AN EXPLORATORY STUDY OF THE EFFECTIVENESS OF THE CRITICAL REASONING TEST BATTERY AS A DEVELOPMENTAL SELECTION TOOL: A CASE OF A RETAIL COMPANY

By Michelle Hendricks

Mini-thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF COMMERCE

in the

DEPARTMENT OF INDUSTRIAL PSYCHOLOGY

at the

UNIVERSITY OF THE WESTERN CAPE

SUPERVISOR: PROFESSOR ABRAHAMS

NOVEMBER 2022

ABSTRACT

Companies around the world strive to ensure that the best is employed to perform the jobs and tasks needed by them and therefore many companies use psychometric tests as a selection tool to aid their appointment decisions. Although the psychometric test is scientific and provides insights about the strengths and weaknesses of individuals, it is important that the tests that are used are valid, reliable, and fair to be worthwhile for the organisation, and are perceived as such by the individuals that are taking the test.

This research examines the test results of respondents that completed both the verbal and numerical aptitude tests as part of a selection test battery that is used at a retail company based in the Western Cape. The study harnessed a quantitative research approach that used the secondary data of the company. Various statistical analyses, which included a t-Test, Analysis of variance (ANOVA), and correlations, were conducted. The data sample (n=202) of the raw scores of the Critical Reasoning Test Battery was analysed to establish whether any disparities existed across the scores achieved by different race groups, age groups, and gender groups.

Significant differences were found amongst the different race groups and age groups. The results are discussed in relation to their limitations as well as their implications in terms of both the practices and future research.

KEYWORDS

Critical reasoning; verbal reasoning; numeric reasoning; selection; employment equity; retail; correlations; race groups, gender groups, age groups

DECLARATION

I hereby declare that the research **"AN EXPLORATORY STUDY OF THE EFFECTIVENESS OF THE CRITICAL REASONING TEST BATTERY AS A DEVELOPMENTAL SELECTION TOOL: A CASE OF A RETAIL COMPANY"** is my own work, that it has not been submitted for any degree or examination at any other institution of higher learning, and that all the sources I have used or quoted have been indicated and acknowledged by complete references. It is being submitted for the degree of Masters of Commerce at the University of the Western Cape.

Name: Michelle Hendricks

Date: November 2022

Signature: ________________

ACKNOWLEDGEMENTS

"…And let us not grow weary of doing good, for in due season we will reap, if we do not give up" (Galatians 6 vs. 9). All things were made through him, and without him was not anything made that was made (John1 vs. 3).
These acknowledgements serve to express my gratitude to the people and entities that played a role in enabling me to complete this mini-thesis:

Firstly I need to give all honour and praise to God as He has been by my side throughout and blessed me with the strength and wisdom to achieve this milestone.
There are too many to mention by name, but to all my family, friends and acquaintances for their support and encouragement throughout my studies, sincerely thank you especially:

To my partner, Michael Hendricks, thank you for patience and supporting me in pursuing my career and study goals. I could not have done it without you and I will always be grateful to you for this. Also, thank you for all the cups of coffee to keep me awake during the completion of my thesis and studies and the day-to-day cups.☺.

To my daughter, Monique Hendricks, thank you for being understanding and patient when things were not always easy. Your support and love carried me through and you inspire me daily when I look at how much you have grown into the beautiful person you have become.

To my parents, Willie and Catherine Caesar, thank you for your love, support, encouragement and interest throughout my studies. You instilled hard work in me and the value of education to realise my full potential. The generational curse is broken. Thank You!

To my employer, thank you for granting me permission to conduct my research within the organisation and using their data to complete my thesis.

To Dr Mahembe, thank you so much for your help with everything. I really appreciate it.

And, lastly to my supervisor Fatima Abrahams, thank you for your guidance, insight and support throughout my thesis. I am grateful for all that I learnt from you and it was a privilege and a pleasure to work with you on my research, THANK YOU!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Human resource professionals make selection and recruitment decisions by considering the competencies that individuals hold as well as determining which of these competencies are required or need further development for success in an organisation. Organisations can make more informed and effective decisions to hire and develop employees, by motivating talent and addressing change by analysinga an individual's current capabilities and comparing them to the skills needed to perform (Roman & Mason, 2015).

To remain competitive in South Africa, organisations must develop and use workplace-related, culturally appropriate, open, and pragmatic selection and assessment methods (Foxcroft & Roodt, 2018). Particularly, South African labour law advocates for the responsible and fair use of psychometric assessments (Employment Equity Act, 1998). Furthermore, various tools, instruments, or test measures are available to assess human behavior. Sources of information such as interviews and collateral sources are tools to make sound and conclusive decisions. The modern era of testing has also seen changes in the assessment process due to the growth of computer-based examinations via the internet, the increasing impact of technological advances on examinations, and the proliferation of competency-based assessments. Since South African society is multicultural and multilingual, an African-centred approach should be adopted. The need to use measures that are both reliable and valid is the foundation of the African-centred approach. The technical test manual of the instrument usually provides this evidence (Foxcroft & Roodt, 2018).

In South Africa, the Employment Equity Act (No. 55 of 1998) stipulates that validation is required. Psychological testing and other similar evaluations of employees are prohibited by this Act unless it is scientifically proven that the test or evaluations used are valid, reliable, and can be applied fairly to all employees, without bias against any employee or group.

An international study conducted by te Nijenhuis, Al-Shahomee, van den Hoek, Allik, Grigoriev and Dragt (2015), concluded that mean differences exist and that a significant correlation of "IQ" with many educational, economic, and social criteria is well established.

Therefore, cognitive testing is widely used in the selection and placement of educational institutions and organisations. This has also led to a long-standing interest in the average difference in “IQ” test scores across different population groups.

According to Donald, Thatcher and Milner (2014), psychological assessment instruments can be important in the transformation of an organisation. Specifically, transformation to redress previously disadvantaged groups and to determine who will be employed by the organisation. This includes promotion, transfer, and access decisions for development. Therefore, these tests and other forms of assessments act as gatekeepers and determine both entries into the labour market and career movement.

Due to the apartheid legacy in South Africa, social structures that were based on racial groups, remain. Furthermore, in South Africa, testing is likely to be biased or partial toward certain groups; therefore, the importance of this research study should not be ignored in the new South Africa. Normally, to determine the bias or fairness of a test, it is required to run that test through trial and error (Foxcroft & Roodt, 2018). The implications of this are broad, as test users and publishers may be required to show or prove in court that certain tests do not discriminate against certain groups of people (Abrahams, 1996; Dupper, Bhoola & Garbers, 2009; Foxcroft & Roodt, 2018; Görgens-Ekermans & Herbert, 2013).

Traditional psychometric assessment batteries indicate proficiency or the lack of a particular skill or ability. They use scientific rigor to provide reliable and valid assessments of various psychological states and behaviors. Therefore, psychometric tests are often used as complementary sources of information when making selection decisions to recruit, promote or develop employees (Foxcroft, 2011; Murphy & Maree, 2006; Steyn & De Bruin, 2020).

Good and ethical test development practices dictate that it is important that tests developed in countries other than the ones in which they are intended to be used, be tested for transportability and alignment. This transportability and alignment should be based on a common cultural experience that gives each candidate an equal chance to understand the underlying logic. Furthermore, the assessment must be designed and used for the specific technical area for which it is intended (PsytechSA, 2017).

Contemporary studies have focused research attention on changes from previous practices used primarily in South Africa, which is primarily taking psychological measures from abroad with minimal attention to the ability of these instruments to confer psychometric properties. For these instruments to be reliably applied in local settings, the psychometric properties of the imported measures must be examined and demonstrated to meet the minimum standard requirements for validity and reliability (Arendse, 2021; Görgens-Ekermans & Herbert, 2013).

In South Africa, not enough research has been conducted on the actual psychometric test batteries that were commonly used for selection. Some of the test listed as aptitude tests includes the South African Wechsler Adult Intelligence Scale (SAWAIS), the Intermediate Battery (INT Bat), the Senior Aptitude Test (SAT), and the Raven's Progressive Matrices and these were seldom used for selection in organisations (Foxcroft, 2011; Steyn & De Bruin, 2020).

According to Laher and Foxcroft (2014), an older study that was conducted by Van der Merwe (2002), one of the few authors who has dealt with broader evaluation issues, particularly how evaluations are conducted in South African organisations. This small, qualitative study (although outdated) provides insight into what tests are being used and for what purposes in the industry. Van Der Merwe (2002) conducted exploratory qualitative research on the various assessment instruments that were used by various organisations, across different sectors in the Eastern Cape, and looked at which tests were used in the different industries. The study provides a cumulative report of which psychometric tests were used, and for what purposes these were used to assist human resources management in various organisations. The author's findings include a myriad of different test instruments that were used, with several tests used by only one or two organisations, whereas other tests are used by three or more organisations (Van der Merwe, 2002).

Besides the problem relating to the incompatibility or lack of alignment relating to tests developed in contexts other than the ones in which they are being used, adverse impact is another problem often experienced in the South African context (Mahembe, 2014; Theron, 2009). According to Odendaal (2015), adverse selection can occur when certain selection strategies cause certain group members to be less selected than other groups. It is widely believed that adverse impact has its roots either in the psychological measures used or in the differences in the criterion being measured. According to Theron (2009), understanding these negative consequences leads to a misdiagnosis of the problems leading to differences in criterion performance.

