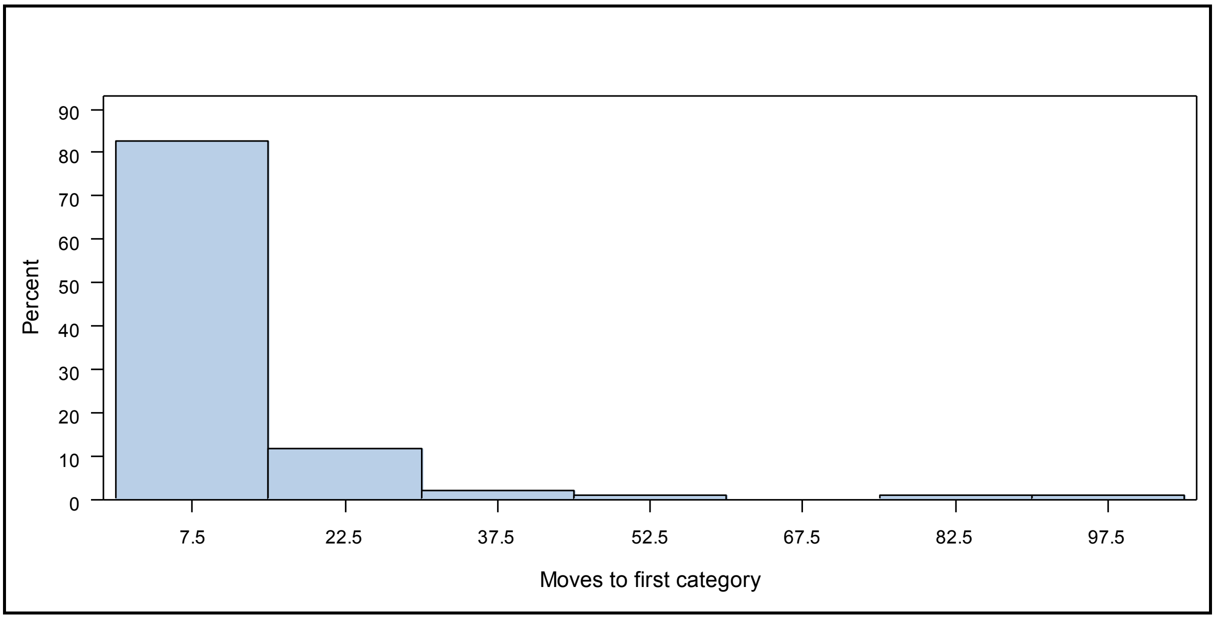

Figure 7 shows a histogram depicting the distribution of the *Trials to complete first category* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 8 Histogram depicting the distribution of the *Failure to maintain set* scores for the entire data set