Besides the problem of adverse impact on selection, other tests such as personality tests, are also plagued by a different problem, that occurs during data collection or questionnaire completion. Even though the use of personality tests for selection remains a popular tool, there has been consistent criticism about running the risk of individuals providing socially desirable responses to improve their chances for selection. In South Africa, common standardised tests of cognitive skills are used to assess verbal skills, quantitative skills, and deductive reasoning (Arendse, 2021; Cockcroft, Alloway, Copello & Milligan, 2015; Foxcroft, 2011; Steyn & De Bruin, 2020; Theron, 2009).

According to Roodt and La Grange (2001), measures of cognitive ability do not predict task performance. This finding was based on research conducted using the Verbal Evaluation Test (VCC3) and asserted that it could not be generalised to other industries or roles, but a limitation of this study was that the population consisted of insurance sales people and should be generalised for this group.

Other South African studies have shown that racial and ethnic group differences are partially explained by group differences in cognitive abilities. Limited studies have been conducted in South Africa on this specific topic, with a few research studies by Huysamen (2002); Maola and Bux (2015); Salgado and Anderson (2003), which investigated the relationship between cognitive ability and task performance. This research above shows that general cognitive ability (or general mental ability, "***g***") is the strongest predictor of occupational learning and mastery, as well as overall job performance for nearly all occupations. Further research documents well that the "***g***"-tests shows mean subgroup differences on cognitive ability tests by race and ethnicity, sex, and age, in local and international research. Results concluded that in multicultural societies such as South Africa, the home environment, schooling, language skills, diet, and other factors could influence measures of cognitive ability. Other South African studies have shown that race, education, socioeconomic status, language, and English language comprehension are key factors influencing the comparison of constructs and items on cognitive and personality tests (Odendaal, 2015).

According to De Goede and Theron (2010), tests that report differences in standardised mean scores across racial groups, particularly in cognitive ability, should not be viewed as the villains, reinforcing the effects of socio-political discrimination and adversely affecting personnel selection. The systematic difference in the allocation of group criteria may be due to how the HR function responds to the problem of selection bias. In South Africa, it appears reasonable to assign these differences to socio-political conditions that subdued the increase in skills and knowledge development of particular groups to succeed in business. The new democratic South Africa provided easy access to opportunities, to develop the skills necessary for groups to be successful. The disadvantaged group remains disadvantaged and does not get the opportunity to develop the required coping strategies, knowledge, skills, and abilities, while the advantaged group is given more opportunities to choose from and develop. Thus, in South Africa, there is a need for a way to identify individuals who will benefit most from positive developmental opportunities, particularly cognitively demanding developmental opportunities.

1.2 SIGNIFICANCE OF THE STUDY

Statistically evaluating and understanding whether any bias or disparities exists between different groups is the key focus of this research to gauge the impact on key deliverables of organisations such as employment equity, personnel selection processes, and much more. Employment equity should ensure that competent members from designated groups in South Africa are hired, promoted, and retained equally; instead of having an organisation suffer undue hardship or hiring and promoting unqualified candidates. The onus rest with the organisation to plan and ensure Employment Equity success, and there is no silver bullet or single solution to this challenge (Roman & Mason, 2015).

Transformational goals can conflict with an organisation’s performance goals, creating significant challenges for companies to achieve their employment equity targets and having huge effects on the personnel selection procedures used by the organisation to get this balance correct. Furthermore, the quota system, which is used in South Africa to redress inherent past disparities, and is known as affirmative action, depersonalises people and sees them as numbers (Oosthuizen & Naidoo, 2010; Reuben & Bobat, 2014).

According to Roman and Mason (2015), top management is the second most targeted group of organisations' employment equity focus. Black recruits are officially or casually excluded through propriety networking practices, and a male-dominated organisational culture is maintained. Little real responsibility or decision-making authority remains; therefore, Black people are selected as tokens and not fully integrated into companies.

Part of the process includes developing and testing effective selection procedures that HR is responsible for to ensure that people are selected to perform effectively. Therefore, sound selection procedures ensure that valid inferences occur about future job activities from available measured scores. Any assessment procedure should provide sufficient evidence that it applies to the position or task. This evidence refers to the validity of the assessment and the degree to which the accumulated data and theory support the analyses of the test scores. Therefore, in South Africa, the Employment Equity Act (No. 55 of 1998) requires that validated tests are used as selection tools (Arendse, 2021; Foxcroft, 2011; Oosthuizen & Naidoo, 2010).

1.2.1 PROBLEM STATEMENT

Therefore, because of the above, this study was conducted on the Critical Reasoning Test Battery, 2nd Edition (CRTB2), which was one of the selection tests used, in conjunction with other assessment tools, in an established retail company based in the Western Cape province. It was specifically used to select applicants for placement in management development programs.

Due to the focus on employment equity and transformation in the sector, it is pertinent to explore the relevance of the validity and reliability of the measurement instrument used for selection. Particularly to explore whether any significant differences exist amongst gender groups and race groups and to justify the fairness of the test in terms of the employment equity requirements in South Africa.

Ineffective talent management causes unsystematic training and development of Black staff. It has far-reaching consequences for selection practices and procedures used and consequently should impact the decision of what to look for or test when recruiting for the required target (Roman & Mason, 2015). It has an important link to this research as it focuses on the results of the tests used for recruiting and selecting for the organisation’s management development pool.

The use of cognitive ability tests assists the retail organisation in selecting candidates for management development programs. For the organisation to have confidence in the fairness of their process, as well as to satisfy the fairness requirements of the employment equity legislation, it is pertinent to statistically evaluate the secondary data of the organisation to understand if any significant differences exist amongst gender and race groups concerning their cognitive ability.

Research is a tool to add information based on sound scientific principles to the existing body of knowledge. While there has been a large amount of research conducted on psychometric testing for selection purposes, much of the research done were international studies, with a limited amount of research done locally or in Africa. While conducting this research, there were no other studies that the researcher found locally or internationally that focused on the validity and reliability of the specific Critical Reasoning Test Battery. This research aimed to report on the findings that examined whether significant differences between groups exist for the critical reasoning test battery of a retail organisation in the Western Cape province of South Africa to address the research questions.

1.3 RESEARCH OBJECTIVES AND AIMS

The primary objective of the study was to determine whether there were significant differences in the average scores achieved on the critical reasoning test battery compared to group differences.

Specifically, the aims of this study are to:

1. Determine if there are differences in the average scores across males and females.
2. Determine whether differences exist across racial groups in terms of mean scores.
3. Observe whether or not age moderates the relationship of either verbal or numeric ability.

1.4 HYPOTHESES

The main points of this study were formulated within the research objectives mentioned above. The assumptions are:

H_1: For verbal reasoning, there are significant differences in mean test scores across the different gender groups.

H_2: For numeric reasoning, there are significant differences in mean test scores across gender groups.

H_3: There are significant differences in the mean verbal reasoning test scores across the different racial groups.

H_4: There are significant differences in the mean test scores across racial groups for numeric reasoning.

H_5: A relationship exist accross specific age groups for both verbal and numeric reasoning.

1.5 CONCLUSION

In this section, an introduction to the research was given to aid understanding. It also looked at research conducted on psychometric testing for selection practices internationally and in South Africa. This chapter provided an overview of the problem statement and identified the purpose of the study and the subsequent hypotheses formulated to answer the research questions.

1.6 STRUCTURE OF THE THESIS

Chapter 1 presents the research conducted for this study. The significance of the study, the problems identified, and the research objectives were all considered. Chapter 2 will look at the literature review, which will provide the theoretical background of this study. It will also define the different theoretical constructs, and the overview of the history of intelligence and aptitude testing. Chapter 3 will describe the research methodology. It will detail the sampling method and the statistical properties of the instrument. Chapter 4 will present the results. Tables and figures will be included in this section to indicate the descriptive and inferential statistics conducted for this study. Chapter 5 will discuss the results of this study. It will have recommendations provided and conclusions drawn.

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

This chapter examines the theoretical framework of this study based on the relevant literature. This study investigated the utility of the Critical Reasoning Test Battery (CRTB2) as an instrument of choice to select candidates considered for management development programs at a retail company in South Africa.

The literature review focuses on the history and the theories that discuss and define cognitive ability tests, intelligence testing, and the constructs of reliability and validity to better understand the study and the relevance these have to the selection process. The chapter reviews the South African landscape with specific reference to the legislation that governs fairness with a focus on the labour market to understand the impact of diversity and cultural differences on psychometric testing since this research study is on the diverse South African population. The selection processes used in the industry are also discussed to understand their application in different settings as well as the use of aptitude tests in South Africa.

2.2 HISTORY AND DEFINITIONS OF INTELLIGENCE, APTITUDE AND ABILITITES

It is pertinent to highlight that while literature sometimes uses the terms ability, cognitive function, intelligence, and aptitude interchangeably, to denote the behavior that is used to predict the future learning or performance of an individual, these are different constructs (Cockcroft et al., 2015; Macklem, 1990).

Therefore, before delving into the various definitions and theories, it is important to point out the similarities and nuances between aptitude and intelligence so that it becomes easier to follow and understand why these theories are discussed. According to Owen (1996), the ability measured by the aptitude tests corresponds to the intellectual ability measured by the general aptitude test. The fundamental difference is that specific measures of selected general cognitive functioning are designed primarily as unitary measurements and do not necessarily represent different abilities. In addition, certain functional domains directly relates to the educational and occupational contexts and are not included in general measures of cognitive function (Foxcroft

& Roodt, 2018). Aptitude is the natural ability to do something. However, there has often been confusion about the skill of someone and the intelligence of someone (Goldstein, Allen & De Luca, 2019).

According to Macklem (1990), ability is the skills or the talent to do something or the proficiency in a specific area. These skills are developed over time by exposure to learning or being shown how to do things. Aptitudes, on the other hand, are those skills that the person already has and are innate or they are born with to do a specific kind of work naturally; but can potentially be developed further with more exposure. These aptitudes are either physical or mental, and different aptitudes have been identified and are testable. Therefore, aptitude is more than just the ability, but rather the ability plus the suitability for the performance of a task. Furthermore, the tests designed to measure these different characteristics differ in several important ways (Arendse, 2021; Foxcroft 2011).

Aptitude tests include items that measure specific skills, such as verbal and numerical skills. These skills forecast theoretical performance in instructive programs or occupational success. It contrasts with achievement tests, which measure recent learning and are associated with the specific content or the curriculum. Thus, aptitude tests inform what the individual can do, irrespective of the particular curriculum exposure that the individual had (Macklem, 1990).

According to Baron and Norman (1992), aptitude tests assess the developed abilities and skills of individuals acquired through years of education and experience in oral and numerical material. The test score achieved by the individual informs us what the person is capable of at the time of being tested but does not inform us why individuals perform as they do, as we would require knowledge about each person's experiential background to answer that. Furthermore, how the test scores are used may differ. Generally, traditional achievement tests are designed and used to assess current performance. Traditionally aptitude tests are designed to forecast future performance after the completion of specific learning. It is important to note that all psychometric tests evaluate the current performance, regardless of whether the aim is to evaluate the individual's performance on a specific test or to predict future performance.

According to Nugba and Quansah (2021), aptitude tests measure an individual's present skills and abilities, therefore, assessing their existing abilities at the time of measurement and thus reflecting past learning. These tests assess intellectual skills that are, in most cases, not particularly taught at school (Nugba & Quansah, 2021). According to Katz and Brown (2019), both achievement and aptitude tests assess the current status of abilities, but aptitude tests also focus more on the predictive nature of the data. Furthermore, while intelligence tests measure general intelligence that looks at overall mental capacity performance, aptitude tests are also designed to measure a narrower range of abilities than intelligence tests (Katz & Brown, 2019).

According to Reynolds, Altmann and Allen (2021), aptitude tests are commonly referred to as cognitive ability tests, which test logical reasoning and mental ability. The scope of aptitude tests is broader than that of achievement tests. Achievement tests measure knowledge and skills in specific areas of education, while aptitude tests measure the cognitive abilities, skills, and knowledge gained through life experiences. Achievement tests measure what was accomplished at a particular point, while aptitude tests predict future performance (Reynolds et al., 2021).

Intelligence is defined as the ability to acquire the knowledge and skills to solve and understand complex problems. The person then uses their aptitude or capacity and applies this learned knowledge or intelligence to solve the problem (PsytechSA, 2017).

The history of intelligence and aptitude testing dates back to around the 19th century; therefore, the literature related to this is older. There are limited and, few recent critiques on these theories and the researcher has endeavored to find the most recent literature related to this topic. Sir Francis Galton published his book *Hereditary Genius* in 1896, which described the first known testing as intelligence aptitude tests. These experiments were mainly physical and sensory and differ significantly from modern testing (Goldstein, Allen & De Luca 2019). At the turn of the 20th century, Alfred Binet, a French professor, and Theodore Simon printed the first intelligence test scale, known as the Simon-Binet Scale, in 1905. This scale focused on verbal abilities and became the prototype that resembles modern-day intelligence tests (Goldstein et al., 2019).

The German psychologist, William Stern, devised the term intelligence quotient or “IQ” in 1912, which used the intelligence scale that was developed by Binet and Stern, and defined intelligence as a broad capability or mental adaptability of a person to modify and change their thinking to deal with new requirements or problems and conditions of life. The “IQ” quotient was represented as the ratio between mental age and chronological age (Goldstein, Pulakos, Passmore & Semedo, 2017).

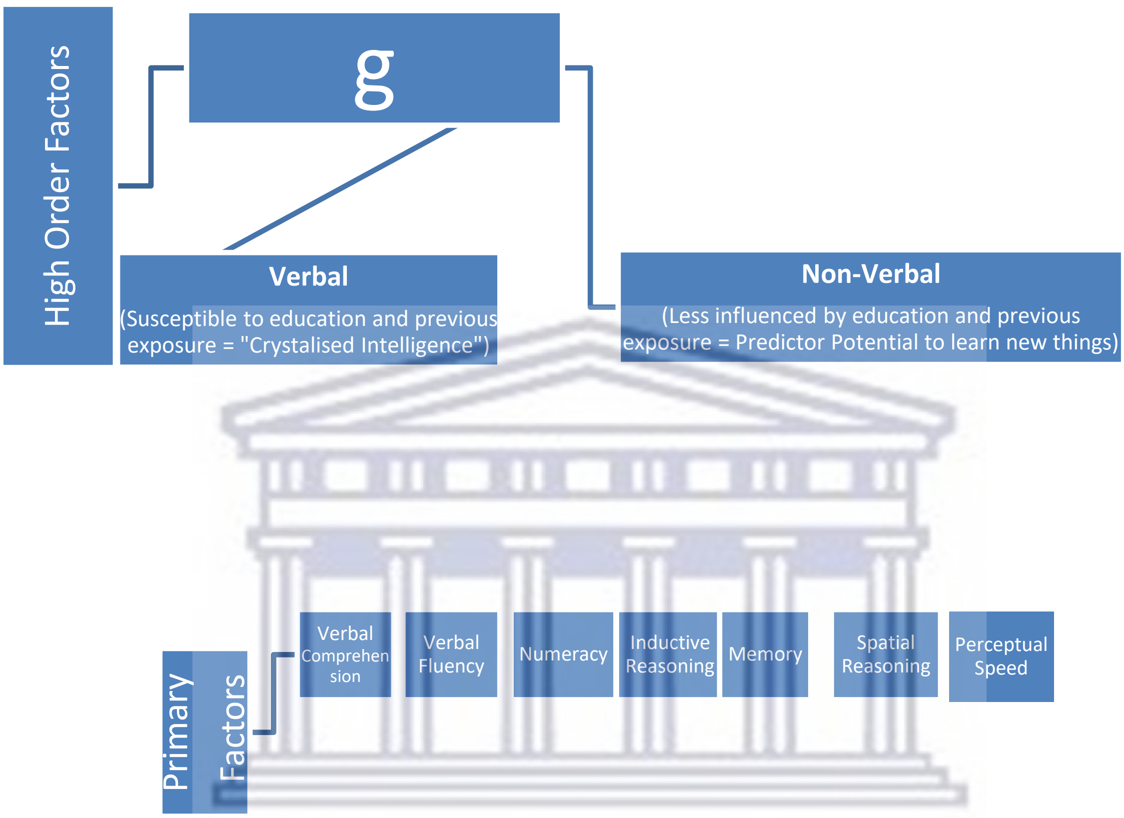
The measurement of cognitive functioning and the study of intelligence is marred by problems and controversies. There is still no agreement between psychologists about how to define these terms and explain exactly how they function and how they should be measured. Therefore, many different approaches and theories have been developed over the decades and remain controversial as each has its supporters and critics (Foxcroft & Roodt, 2018).

The coining of the term "IQ" gave rise to the development of the first theory of intelligence in 1904 by Charles Spearman, an English psychologist. Spearman used many technical assertions to explain the different aspects of his theory but in laymen's terms, people are born with general intelligence or "***g***". The degrees to which they have it differ and are displayed by how they complete intellectual tasks, and because it is innate, it is immune to remediation or improvement (Gardner, 2012; Goldstein, Allen & De Luca, 2019). The two-factor theory of Spearman identified two components of intelligence. The first component or factor is referred to as the "***g***" factor, which is the general capacity for reasoning and problem solving, and the other factor is referred to as the "***s***" factor, which is the ability that is specific to solving a single aspect of a problem (Omoluabi, 2006). By the 1930's most of the revised tests and editions focused on providing a general estimate of either verbal or non-verbal performance (Goldstein et al., 2019).

According to the Psytech test manual (2017), the larger the quantity of data that needs to be handled, the more vital "***g***" becomes. Therefore, the information-handling load of a job increases, and a person with lower general mental ability is unlikely to succeed when compared to a person with a higher "***g***" (Psytech SA, 2017). To simplify the understanding of this concept, please see Figure 2.1 presented on the following page.

Figure 2.1

Structure of abilities



Source:

Adapted from Psytech S.A. (2017). *Technical Manual for Critical Reasoning Battery*. Psytech South Africa catalogue: products, services and training offered by Psytech South Africa. Johannesburg: Psytech SA.

In 1916, Lewis Terman from Stanford University adapted the Binet-Simon Scale and standardised it for American use. The subsequent Stanford-Binet Scale was not only used and popular in the United States of America, but also in France. It also sparked the start of significant issues regarding the testing of minorities, as the initial standardisation sample comprised only middle-class Whites (of Western European descent). The deliberate exclusion of Mexican-Americans, African-Americans, and other groups of colour, was criticised for decades as revisions of the scale continued to include only Whites. The first revision that included other groups of colour was more than half a century later, in 1972. Furthermore, the views and findings of Terman also gave rise to the perception that minorities and groups of colour had a lower intellectual level and were less educable (Reynolds, Altmann & Allen, 2021).

In 1931, Thurstone, a psychologist at the University of Chicago, differed with Spearman about the existence of an overarching "***g***" factor. Thurstone became the leading advocate of the Multiple-Factor Theory, which extended Spearman's theory by replacing the single-factor of "***g***" with a set of group factors that were correlated. In so doing, Thurstone identified seven basic mental abilities: language comprehension, general reasoning, verbal fluency, memory, numbers, spatial speed, and perceptual speed (Gardner, 2012).