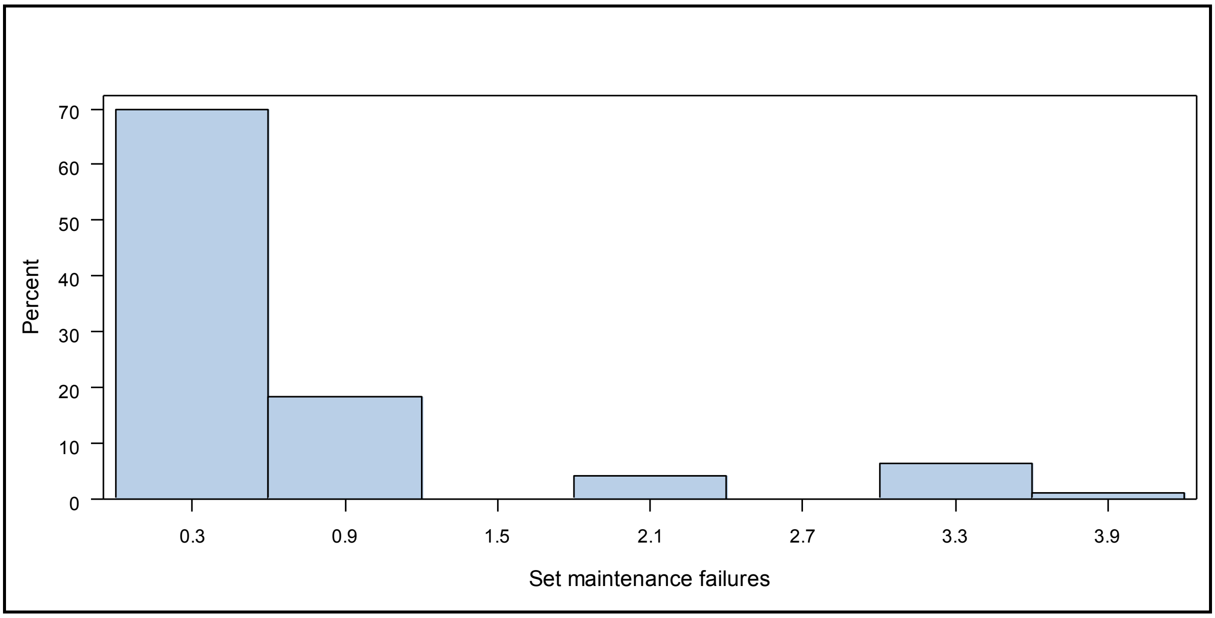

Figure 8 shows a histogram depicting the distribution of the *Failure to maintain set* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

Figure 9 Histogram depicting the distribution of the *Learning to learn* scores for the entire data set

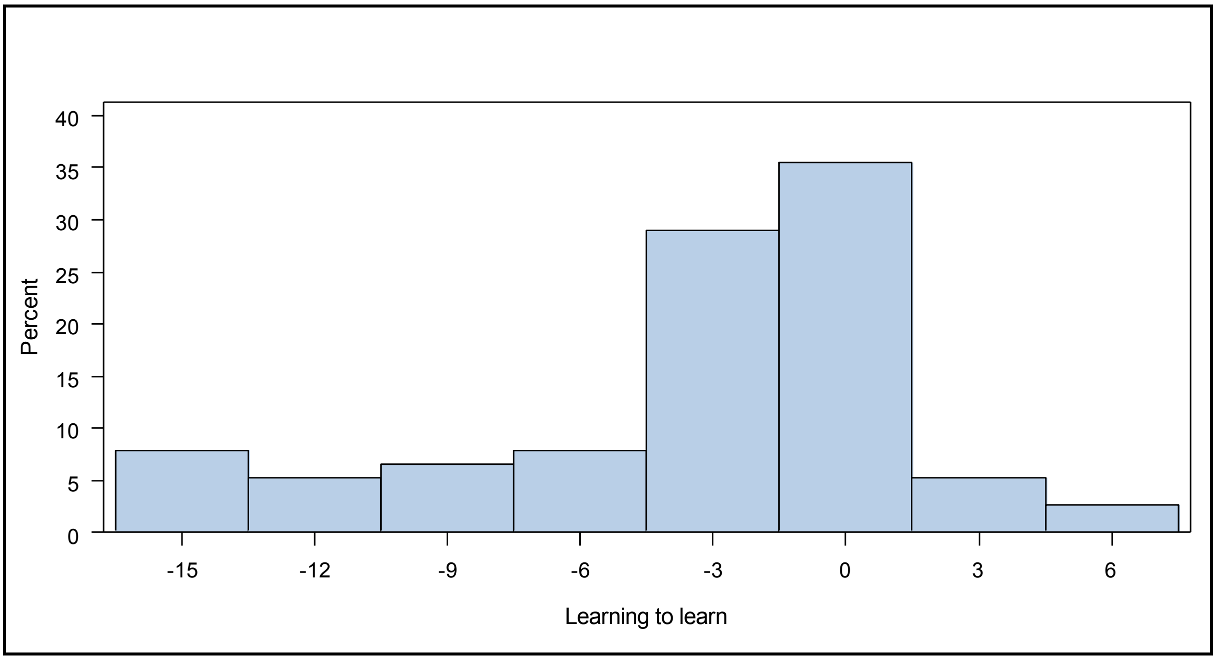

Figure 9 shows a histogram depicting the distribution of the *Learning to learn* variable for the entire data set. From the histogram it is evident that the data is not normally distributed. The histogram is based on the 76 particpants for whom a *Learning to learn* score could be calculated.