According to Foxcroft and Roodt (2018), the Swiss psychologist, Piaget (1936, 1950), developed the intelligence model, which consists of the four phases of human growth. Piaget's theory asserts that the four stages of development differ qualitatively and are distinct from each other, although patterns are associated with each growth stage. The earliest phase is the Sensorimotor phase, which happens from birth to two years old. The next phase is from two to six years old and is called the Pre-operational phase. The third phase is the Concrete operational phase, which happens between seven and twelve years old. The final phase is called the Formal operational phase, which occurs from twelve years onward.

In the early 1940s, the theory of fluid and crystallised ability was developed by Raymond Cattell. Fluid intelligence is defined as the capability to examine problems, recognise patterns and associations that support the problem, and uses logic to solve it. This form of intelligence is influenced by genetic factors and is not dependent on educational experience. Another distinction is that fluid intelligence is susceptible to the decline of the central nervous system, whereas crystallised intelligence remains relatively intact. Crystallised intelligence is based on the cultural and educational experiences acquired by interacting with fluid intelligence, and these get assessed by the verbal and numeric reasoning ability tests (Gardner, 2012; McGill & Dombrowski, 2019; Wasserman, 2019).

The Cattell-Horn-Caroll model derives from the combined work of Raymond Cattell in 1941, John Horn in 1965, and John Carroll in 1993 and combines the theoretical models of fluid intelligence, crystallised intelligence, and the three-stratum model (McGill & Dombrowski, 2019; Wasserman, 2019). Initial investigation of the model in collaboration with John Horn gave rise to the Cattell-Horn model; further refinement of this model by John Carroll gave rise to the Three-stratum theory, which later was called the CHC model. The CHC theory holds that the relationship between fluid and the crystallised intelligence can be found by categorizing it into three different strata. These strata are known as the narrow abilities, the broad abilities, and a single general ability known as "***g***" which is at the highest stratum of the model (Cormier et al., 2017; Flanagan et al., 2002). According to Alfonso et al., (2005), the CHC model had substantial influence over the assessment of cognitive abilities and the analysis of intelligence tests and became widely regarded as the most influential theory in the study of human intelligence.

According to Foxcroft and Roodt (2018), the CHC taxonomy has received increased support from psychologists and is grounded in empirical evidence with implications for intervention from a theoretically grounded framework. This view is shared and corroborated by researchers who assert that due to increasing validity evidence and the revisions as well as the modifications to the theory of fluid and crystallised intelligence, intelligence test development are influenced (Cormier et al., 2017; James et al., 2015). Intelligence test developers acknowledge the prominence of CHC theory in outlining and understanding cognitive ability theories. Due to empirical research and literature reviews conducted by them, the CHC theory is widely used. It forms the basis for choosing, arranging, and deducing tests of intelligence and cognitive abilities (e.g., Flanagan et al., 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998). Furthermore, the CHC theory was used to classify achievement so that it both aids the understanding of theoretical abilities as well as to provide a basis to organise assessments for individuals that potentially have learning disabilities (Flanagan et al., 2002; McGill & Dombrowski, 2019; Wasserman, 2019).

Due to the limitations of the Stanford-Binet Scale, in 1955, the American psychologist David Wechsler developed and published his intelligence measure called the Wechsler Adult Intelligence Scale (WAIS). Test scoring is not based on chronological and mental age, but it equates the test user's score to the scores of others in that age group. This scoring method became the custom in intelligence analysis and is utilised in modern revisions of the Stanford-Binet test (Reynolds, Altmann & Allen, 2021).

Howard Gardner (1999) postulated a “theory of multiple intelligences” that not only relies on the context of cognitive processes recognising numerous mental abilities or talents that constitutes intelligence. Gardner asserted that the progress of “artistic” skills was restrained in the models of cognitive and developmental psychology. Therefore, he suggested seven relatively independent types of intelligence. These are musical, physical kinaesthetic, logical-mathematical, semantic, spatial, interpersonal, and intrapersonal skills. While this method included different facets as part of intelligence and thus broadened the view of intelligence, the practical consideration limits applying this view (Foxcroft & Roodt, 2018).

In the mid-1980s, Sternberg proposed the triarchic theory of intelligence, which suggests that instead of viewing intelligence as something obtained from a test result, to rather view it in the contexts in which it occurs, since people adapt to the real world and environment that shapes them. Therefore, this theory fits in the broader category of contextual intelligence. This theory includes analytical or componential intelligence, creative or experiential intelligence, and practical or contextual intelligence. Componential or analytical intelligence deals with the way a person analyses information or data, such as the strategic speed and processing and acquisition of knowledge to get a solution to the problem. Experiential or creative intelligence refers to solving a problem quickly without error if the situation is known from before, routine or if the situation is new. It, therefore, follows that the more intelligent the person is, the quicker they will solve the problem with fewer mistakes. Practical intelligence looks at how well the person fits in their external environment. The more intelligent the person is, the better they either adapt to their environment, or select an environment more suited to them or change or are more equipped to shape the environment to suit them (Gardner, 2012; Goldstein et al., 2019; Reynolds et al., 2021).

Foxcroft and Roodt (2018) describe another concept of theoretical intellect and the system or data handling style. According to this style, intellect comprises three components, namely attentional practices, information practices, and planning practices, and this theory has various variations and models that fall into this category. The Sternberg Triarchic Abilities Test (STAT) is grounded in conceptual intelligence theory or practical intelligence, commonly referred to as street-smart. The System or Information-handling approach became known as the Cognitive-processing approach, and the PASS (Planning, Attention, Simultaneous and Successive) model is an example of this approach. Furthermore, Gardner (2012) discusses a category of biological theories, which covers numerous models and studies conducted on the human biology of the brain to understand how intelligence works.

Warne (2020) asserts that of all the scientific facts related to intelligence research, the one he would give anything to change is the existence of the average "IQ" scores across different racial or ethnic groups. Researchers do not argue about group differences in intelligence test scores but over what causes these differences. The gap in the "IQ" between White and Black students was a subject of this argument, especially in the United States. This argument has elicited comments and research from people that support and share polar opposite views, namely, that these differences are not genetic, that environmental factors could be responsible for these differences, or that these differences indeed are genetic (Warne, 2020). Jensenism was one of the theories that emerged from this debate when an educational psychologist, Arthur Jensen, published his work in 1969 that suggests that "IQ" is inherited and that the 15-point score difference across the racial groups was genetic (Clay, 2017). In 1984, James Flynn published his work that documented that "IQ" scores have increased steadily over two generations, and an 18-point difference existed between different race groups. He attributed this difference to cultural factors and named it the Flynn effect (Trahan, Stuebing, Fletcher & Hiscock, 2014).

From this vast array of theories discussed in this section, it becomes clear that many models exist on intelligence and cognitive ability, but also that these range from vastly simple to complex. Furthermore, no single theory of intelligence was accepted or endorsed as universally accepted in psychology. What all these theories do, however, shows that there undoubtedly are individual intellectual differences in test performance, and the unending development of models continue to shape the controversial discussions and the prospect of intelligence testing in psychology.

The greater the amount of data handled, the more vital "***g***" becomes. Concerning the current research, the CRTB2 version of the test instrument is recommended for South Africa, and the theory applied to this test is that of Spearman (PsytechSA, 2017). Despite the myriad of models, the researcher's preference would be to accept the theory of Spearman as this theory attempts not only to quantify or differentiate what intelligence is, but the notion of "***g***" tries to measure and quantify general intelligence. Even if some theories accept the idea of general intelligence, while others argue that it is too broad and focus on specific constructs, the tests used and data generated often measure constructs related to intelligence. Furthermore, considering how interchangeably many of the concepts are and how similar the constructs measured by the different tests are, at least the "***g***" factor provides a means that statistically accounts for any variance that may exist. Despite this, the most contentious issue for the researcher is the tainted past that intelligence testing has because of the history of individual differences and elitism. These cannot be ignored, regardless of the political views of a country, because democracy and equality are at the heart of these issues.

2.3 THE SOUTH AFRICAN LANDSCAPE

2.3.1 BACKGROUND

The data used for this study is secondary data from a retail company based in the Western Cape. It is necessary to understand the South African legislative requirements regarding testing because the usage of psychometric tests is under statutory control. It is vital to visit the history of the country and the developments of testing historically; due to the past discrimination legacies of the country and the known disparities across the different population groups. In addition, because South Africa is multi-cultural and has diverse languages, it is crucial to understand the impact that these may have on the study.

According to Foxcroft and Roodt (2018), psychological testing in South Africa dates back to 1915. Initially, testing was mainly for researching individual differences in intellectual and cognitive functioning. Given the historically uneven dissemination of wealth based on racial classes of White, Coloured, Asian and Black, subsequent developments in psychometric assessments mirrored and reflected this while it was evolving. The legacy of apartheid had a profound impact on test development, given that the earliest tests were used by the Department of Education to accommodate White scholars in superior schooling. The tests were standardised for Whites only and compelled by the political ideas of the time to differentiate between racial groups to display dominance of one population group over other groups (Clay, 2017; Foxcroft & Roodt, 2018).

The end of World War II ushered in a crucial requirement to identify the appropriateness of employment predominantly in mines, which saw a need to assess Blacks with very little formal education and who spoke various dialects. Education was segregated along racial lines as the apartheid regime practiced job reservation. Therefore, different tests were used for different population groups, as different population groups did not contend with each other. Western tests imported were standardised mainly for Whites, and tests developed locally were predominantly for Whites (Foxcroft & Roodt, 2018).