Figure 10 Histogram depicting the distribution of the *Percent conceptual level responses* scores for the entire data set

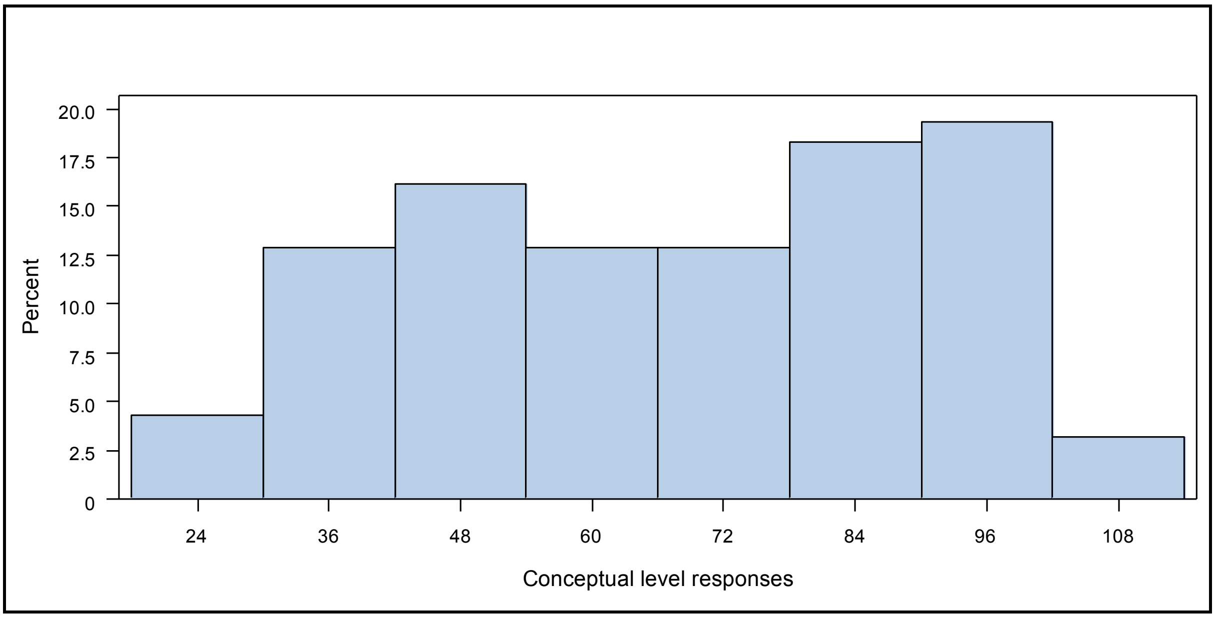

Figure 10 shows a histogram depicting the distribution of the *Percent conceptual level responses* variable for the entire data set. From the histogram it is evident that the data is not normally distributed.

From the presented histograms (Figure 1 to 10), it is concluded that none of the WCST variables displays a normal distribution for the entire data set.

Appendix E

Expanded Preliminary Normative Data

Table 9(a) will now present the preliminary normative data for the entire data set including the median, mode, range and variance in addition to the mean and standard deviation presented in Chapter 4. Tables 9(b) and 9(c) will then present the expanded preliminary normative data for the separate age groups for the variable *Trials to complete first category*.

Table 9(a) Expanded preliminary normative data for Setswana-speaking university students between the ages of 18 and 29 years

N=93	Mean	Median	Mode	Range	Variance	Standard Deviation
Number of categories completed	5.16	5.00	4.00	9.00	5.57	2.36
Total number of correct responses	81.67	83.00	83.00	64.00	318.40	17.84
Total number of errors	46.34	45.00	45.00	64.00	317.84	17.83
Perseverative responses	27.04	23.00	11.00	71.00	175.46	13.26
Perseverative errors	17.18	14.00	7.00	56.00	120.48	10.98
Non-perseverative errors	29.15	28.00	22.00	47.00	114.26	47.00
Failure to maintain set	0.51	0.00	0.00	4.00	0.86	0.92
Learning to learn[1]	-3.48	-1.82	0.00	20.54	24.31	4.93
Percent conceptual level responses	67.03	68.00	39.00	85.00	533.62	23.10