In the early 1940s in South Africa, a group of researchers worked primarily with intelligence tests. They consistently used tests that were normed and developed for White children and used these as a standard to compare the results achieved by Black children (Laher & Cockcroft, 2014).

Before 1994, South Africa was not a democracy because under the rule of the previous government dispensation, apartheid was practiced, which caused certain racial groups within the country to be marginalized and not allowed to enter particular jobs, especially managerial positions, within the labour market. When South Africa had their first democratic election in 1994, much of these practices changed when the country transformed. The ruling government was keen to correct past mistakes and ensure that all the citizens were treated equally and had the same basic rights. As a result, candidates from diverse ethnic groups started to contend for the same positions. Tests were designed for many racial groups, and norms were included for more race groups to address this. Alternatively, the tests which were standardised on White-only groups and the tests imported from abroad, were administered to various groups. Due to the lack of suitable norms, it became a common practice to interpret these test results with caution or adjust scores for specific race groups. The research was lacking to determine whether these tests were appropriate for a diverse South Africa and define whether the measures were prejudiced (Arendse, 2021; Clay, 2017; Cockcroft et al., 2015; Foxcroft, 2011; Venter & Levy, 2009).

During the apartheid years, it was common for separate tests to be used to test different population groups, because these groups rarely competed for the same job. When the socio-political condition in South Africa improved, the use of particular tests for diverse groups was no longer constitutionally suitable. Therefore, tests were being used cross-culturally, and an investigation was required to define the appropriateness of these tests (Abrahams, 1996; Arendse, 2021; Foxcroft, 2011).

2.3.2 LEGISLATION

Due to the extensive use and abuse of possibly culturally biased instruments, the perception increased that the apartheid Government could exclude specific race groups from employment and occupational opportunities. Negative perceptions about the usefulness of psychometric testing grew, and the South African populace started rejecting testing. That resulted in psychological assessments becoming statutory controlled, and the Health Professions Act (No.56 of 1974) was introduced to restrict the use of psychological measures to properly registered psychology professionals. Regulating this also prevents test developers from prematurely releasing these tests before validity and reliability are established (Clay, 2017; Foxcroft & Roodt, 2018).

The post-1994 democratic South Africa developed and adopted, what many describe as, a progressive Constitution that gave rise to other legislative documents to ensure that all their citizens are treated fairly and with dignity. To rectify and control past injustices in the employment sectors, the Labour Relations Act (No 66 of 1995) was promulgated. This legislation became the cornerstone of all labour-related matters and aimed to regulate all aspects of employment to a minimum standard that is in the interest of both employers and employees but focuses on protecting all employees from any injustice or discrimination (Foxcroft, 2011; Reuben & Bobat, 2014; Steyn & De Bruin, 2020; Venter & Levy, 2009).

In the education sector, many reforms were passed, which saw school-readiness tests and systematic group tests banned in schools. The usefulness of testing practices and test results were queried and published for public comment in the *Education White Paper 6, Special Needs Education: Building an Inclusive Education and Training System* (Department of Education, 2001). In addition, the role of the school psychologist and support staff is changing. This change will offer subsidiary care to students, assist instructors with identifying the needs of learners, and assist with developing plans to remove learning barriers (Foxcroft & Roodt, 2018).

The apartheid legacy prevented Black people and females gaining access to equal schooling and competing in the labour market. It would not be possible to simply appoint and promote people due to past irregularities to fix things. Therefore, the South African Government introduced an affirmative action policy to redress these past discriminations, and later the Employment Equity Act (No. 55 of 1998) was promulgated. The purpose of employment equity was to eradicate prejudice in the workplace and to address disadvantages in the workplace by ensuring equal depiction of elected groups in all groupings and ranks of the organisation (Dupper, Bhoola & Garbers, 2009; Reuben & Bobat, 2014). Section 8 of the Employment Equity Act (No. 55 of 1998), stipulates that:

> "Psychological testing and other similar assessments are prohibited unless the test or assessment used: (a) has scientifically been showed to be valid and reliable; (b) can be applied fairly to all employees, and (c) is not biased against any employee or group" (p16).

Employment equity provides workers and trade unions, which historically had no protection, with much-needed support, by legislating the requirements for test usage and focusing on fair assessments. This Act has implications for companies and practitioners, as many of the measures used currently have not been investigated for bias or cross-culturally validated (Clay, 2017). It forced practitioners and companies to evaluate which test they are using and the quality and appropriateness of the measures, as they may need to prove in a court of law that the assessment measures do not discriminate against any specific groups (Clay, 2017; Foxcroft & Roodt, 2018).

The Employment Equity Act provides a legal structure for implementing affirmative action in South Africa. Preferential measures ensure that the designated group of appropriately qualified individuals have equal job opportunities and are represented equally at all job levels in the designated employer's workforce (No. 55 of 1998). In 2013, additional legislation, the Broad-Based Black Economic Empowerment Amendment Act (No. 46 of 2013), was passed and dealt with the inclusion of Black people, which includes the Black, Asian, and Coloured race groups. Note: B-BBEE compliance for businesses is not a legal obligation, and companies are not obligated to disclose or verify their B-BBEE status. It is viewed as a business prestige and attempts to include people in the economy, empower them, and targets historically disadvantaged groups such as Black people, women, the youth, rural communities, and disabled people (No. 46 of 2013).

Even though the scope of this research did not include gauging the perceptions of staff, it is worth adding the study of Reuben and Bobat (2014), which conducted research in the retail sector of South Africa. There is limited research conducted in this sector, and this research gauged the perceptions held by staff about affirmative action. The main finding of this research was that the employees and the company continued at that stage to rationalise the racial order of competence that was prevalent during the apartheid years.

It is noticeable that there are many challenges that are specific to the South African context. There are numerous legislative pieces in place in an attempt to govern and control many of these challenges. These are inherently due to the ramifications of apartheid and the dark history of psychological testing and aim not to disadvantage any groups further. It has implications for individuals, companies, and researchers. It is good to have controls in place if it leads to improvement. However, this can only be achieved by legal means if there is effective policing by the government that ensures compliance and implementation of these legislative pieces.

2.3.3 OTHER ISSUES

The social context in which people live is an important aspect to consider, especially when looking at aptitude and intelligence testing. The variation in the external environment, such as socioeconomic status, influences a person's performance on measures of ability. Socioeconomic status refers to the social prestige of an individual, and the key indicators are education, occupation, and income. What an individual has learned during life is derived from their school experience. Schooling forms the basis of problem-solving, which is also the basis of intelligence testing. Therefore, the test score that an individual achieves on an intelligence test can indirectly be an indication of their level of education. Disparities experienced in South Africa led to large imbalances and poverty. Socioeconomic status plays a significant role because it influences the experiences that moderate abilities, behavior, and attitude, which can directly impact test performance. It may influence the facilities that the person can access, such as schools, libraries, opportunities that arise, and the attitudes of those in the same group. However, in South Africa, this condition is complex because the quality of education in poorer communities is not necessarily at the same standard as those in more affluent areas. Therefore, people in disadvantaged areas may have different awareness and experience than learners in the same grade (Clay, 2017; Foxcroft & Roodt, 2018; Shuttleworth-Edwards, 2012).

Psychological assessments can add to the proficiency of selection, placement, development, and the controlling of HR procedures; therefore, to safeguard society against misuse, only classified tests should be used. Traditionally, tests are procured from countries abroad and used in different sectors and industries. Initially, tests were established separately for Afrikaans and English-speaking groups but omitted people who spoke African languages, even though this is the largest population group in South Africa (Arendse 2021; Van De Vijver & Rothmann, 2004).

Performance on assessments can be influenced by language if tests administered are not in the test-takers home language. It may be experienced as the individual being slower to complete the test in another language or may feel they can respond better in their dialect. Language can thus become a possible basis for unfairness. Well-validated tests may not be appropriate for a country wherein 80 percent of its people use English as their second language. South Africa has 11 officially recognised languages, and while ones home language may not be English, most schooling is in English. Residents in townships may speak a vernacular form, referred to as slang, yet have a different home language. Test translation becomes another challenge to translate the test in all 11 languages. There is no easy answer on how to overcome this thus language remains a challenge in psychometrics in South Africa (Clay, 2017; Foxcroft & Roodt, 2018).

When administering tests, consider that interns and clinicians do not automatically speak the native languages of the customers they are attempting to attend. Higher education is taught in English, but many students learn English as a third or fourth language. Therefore, it may be worth exploring a strategy that uses qualitative approaches to complement quantitative information and shift the focus from test scores to descriptive meaning (Laher & Cockcroft, 2014).

The validation study on the General Reasoning Test (GRT2) provided a range of results and findings. The majority of the results of the research conducted in this area show that race, education, language, and English language comprehension are important factors that influence items and construct relevance and that the quality of education received in high schools has an influence as well (Abrahams, Friedrich & Tredoux, 2012).

According to Long (2017), in psychology, questions about the social context, economic and structural violence, and how it manifests itself in peoples' lives tend to be ignored when administering tests. It often leads to inappropriate interventions administered for the challenges that people face.

A psychologist who scores tests uses a deviation score method. It derives from the fact that scores on intelligence tests create a bell-shaped curve called a normal distribution; the scores of the person tested are compared to their peers in the same group, which is called a norm group (Warne, 2020). The published Jensenism research data suggests the existence of a 15-point score difference across the racial groups over time, which illustrates how important it is that norm groups are updated continuously. Furthermore, because intelligence research and testing originates in Western culture, many of the norm groups and cut-off scores that are applied for selection purposes are derived from these norm groups of the existing test (Laher & Cockcroft, 2014).

According to Warne (2020), the standard for the usefulness of a test is not whether the test is perfectly accurate, but whether they are more accurate than the alternative decision-making strategies. The researcher's view is that, while the various tests may indeed measure "***g***", it is incorrect to presume that the instrument is always ethnically suitable, especially if the test was standardised for use in a different country than the country of origin. The decisions made from these test results directly affect the public livelihood and education opportunities if they are not selected based on the score they achieved. Tests need to be proven to be scientifically reliable and the correct norm group must be used, to be a good decision-making tool. Some of these tests are not scientifically sound and should not be used. A further retardant should be the enormity of the impact it may have on the livelihoods.

2.4 PSYCHOMETRIC TESTING PROPERTIES AND TEST STANDARDISATION

Psychometric testing in South Africa is legislated, and the promulgation of the South African Employment Equity Act was introduced in 1998, amongst some of the legislation, to assist with compliance and re-instill trust in psychometric testing (Republic of South Africa, 1998). The law regulates the use of psychological tests and other similar instruments. Tests may be used if valid and reliable, can be used equally by any group, and only if they are not partial against employees or groups (No. 55 of 1998).

The view of Van der Merwe (2002) is still relevant when the author reported that efforts are needed to develop new tools to meet South African legal requirements and practices and to test existing tools for use by multicultural groups. The infancy stage of research about the likeness and partiality of assessment instruments in South Africa necessitates more studies before psychology as a profession meets the desires of labour law (No. 55 of 1998).

A more recent view by Clay (2017) echoes this view of Van der Merwe in the article about decolonizing psychology in South Africa. It highlights the dark history that the discipline of psychology had in South Africa and what some of the initiatives are that psychologists are doing post the aftermath of apartheid to try to make psychology relevant and valuable for the majority of South Africans. It asserts that the goal is not only to undo the remnants of White supremacism but also to warrant that psychology is beneficial in a diverse society that continues to battle against disparity (Clay, 2017).

It is crucial to look at the efficacy of psychometric testing since research notes concerns raised about this. Therefore, some of the vital concepts of psychometric testing are described with an examination of psychometric testing conducted in South Africa. Foxcroft and Roodt (2018) duly suggest that psychological assessment is entrenched in global education, recruitment, and civil and criminal justice and that various individuals in South Africa demand detailed psychological assessment.

According to Laher and Cockcroft (2014), most studies investigated were grounded firmly within the prediction paradigm. They point out that unlike the situation in other countries, the proliferation of reliability and validity studies was not just a psychometric activity but also rather a response to the criticisms of the South African testing industry. The notion was that psychometric tests were used to support the racist agenda of White supremacy and ignored the systematic prejudices that favored White individuals. It is reinforced by numerous studies that point to the legitimacy of work capacity legislation as a reason for their research (Laher & Cockcroft, 2014).

According to Bryman and Bell (2014), some criticism against scientific research is that it often relies heavily on instruments and procedures, which hinders the connection between the study and everyday life. It, therefore, poses the question of how one can be sure that if respondents answer a set of questions in a specific manner, one can ascertain that those responses reflect and connect to their everyday life. Therefore, in research, the validity of an instrument establishes whether the concept that is supposed to be measured is really measured.

Various types of validity are used to ensure that conclusions generated from research have integrity. In lay terms, the validity coefficient indicates whether the test processes the concept that it theoretically should measure (Babbie, 2021). The higher the validity coefficient, the more confidence is assured that the test theoretically measures what it should measure (Bryman & Bell, 2014). This research does not aim to compare the various types of validity; it tries to establish whether differences exist among the scores achieved by the different groups, since the test was standardised. It should assist when comparing this to other research conducted and drawing conclusions about the validity, fairness, and consistency. It is essential to know and comprehend the difference between reliability and validity so that the interpretation of the research results is understandable. These concepts are referred to in Chapter 4, which deals with a discussion of the results.

According to Odendaal (2015), cognitive ability psychometric tests indicate either the proficiency or deficiency of specific skills or abilities. In South Africa, standardised tests of reasoning ability evaluate verbal ability, numerical ability, and logical reasoning. The scientific accuracy of psychometric tests helps establish reliable and valid evaluations of various skills or abilities. For this reason, psychometric tests are often used as complementary sources of information when companies need to make selection decisions to recruit, promote or develop employees.

According to Laher and Cockcroft (2014), inadequate research attention to testing practices and the technical concerns of test creation and validation in South Africa exist. Furthermore, many psychometric instruments are not standardised in South Africa, which results in inherent cultural bias. Standardisation of the test refers to the standardisation of the test itself to ensure that the test items are relevant to the country using the test. Appropriate norm groups must be established for the test scores and administering assessments consistently. However, standardisation cannot replace validity because groups have different standard scores, and the estimated goodness-of-fit curves may also differ (Laher & Cockcroft, 2014).

When predicting the criteria on which the test is constructed, the test is biased toward members of a population subgroup if the baseline scores forecasted from the regression line are consistently too high or too low for subgroup members. It may mean that subgroup members are not selected when they achieve adequately (Mahembe, 2014).

When evaluating psychometric tests, the concerns of prejudice and fairness are crucial, as this attempt to answer the question of whether test scores attained by diverse cultural groups can comparatively be read. Different types of prejudice may explain the answer to this question (Foxcroft & Roodt, 2018).

Language can be a source of prejudice because participants have different levels of language proficiency when tested in a second or third language. It can be a crucial factor in multicultural assessments, even if language ability is not assessed (Arendse, 2021; Van De Vijver & Rothmann, 2004).

Besides the origin and local development of the psychometric test, which remains challenging in South Africa, other biases exist. These are related to racial groups, language proficiency, and others mentioned in the previous section. There is a long history of imported test modification for use in South Africa, and the indigenous measures developed during apartheid were normed against White South Africans. Many of the tests used locally in South Africa were standardised so that it is harnessed for use in South Africa, which also presents challenges. This practice differs from validating tests, which have new norm groups specifically developed from populations in South Africa. It is an accepted practice, but cautions regarding this make it more pertinent to examine the research on the actual psychometric test, specifically those used in the South African context, which was not locally created. The above research informs us that it is vital to know about and look at the psychometric properties of the instruments used and to use similar tests in different contexts and geographical regions, which can produce related or vastly different results. Therefore, it is critical to remember this when conducting research, especially when interpreting the results and drawing conclusions or generalising.

2.5 APTITUDE TESTING RESEARCH AND FINDINGS

This section examines research on psychometric test instruments used for aptitude testing to aid comparisons of the findings to research already conducted and related to this research's topic. It may assist in understanding and interpreting the results of this research. The studies included were arranged chronologically from the oldest to the most recent unless particular studies to substantiate the findings of local studies were added and indicated accordingly. The order of studies started with the studies conducted on the South African populations, then those on African countries, and finally, international studies. It assists in understanding the research conducted locally and comparing these to international research. Note that this research topic dates back to the 19th century, but the researcher endeavored to include the most recent research studies. Older studies cited were those conducted locally, as the research on this topic in South Africa is limited. Also, the research topic dates back to research conducted long ago, and specific attention is focused on studies in this area of research looking at the current trends related to this research topic.

2.5.1 SOUTH AFRICAN STUDIES AND SUB-SAHARAN STUDIES

A study by Lynn and Owen (1994) aimed to determine whether actual differences in "***g***" exist for Black-White differences in South Africa, similar to those found in a United States of America study. Their study used the South African Junior Aptitude Tests (JAT), created in South Africa. It was standardised for White students in Grades 7-10, consisting of 10 tests of basic skills. School psychologists conducted the test between 1985 and 1986, and the sample used for that study was 15-16 year-old adolescents in South African Grade 9 secondary schools. The study found substantial test score differences among the three groups and that all the White-Asian and White-Black differences were statistically significant (Lynn & Owen, 1994).

Rushton and Jensen (2005) conducted a thirty years review study of racial differences in cognitive ability. Scientific research debates about individual and group differences in cognitive abilities started in the mid-19th century when there were widespread uses of standardised mental tests in World War I. A 15-point difference in Black-White "IQ" scores was found in a study conducted in the United States of America, which was believed to be genetic or hereditary. Herrstein and Murray (1994) subsequently published the Bell Curve Study, which provided evidence that there was a significant 1.1 standard deviation difference between Black and White Americans, which is accepted as empirical evidence and not disputed, but the cause for the difference remains disputed and continues the debate. Rushton and Jensen (2005) found the average "IQ" scores worldwide of Whites to be about 100, Black Americans about 85, and Blacks in sub-Saharan Africa to be the lowest at 70. This study concurs with the research findings by Lynn and Owen (1994), which is that Blacks in South Africa scored lower than their peers in the United States of America.

Kemp (2000) investigated the properties of socioeconomic status and educational quality of a stratified sample of Grade 12 learners and graduates scored on the Wechsler Adult Intelligence Scale (WAIS-III) test. The study showed a decidedly positive association between high socioeconomic status and WAIS-III score performance and that the influence of moderately poor quality of schooling and reduced level of education on WAIS-III score performance was considerable (Kemp, 2000).

A study conducted by Rushton, Skuy and Fridjhon (2002), administered the Raven's Standard Progressive Matrices Test to engineering students at the University of the Witwatersrand in South Africa. Using 1993 Unite States norms, their findings report the "IQ" equivalents of Africans were 97, Asians were 102, and Whites were 110, in that order. The African – Asian – White variances indicated a variance in "***g***" or the common factor of intelligence, and they were interpreted as the ''Jensen Effects''. The study did not find any performance differences across gender groups (Rushton et al., 2002).