Table 9(a) shows that the overall data set (N = 93) displayed, for example, a mean of 5.16, a median of 5.00 and a mode of 4.00 for the variable *Number of categories completed*. The mean is greater than the median and mode, indicating that, as previously discussed; the data is not normally distributed, but rather positively skewed. It can thus be suspected that more participants did relatively well than relatively poor. Furthermore, the variable *Number of categories completed* showed a range of 9.00, variance of 5.57 and a standard deviation of

[1] Note the all calculations for the *Learning to learn* score are based on the 76 subjects for whom a learning score could be calculated.

2.36. This indicates that the data is narrowly distributed with relatively low discriminatory value.

As a further example, Table 9(a) shows that the variable *Total number of correct responses* displayed a mean of 81.67, a median of 83.00 and a mode of 83.00. In this case the mean is lower than the median and mode, indicating that, as previously discussed the data is not normally distributed, but rather negatively skewed. It can thus be suspected that relatively more participants will do poorly than well on this sub-score. Furthermore, the variable *Total number of correct responses* shows a range of 64.00, variance of 318.40 and a standard deviation of 17.84 and the variable. This indicates that the data is widely distributed and a more sensitive indicator with higher discriminatory value.

Table 9(b) Expanded preliminary normative data for Setswana-speaking university students between the ages of 18 and 19 years for the variable *Trials to complete first category*

N=31	Mean	Median	Mode	Range	Variance	Standard Deviation
Trials to complete first category	12.00	11.00	11.00	9.00	5.67	2.38

Table 9(b) shows that the 18- to 19-year-old age group (N = 31) displayed a mean of 12.00, a median of 11.00 and a mode of 11.00 for the variable *Trials to complete first category*. The mean is greater than the median and mode, indicating that, as previously discussed; the data is not normally distributed, but rather positively skewed. It can thus be suspected that more participants did relatively well than relatively poor on this variable. Furthermore, the variable *Trials to complete first category* shows a range of 9.00, variance of 5.67 and a standard deviation of 2.38. This indicates that the data is narrowly distributed with relatively low discriminatory value.

Table 9(c) Expanded preliminary normative data for Setswana-speaking university students between the ages of 20 and 29 years for the variable *Trials to complete first category*

N=62	Mean	Median	Mode	Range	Variance	Standard Deviation
Trials to complete first category	16.34	11.00	11.00	92.00	237.28	15.40

Table 9(c) shows that the 20- to 29-year-old age group (N = 62) displayed a mean of 16.34, a median of 11.00 and a mode of 11.00 for the variable *Trials to complete first category*. The mean is greater than the median and mode, indicating that, as previously discussed; the data is not normally distributed, but rather positively skewed. It can thus be suspected that more participants did relatively well than relatively poor. Furthermore, the variable *Trials to complete first category* shows a range of 92.00, variance of 237.28 and a standard deviation of 15.40. This indicates that the data is widely distributed with relatively higher discriminatory value.

References

Anderson, S. J. (2000). *Neuropsychological assessment procedures: A practical guide.* Pietermaritzburg: University of Natal, School of Psychology.

Anderson, S. J. (2001). On the importance of collecting local neuropsychological normative data. *SAJP,* 31(3).

Anderson, S. W., Demasio, H., Jones, R. D. & Tranel, D. (1991). Wisconsin Card Sorting Test as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology*, 16(9): 902–922.

Aron, A. & Aron, E. N. (1999). *Statistics for Psychology* (2nd edition). New Jersey: Prentice-Hall.

Atkinson, R. & Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update*, (33): 1–4.

Babbie, E. (2005). *The basics of social research*. Stamford, CT: Thomson.

Babbie, E., Mouton, J., Payze, C., Vorster, J., Boshoff, N. & Prozesky, H. (2001). *The practice of social research* (South African edition). Cape Town: Oxford University Press South Africa.

Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice Effects on The Wisconsin Card Sorting Test – 64 Card Version Across Twelve Months. *The Clinical Neuropsychologist, 15*, 471-478.