Previous studies of ability tests show differences observed in mean score performance across groups and could not exclude item bias. In some cases, Black subjects assessed in English revealed that language and reasoning style were possible sources of bias. It suggested that using distinct norms for diverse population groups reverses the aim of a general test. Research supports that although South African tests are reliable and valid for a designed and standardised group, it necessitates score comparisons of the individuals from different groups. This finding was made by the author after conducting an electronic literature review of research done in South Africa, and comparing this with research conducted at institutions, which included technical colleges and universities, to probe the concept of dynamic assessments (Murphy & Maree, 2006). However, Foxcroft (2011) cautions that cross-cultural validity was not explicitly established in several tests. It may mean that those score comparisons with other cultures may provide discriminatory information if used unreservedly (Foxcroft, 2011).

Research conducted on university students by Rushton (2008) used Raven's Standard and Advanced Progressive Matrices Test. The study confirms forecasts that differences exist in the average scores of the different race groups and that the group average remains consistent over time and place. The study used the following categorisation of race groups and showed the ranking order of the group "IQ" averages are East-Asians, Whites, South-Asians, Coloureds, and lastly Blacks, with the "IQ" scores of this ranked order as = 116, 113, 106, 103, and 98, respectively (Rushton, 2008).

A study that reviewed the published data of sub-Saharan African studies using the Raven's Progressive Matrices test examined the Flynn effect. It found that the although the reliability and predictive validity equated to the Western samples, the test itself was a fairly weak indicator of the general intellect among Africans. Through factor analysis, it found that the test often measures additional factors and determined that the Flynn effect has not yet appeared in the sub-Saharan African groups (Wicherts, Dolan, Carlson & van der Maas, 2010).

An article published by Foxcroft (2011) aimed to share the learnings of a decade of research studies conducted to share insights related to ethical issues in psychological testing in Africa. The main finding and personal lesson shared by the writer was that to ensure ethical practice for test interpretation, establishing local norms before testing should occur. This practice is sometimes not adhered to in South Africa, as appropriately developed native standards are absent. Furthermore, as part of ethical test practice, it is required that locally developed test norms are available to ensure accurate interpretation of results. However, the standards used were those from Europe, the United Kingdom, and the United States of America and were applied with caution but with little attempt to develop new local norms. It is problematic and can lead to incorrect decisions taken based on the test score achieved.

A study of the General Reasoning Test (GRT2) test was conducted by Abrahams, Friedrich and Tredoux (2012), to assist with identifying potentially successful students at a university in South Africa. A validation study was conducted on the GRT2 test and recommended to take care when interpreting test results due to the differences between language and gender groups.

A study conducted by Odendaal (2015), found that culture and language moderate the relationship between proficiency tests. Secondary data sets from the test publisher were used, which included results from 1640 job applicants of various industries. The results indicate group differences in general reasoning, with the Nguni and Sotho groups scoring lower on general reasoning and higher on social desirability than the Afrikaans and English groups. This study supports the research conducted in the United States of America. According to Van der Merwe (2002), the research conducted in the United States of America showed that the American Black population scores were below that of the White population. These differences were not due to the traditional features of the tests, for example, if the test was verbal or non-verbal or if the tests administered were for individuals or groups.

A study conducted by Cockroft et al. (2015) compared the performance of multi-lingual and low-socioeconomic South African university students against that of mono-lingual and high-socioeconomic British students on the WAIS-III intelligence test. The norm population of this test is Western-based, and it is one of the most widely used tests in the South African industry.

The findings confirmed that bias exists in verbal and non-verbal subtests and that this test is not necessarily culturally fair.

Statistical analysis of secondary data that was supplied by the Umalusi Council to study the performance of Grade Twelve learners in the Province of the Western Cape, South Africa (Letsoalo, 2018). Results of the research show that female students scored significantly more, at least 0.45% more than male students. It recommended that institutions consider the difference in scores achieved between male and female students to improve the student's abilities to learn. This study, therefore, suggests that females achieve higher academic results in the Western Cape Province than their male counterparts (Letsoalo, 2018).

A recent study by Nieuwoudt, Dickie, Coetsee, Engelbrecht and Terblanche (2019) suggested that Coloured women display lower cognitive abilities than women of other racial groups. This study drew much criticism and was retracted by the publisher, who cited that the article was peer-reviewed and published but subsequently retracted due to flaws existing in the methodology and reporting of the original study.

According to Steyn and De Bruin (2020), samples collected from 52 South African organisations with 60 employees representing 3143 respondents were 56.4% males and 43.6% females. Four instruments were used in the study, for which the measurement variability was determined and used on both gender groups. Correlation and regression analysis were performed separately for men and women, and the results were compared to assess the association between the antecedents. No differences were observed between the sex groups.

More recently, Arendse (2021), investigated gender discrimination in cognitive testing by conducting a validation study of the newly developed English Comprehension Test Battery. Not only did the findings of this research contradict the belief held, which is that women are more accomplished at verbal tasks than men, but it highlighted that language cause differential performance in assessments in the multi-cultural setting of South Africa. The study used a non-probability sample of 881 participants. The results are inconsistent with many international studies showing that women outperform men on language cognitive assessments because two-way ANOVA did not differ statistically between men and women (Arendse, 2021).

Even though the following study is not current, it is one of few studies undertaken that directly compare how the different race groups within South Africa scored against their peers in other countries. It compared the scores achieved by the South African group against the scores achieved by the race groups in the United States of America. Due to the uniqueness of this study, the author felt it was valuable insight and information to include, even if it is old. The study utilised the South African Junior Aptitude Tests (JAT), created in South Africa and standardised on White learners in Grades 7 to 10 and consisted of 10 tests of basic skills. The school psychologist conducted the test between 1985 and 1986, and the sample used for this study consisted of 15-16-year-old adolescents in South African Grade Nine secondary schools. Substantial test score differences existed among the three groups, and all the White-Asian and White-Black differences were statistically significant, $p < .0001$ (Lynn & Owen, 1994). This research concurs with the mainstream findings, which is that a 1-point standard deviation difference exists between the Black-White race scores and that the South African population tested scored a 1-point standard deviation lower than the population of the United States of America. Thus, giving credence to the 15-point score difference in intelligence testing.

2.5.2 INTERNATIONAL STUDIES

According to Lynn (2006), the three largest races in East Asia, Europeans, and Africans, are the focus of most research on racial differences in intelligence. His book discusses more than 500 "IQ" studies published worldwide since the early 20th century. The highest "IQ" scores range from 105 for East-Asians (Chinese, Japanese, and Koreans), followed by the Europeans ("IQ" 100), then the Inuit or Eskimos ("IQ" 91), South East Asians, and Native American Asians ("IQ" 87), Pacific Islanders ("IQ" 85), South Asians and North Africans ("IQ" 84). The lower "IQ" scores are the sub-Saharan Africans ("IQ" 67) followed by the Australian Aborigines ("IQ" 62), with the lowest scoring Bushmen of the Kalahari Desert together with the Pygmies of the Congo rain forests ("IQ" 54). He concluded that further evidence might need to confirm some estimates of the racial "IQ" differences, but have high confidence, because other studies show similar results. Lynn shows a strong correlation between mathematics and science achievement exists in international research studies. It helped reverse a century-old consensus that there was a difference in "IQ" scores between men and women, suggesting that men score 4 to 5 points higher than women. He concluded that racial differences in intelligence are 50% genetic and 50% environmental and consistent with other recent reviews (Lynn, 2006).

An international study by te Nijenhuis (2013), states that the occurrence of the Flynn effect is determined by environmental factors, raising the question of whether these factors are responsible for group differences in intelligence. It argues that it has clearly been shown that group differences correlate strongly with "***g***" loadings and that empirical studies have conflicting findings. Therefore, this study aimed to present new evidence by using datasets from the United States and the Netherlands and found that there is a small negative correlation between the Flynn effect gains and "**g**" loadings. It concluded that the Flynn effect and the group differences have different causes (te Nijenhuis, 2013).

The positive correlations were well established between measures of different mental abilities. Empirical findings and reputed evidence are available of a general factor in different measured mental abilities. It is termed the "***g***" by Spearman and refers to general factors reflected in individual differences on all psychological, mental tests regardless of their content. They compared Blacks, Whites, and Asians from South Africa and compared them to a sample of White Americans and a sample of Black people from Zimbabwe (te Nijenhuis et al., 2015).

Spearman's two-factor hypothesis tested in this study was that there would be small differences across racial groups in the lower "***g***" subsets because differences across racial in a subset of "IQ" batteries were a function of the "***g***" loadings of that subset. The study included comparisons of young Korean adults to several groups of adults from Canada, the United States, Russia, Peru, and South Africa and was conducted by te Nijenhuis, Choi, van den Hoek, Valueva, and Lee (2019). It had a sample of 4770 young adults that took the Advanced Progressive Matrices Test. The study concluded that Spearman's hypothesis was still a valid phenomenon (te Nijenhuis et al., 2019).

The above research informs us that similar instruments used in the same or different contexts and different geographical regions can produce similar or vastly different results. Therefore, it is crucial to remember this point when conducting research and considering it when interpreting the results and drawing conclusions or generalising. The above studies were all based on psychometric test instruments for aptitude testing to assist with comparing the findings of research already conducted to this research study. Including research conducted locally and internationally assists with comparing similarities and differences, assists the researcher's discussion of these research results, and reaching conclusions about this study.