Baker, D. S., Denburg, N., Fonseca, R. P. & De Mattos, M. A. (2010). A cultural study on decision making: Performance differences on the Iowa gambling task between selected groups of Brazilians and Americans. *Psychology & Neuroscience*, 3(1): 101–107.

Beers, M. H. (Ed.). (2003). *The Merck manual of medical information: Home edition* (2nd edition). New York: Pocket Books, Simon & Schuster.

Belli, G. (2009). Nonexperimental quantitative research. In:. Lapan, S. D & Quartaroli, M. T. (Eds), *Research essentials: An introduction to designs and practices*. San Fransisco: Wiley & Sons, pp. 59–77.

Bethlehem, D., De Picciotto, J. & Watt, N. (2003). Assessment of verbal fluency in bilingual Zulu-English speakers. *South African Journal of Psychology*, 33(4): 18–28.

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics*. Thousand Oaks, CA: Sage.

Brickman, A. M., Cabo, R. & Manly, J. J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology,* 13(2): 91–100.

Brown, J. D. (2002). The Cronbach alpha reliability estimate. *JALT,* 6(1): 17–18.

Burns, C. N. (1996). A preliminary standardizationof the WISC-R for eight year old Tswana-speaking children: Subtests: Vocabulary and Block Design*.* Unpublished PhD thesis. Pretoria: Medical University of Southern Africa (Medunsa).

Cairns, N. J. (2004). Neuroanatomy and Neuropathology. In: Goldstein, L. H. & McNeil, J. E. (Eds), *Clinical Neuropsychology: A practical guide to assessment and management for clinicians*. West Sussex, UK: Wiley & Sons, pp. 23–55.

Capitani, E. (1997). Normative data and neuropsychological assessment: Common problems in clinical practice and research. *Neuropsychological rehabilitation, 7*(4): 295–309.

Cinan, S. & Tanor, O. O. (2002). An attempt to discriminate different types of executive functions in the Wisconsin Card Sorting Test. *Memory,* 10(4): 277–289.

Claassen, N., Krynauw, A. H., Paterson, H. & Mathe, M. (2001). *A standardization of the WAIS-III for English-speaking South Africans.* Pretoria: Human Science Research Council (HSRC).

Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In: Goldstein, L. H. & McNeil, J. E. (Eds), *Clinical Psychology: A practical guide to* *assessment and management for clinicians*. West Sussex, UK: Wiley & Sons, pp. 121–140.

Delport, C. (2005). Quantitative data-collection methods. In: De Vos, A. S., Strydom, H., Fouche, C. B & Delport, C. (Eds), *Research at grass roots: For the social sciences and human service professions* (3rd edition). Pretoria: Van Schaik, pp. 159–191.

Denckla, M. B. (2007). Executive function: Binding together the definitions of Attention-Deficit/Hyperactivity Disorder and learning disabilites. In: Meltzer, L. (Ed), *Executive functions in education: From theory to practice*. New York: Guilford Press, pp. 5–18.

Dornyei, Z. & Taguchi, T. (2010). *Questionnaires in second language; Construction, administration, and processing*. New York: Taylor & Francis.

Draper, N. R. & Smith, H. (1981). *Applied regression analysis* (2nd edition). New York: Wiley.

Durrheim, K. (1999a). Research Design. In: Terre Blanche, M. & Durrheim, K. (Eds), *Research in practice: Applied methods for the social sciences*. Cape Town: University of Cape Town Press, pp. 29–53.

Durrheim, K. (1999b). Quantitative measurement. In: Terre Blanche, M. & Durrheim, K. (Eds), *Research in practice: Applied methods for the social sciences*. Cape Town: University of Cape Town Press, pp. 72–95.

Durrheim, K. (1999c). Quantitative analysis. In Terre Blanche, M. & Durrheim, K. (Eds), *Research in practice: Applied methods for the social sciences*. Cape Town: University of Cape Town Press, pp. 96–122.

Endres, G. N. (1996). A preliminary standardization of the WISC-R for eight year old black children: Using the Picture Completion and Coding sub-tests. Unpublished dissertation. Ga-Rankuwa: University of Limpopo (Medunsa Campus).

Fallgatter, A. J. & Strik, W. K. (1998). Frontal brain activation during the Wisconsin Card Sorting Test assessed with two-channel near-infrared spectroscopy. *European archives of psychiatry and clinical neuroscience,* 248:245– 249.

Foxcroft, C. & Roodt, G. (2005). *An introduction to psychological assessment in the South African context* (2nd edition). Cape Town: Oxford University Press.

Foxcroft, C., Paterson, H., Le Roux, N. & Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis. The test use patterns and needs of psychological assessment practitioners.* Pretoria: Human Science Research Council (HSRC).

Gasquoine, P. G. (1999). Variables moderating cultural and ethnic differences in neuropsychological assessment: the case of Hispanic Americans. *Clinical Neuropsychology*, 13(3).

Gasquoine, P. G., Croyle, K. L., Cavazos-Gonzalez, C. & Sandoval , O. (2007). Language of administration and neuropsychological test performance in neurologically intact Hispanic American bilingual adults. *Archives of Clinical Neuropsychology*, 22(8): 991–1001.

Gatewood, R. D., Feild, H. S. & Barrick, M. (2008). *Human Resource Selection* (6th edition). Mason, OH: Thomson.

Goldstein, F. C. & Green, R. C. (1995). Assessment of problem solving and executive functions. In: Mapou, R. L. & Spector, J. (Eds), *Clinical neuropsychological assessment: A cognitive approach (Critical issues in neuropsychology)*. New York: Plenum Press, pp. 49–79.

Goldstein, L. H. & McNeil, J. E. (2004). General introduction: What is the relevance of neuropsychology for clinical psychology practice? In: Goldstein, L. H. & McNeil, J. E. (Eds), *Clinical neuropsychology: A practical guide to assessment and management for clinicians*. West Sussex, UK: Wiley & Sons, pp. 3–20.

Grieve, K. (2001). *Guidelines for basic neuropsychological assessment procedures: A manual for professional students at the University of South Africa.* Unpublished manual, Pretoria.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th edition). New Jersey: Wiley & Sons.

Haynes, M. A., Lee, A. B. & Drew, C. R. (1995). Medunsa and the training of black doctors for South Africa. *Academic Medicine,* 70(2): 115–121.

Haynes, S. N., Richard, D. C. & Kubany, E. S. (2005). Content validity in psychological assessment:A functional appriach to concepts and methods. *Psychological assessment,* 7(3): 238–247.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G. & Curtiss, G. (1993). *Wisonsin Card Sorting Test manual: Revised and Expanded.* Odessa, FL: Psychological Assessment Resources.

Horton, A. M. & Soper, H. (2008). Children's Memory. In: Reed, J. & Warner-Rogers, J. (Eds), *Child neuropsychology: Concepts, theory and practice*. Oxford: Wiley & Sons, p. 228.

Hughes, C. & Graham, A. (2008). Executive functions and development. In: Reed, J. & Warner-Rogers, J. (Eds), *Child neuropsychology: Concepts, theory, and practice*. Oxford: Wiley & Sons, pp. 272–273.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher,* 30(2): 3–13.

Kolb, B. & Wishaw, I. Q. (2008). *Fundamentals of human neuropsychology* (6th edition). New York: Worth.

Lafayette Instrument. (2002). *Grooved Pegboard test user instructions.* Lafayette, IN: Lafayette Instrument Company.

Leedy, P. D. & Ormrod, J. E. (2005). *Practical research: Planning and Design* (8th edition). Upper Saddle River, NJ: Prentice Hall.

Lezak, M. D., Howieson, D. B. & Loring, D. W. (2004). *Neuropsychological assessment* (4th edition). New York: Oxford University Press.

Loeber, S., Duka, T., Welzel, H., Nakovics, H., Heinz, A., Flor, H., et al. (2009). Impairment of cognitive abilites and decision making after chronic use of alcohol: The impact of multiple detoxifications. *Alcohol & Alcoholism*, 44(4): 372–381.

Louw, D. A., Van Ede, D. M. & Louw, A. E. (2004). *Human development* (2nd edition). Cape Town: Kagiso Tertiary.

Magnania, R., Sabinb, K., Saidela, T. & Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS*, 19(2): 67–72.

Makhele, L. (2005). *Utility of Koppitz norms for the Bender-Gestalt test performance of a group of Sesotho-speaking children.* Unpublished dissertation. Bloemfontein: University of the Free State.

Manly, J. J. (2008). Critical issues in cultural neuropsychology: profit from diversity. *Neuropsychology Review,* 18(3): 179–183.

Maree, K. (2010). Editorial : assessment in psychology in the 21st century - a multi-layered endeavour. *SAJP*, 40(3). 229 - 233

Maree, K. & Pietersen, J. (2007a). The quantitative research process. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 144–153.

Maree, K. & Pietersen, J. (2007b). Sampling. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 171–181.

Medunsa. (n.d.). History of Medunsa. Available at http://www.medunsa.ac.za/faculties/nsph/nsph_about/about_medunsa1.htm (accessed on 13 March 2011).

Meltzer, L. (2007). Executive function: Theoretical and conceptual frameworks. In: Meltzer, L. (Ed.), *Executive function in education: From theory to practice*. New York: Guilford Press, pp. 1–3.

Mitrushina, M., Boone, K. B., Razini, J. & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd edition). Oxford: Oxford University Press.

Montgomery, D. C., Runger, G. C. & Hubele, N. F. (2004). *Engineering Statistics* (3rd edition). New York: Wiley.

Moran, S. & Gardner, H. (2007). Hill, skill, and will: Executive function from a multiple-intelligence perspective. In: Meltzer, L. (Ed.), *Executive function in education: From theory to practice*. New York: Guilford Press, pp. 19–38.

Nieuwenhuis, J. (2007). Qualitative research designs and data gathering techniques. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 69–97.

Onwuegbuzie, A. J. & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2): 281–316.

Onwuegbuzie, A. J., Jiao, Q. G. & Bostick, S. L. (2004). *Library anxiety: Theory, research March and applications.* Lanham, MD: Scarecrow Press.

Ormond-Brown, D. (2010). Computerized version of the WCST. Personal communication, April 2010.

Ormond-Brown, D. (n.d.). *Diagnostic Software* (Ormond Software Enterprises). Available at http://www.ormond.co.za/html/diagnosis.html (Accessed on 21 March 2010).

Orsini, D. L., Van Gorp, W. G. & Boone, K. B. (1988). *The neuropsychology casebook.* New York: Springer-Verlag.

Pedraza, O. & Mungas, D. (2008). Measurement in cross-cultural neuropsychology. *Neuropsychol Rev,* 18(3): 184–193.

Pepping, M. (2003). *Successful private practice in neuropsychology: A scientist-practitioner model.* San Diego: Academic Press.

Phipps, W. D. (1997). A preliminary standardization of the WISC-R for eight-year old Tswana-speaking children:The Digit Span and Mazes subtests. Unpublished dissertation. Pretoria: Medical University of Southern Africa (Medunsa).

Pietersen, J. & Maree , K. (2007a). Statistical analysis I: Descriptive statistics. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 182–196.

Pietersen, J. & Maree, K. (2007b). Statistical analysis II: Inferential statistics. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 197–213.

Pietersen, J. & Maree, K. (2007c). Overview of statistical techniques. In: Maree, K. (Ed.), *First steps in research*. Pretoria: Van Schaik, pp. 224–254.

Preston, J. D., O'Neal, J. H. & Talaga, M. C. (2009). *Consumer's guide to psychiatric drugs.* New York: Pocket Book/Simon & Schuster.

Professional Board for Psychology. (2004). *Rules of conduct pertaining specifically to psychology.* Pretoria: HPCSA.

Purdon, S. E. & Waldie, B. (2001). A short form of the Wisconsin Card Sorting Test. *Journal of Psychiatry & Neuroscience,* 26(3): 253–256.

Razani, J., Murcia, G., Tabares, J. & Wong, J. (2006). The effects of culture on WASI test performance in ethnically diverse individuals. *The Clinical Neuropsychologist*, 21(5): 776–788.

Roselli, M. & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, 52(3): 326–333.

Salkind, N. J., Dougherty, D. M. & Frey, B. (Eds). (2010). *Encyclopedia of research design.* Thousand Oaks, CA: Sage.

Sbordone, R. J. & Saul, R. E. (2000). *Neuropsychology for health care professionals and attorneys* (2nd edition). Boca Raton: CRC Press.

Sherman, E., Strauss, E. & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd edition). New York: Oxford University Press.

Shuttleworth-Edwards, A., Kemp, R. D., Rust, A. L., Muirhead, J., Hartman, N. P. & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III performance. *Journal of Clinical and Experimental Neuropsychology*, 26: 903–920.

Siddiqui, S. V., Chatterjee, U., Kumar, D., Siddiqui, A. & Goyal, N. (2008). Neuropsychology of the prefrontal cortex. *Indian journal of psychiatry,* 50(3): 202–208.

Sinclair, M. & Taylor, E. (2008). Attention development. In: Reed, J. & Warner-Rogers, J. (Eds), *Child neuropsychology: Concepts, theory and practice*. Oxford: Wiley & Sons, pp. 250–251.

Smith-Seemiller, L., Franzen, M. D. & Bowers, D. (1997). Use of Wisconsin Card Sorting Test short forms in clinical samples. *The Clinical Neuropsychologist,* 11(4): 421–427.

Tighe, J., McManus, I. C., Dewhurst, N. G., Chis, L. & Mucklow, J. (2010). The standard error of measurement is a more appropriate measure of quality for postgraduate assessments than is reliability: An analysis of MRCP(UK) examinations. *BMC Medical Education,* 10(40): 1–9.

Touradji, P., Manly, J. J., Jacobs, D. M. & Stern, Y. (2001). Neuropsychological test peroformance: A study of non-hispanic white elderly. *Journal of Clinical and Experimental Neuropsychology,* 23(5): 643–649.

Tredoux, C. & Smith, M. (1999). Jumping to conclusions: An overview of inferential statistical methods. In: Terre Blanche, M. & Durrheim, K. (Eds), *Research in practice:*

Applied methods for the social sciences. Cape Town: University of Cape Town Press, pp. 331–354.

Trochim, W. (2006). Types of reliability (Research methods knowledge base). Available at http://www.socialresearchmethods.net/kb/reltypes.php (accessed on 18 May 2011).

Tshabalala-Msimang, M. (2004). Speech by the Minister of Health. 23rd Graduation Ceremony – Medical University of South Africa. Ga-Rankuwa: Department of Health, pp. 1–4.

University of Limpopo. (n.d.). University of Limpopo (About Us). Available at http://www.ul.ac.za/index.php?Entity=About%20Us (accessed on 13 March 2011).

Van de Vijver, A. & Rothmann, S. (2004). Assessment in multicultural groups: The South African case. *SA Journal of Industrial Psychology,* 30(4): 1–7.

Van Eeden, R. & Mantsha, T. R. (2007). Theoretical and methodological considerations in the translation of the 16PF into an African language. *SAJP,* 37(1): 62–81.

Van Vuuren, D. & Maree, A. (1999). Survey methods in market and media research. In: Terre Blanche, M.& Durrheim, K. (Eds), *Research in practice: Applied methods for the social sciences*. Cape Town: University of Cape Town Press, pp. 269–286.

Walliman, N. (2005). *Your research project: A step-by-step guide for the first time researcher* (2nd edition). London: Sage.

Welkenhuysen-Gybels, J. & Van de Vijver, F. (2001). A comparison of methods for the evaluation of construct equivalence in a multigroup setting. Proceedings of the Annual Meeting of the American Statistical Association. Belgium: Catholic University of Leuven, pp. 1–6.

Whitley, B. E. (2002). *Principles of research in behavioural science* (2nd edition). New York: McGraw Hill.

Williams, T. A., Sweeney, D. J. & Anderson, D. R. (2009). *Contemporary business statistics: with custom selections* (3rd edition). South-Western Territory, Australia: Cengage Learning.

Zillmer, E. A., Spiers, M. V. & Culbertson, W. C. (2004). *Principles of neuropsychology* (2nd edition). Belmont, CA: Thomson Wadsworth.