2.6 SELECTION

Organisations are made up of people that work together toward a common goal. The people supply their work, talent, creativity, drive, and initiative to the organisation in exchange for money. Competent people must be employed at all levels in the organisation to pursue and achieve the goals of the organisation. The workforce also changes over time as people are transferred, promoted, replaced, or resign; therefore, organisations plan and manage their workforce through recruitment and selection processes. The recruitment process attempts to attract large enough groups of potential employees with the abilities and aptitudes for a specific job so that management can make the selection decision and appoint the most suitable person (Zunker, 2016).

People are less predictable than machinery, but they are the resources of a business. Selection aims to bring talented people into the organisation that will perform their roles while achieving self-development and satisfaction through their work. Selection, therefore, matches people to the jobs available to achieve the goals of the organisation. Each organisation usually evaluates potential employees using a preferred set of abilities and aptitudes for specific jobs to evaluate the candidates. Various psychometric tests assist with the selection decision, but for these tests to be effective, they need to be validated by the organisation (Arendse, 2021; Foxcroft, 2011; Oosthuizen & Naidoo, 2010).

There are three broad areas of industry assessments, and the instruments used in the psychological measurement of individual attributes, can be classified as psychological measures. The other two areas relate to the assessment of groups and organisations, and these are not necessarily classified as psychological measures. Valid and reliable instruments should be used to assess the individual differences for both selection and hiring purposes, or to differentiate for training, promotion, and development (Foxcroft & Roodt, 2018).

Making informed and appropriate decisions is part of life, and psychological assessments can assist with this as assessments can serve many purposes. A variety of assessment tools are available to enable us to do this. The controversial field of cognitive assessments and individual differences in cognitive ability is important for life in general, education and occupational success. However, because different rating scales give different scores, which are often used to attack cognitive scales specifically, but are a clear and important predictor of success in both professional and educational contexts (Foxcroft & Roodt, 2018).

Various advantages are directly or indirectly associated with higher intelligence, which in psychometric terms, is defined as a specific level of cognitive functioning. This includes issues related to a good education, a higher level of employability, high positions in society, etc. It is important to understand what each test measures and the intended use of the test should be the basis to limit how the results are interpreted (Foxcroft & Roodt, 2018).

Matsepe, Cross and Fenyane (2020), conducted a study on the regulation of admission and selection of students to universities in South Africa. The research included students from the universities of Witwatersrand, Cape Town, and KwaZulu-Natal. They found that while universities made progress in addressing racial issues in their admissions strategy, their primary target audience remains students from affluent communities, leaving out potentially high-achieving students from marginalized groups attending township and rural schools, especially where funds are scarce. They found that the universities have shifted from issues of race to issues of social class, and survey data confirmed that most institutions agree that some form of affirmative action is needed to address current social disparities. However, the intense debate over turnaround policies in South Africa's higher education sector has made it challenging to develop and implement in practice appropriate recruitment strategies for institutions (Matsepe et al., 2020).

Research conducted by Mangara and Chiyindiko (2019) found a similar trend in the education sector selection process. They found that South African researchers had previously attributed an under-representation of disadvantaged gifted students to their use of traditional identification methods, including intelligence quotient ("IQ") and standardised achievement tests. Furthermore, they point out that international research used innovative non-traditional identification methods, which include non-verbal tests, student portfolios, and emoticon checklists to address under-representation.

The data generated by psychological assessment measures are useful and should supplement the selection decision, not replace it. An input-based selection approach is used when the personal characteristics are consistent with the job requirements. This method is predictive and different from the output-based approach, which relies on the outcome, which determines if the minimum standard is met. An output-based selection approach is used when an individual is compared to the required outputs of the job and this method is also referred to as the competency-based method. In both instances, there is a need to address the challenge of bias and adverse impact. The selection strategy should not afford a particular group less or more likelihood of being selected than the other group (Foxcroft & Roodt, 2018).

The Labour Relations Act enforces defensible selection methods to avoid unfair labour practices, which implies that the employer must be able to prove that they appointed the most suitable person for the job, using measurable criteria. Furthermore, the Employment Equity Bill of 1998 infers that psychometric testing of employees is prohibited unless it can demonstrate that it respects diversity. This can be achieved by ensuring that tests are validated and that steps are taken to ensure that tests are culturally correct and unbiased for marginalized groups, which include previously disadvantaged groups of colour, women, and the disabled (Clay, 2017; Cockcroft et al., 2015; Foxcroft, 2011; Muller, 1999).

An international study conducted to compare testing practices between the United States of America and other countries found that the domain of preferential treatment makes the United States an obvious outlier. Preferential treatment for different cut-off scores or individual test standardisation, or within group selection strategies is prohibited by United States law. In contrast, in South Africa, preferential treatment is allowed and applied. Racial quota allocation is legal and used by many large employers. The practical implication of this, is that it is legal in South Africa to use racial norms or top-down selection strategies within groups to meet the affirmative action goals of the organisations. Furthermore, the circumstances in which preferences are acceptable differ from country to country, and several survey respondents noted that lower standards apply to protected groups, in particular in Australia, India, and South Africa. The study also found documented studies in South Africa, showing that the differences in average cognitive test scores across Black and White groups are greater than in U.S. studies (Myors, Lievens, Schollaert , Van Hoye, Cronshaw, Mladinic, Rodriguez et al., 2008).

In South Africa, one of the primary goals of assessments is to align current practice with legislative demands. This is possible if new instruments are created and existing instruments are validated to use in multicultural groups. South African organisations typically use a range of assessments and interviews as complementary sources to aid their selection decisions (Clay 2017; Foxcroft, 2011; Laher & Cockroft, 2014; Van De Vijver & Rothmann, 2004). This research study used the CRTB test results, which are used in conjunction with other psychometric tests and interviews, to make the final selection decision. An international study conducted by Myors et al. (2008), focussed on legislative environment of selection in a various countries. Psychologists from 20 countries were given questionnaires concerning the aspects of the legal environment in their country, and provided their feedback to these. These enquiries were organised and centred on the description of the legal environment of the United States so that comparisons could be made.

This study enquired whether any specific recruitment approaches were restricted or forbidden due to legislation or court rulings of a specific country. The study found the most outstandingly different style to control recruitment application was found in South Africa, because the collective method is that the employer has the right to use a particular method as the South African legislation places this burden on the employer (Myors et al., 2008).

During 2012, The Society for Industrial and Organisational Psychology in South Africa (SIOPSA) published Guidelines for the Validation and Use of Assessment Procedures for the Workplace for practitioners to safeguard that their assessment tools and processes fulfill the scientific necessities and international best practices. However, these procedures were very recently updated (SIOPSA, 2022). These guidelines were mainly centered on the American Society for Industrial and Organisations (Myors et al., 2008).

Van der Merwe (2002) researched various companies based in the Eastern Cape in South Africa and focused on the types of selection instruments used by different organisations. It further highlights the situation in South Africa, described by the research in the preceding paragraph above. When conducting his study, he found that no permanent or generally recognised, or customary recruitment procedures were used by all organisations. The landscape of the establishment, the duty for recruiting, and the customs of HR management, are all factors influenced by the selection procedures of organisations. These procedures can range from simple to very complex. The mainstream recruitment strategies are based on a technique termed the successive hurdle technique, which entails that candidates effectively complete and pass a variety of selection steps and that unsuccessful applicants are excluded at each stage or obstacle, which they have not passed. Because psychometric tests provide objectivity and validity, which can be proven for the specific test, it is included as a selection tool by most organisations. However, these tests must be viewed as merely a single opportunity in the selection process. It should thus become only an aid and should never be used exclusively, or substitute the entire selection procedure (Van der Merwe, 2002).

An international study by te Nijenhuis et al., (2015) used the Standard Progressive Matrices instrument. The results show that cognitive testing is widely used in organisational selection and placement strategies and is increasingly used in educational settings. Significant correlations of "IQ" (Intelligence Quotient) with many educational, economic, and social criteria are well established. This has continued a long-standing curiosity about the average difference in "IQ" test scores between the largest populations in the United States, individuals of European descent who are socially identified as White, and individuals of African descent who are socially identified as Black or African-American, which are the most studied.

Sampling from diverse populations is challenging because of the reality of group differences in the learning and vocational potential needed to succeed. Organisations in South Africa have difficulty meeting equity targets when selecting candidates from a diverse pool while maintaining production and efficiency metrics. It is necessary to enable those selected for development activities to successfully develop the professional skills needed to succeed in the workplace. To enable accurate prediction of learning outcomes in assessments, it is necessary to understand what the determinants of learning outcomes are and how they combine (Mahembe, 2014).

The assurance that each applicant will receive a fair and equal chance to apply for and selected for a position is called equity. People cannot be excluded from a job, based on arbitrary grounds that include race, gender, religion, etc. Human Resources managers in South Africa are expected to develop personnel selection procedures that can make predictions of job performance for multi-cultural groups from the variety of available selection techniques, even though little empirical research is available to assist their decisions (Clay, 2017; Foxcroft, 2011; Laher & Cockcroft, 2014; Van der Merwe, 2002).

According to Theron (2009), selection decisions should be based on the expected performance of the evaluated criteria without systematic prediction errors associated with prediction groups. Furthermore, under-represented groups may be unfairly disadvantaged by the use of such predictive paradigms. Tests that are valid and reliable are deemed more culturally fair, but in these contexts, reliability and validity only refer to the information collected in the past, so you should remember this when making your final decision (Theron, 2009).

Although the results of a South African research study by Odendaal (2015), showed that the differences between the two groups of Blacks and the two groups of Whites were small, the two groups of Blacks in the study scored significantly lower on average than the two groups of Whites in general reasoning ability. In addition, Nguni and Sotho-speaking participants scored lower on the General Reasoning Test 2 (GRT2) than Afrikaans and English-speaking participants. It suggests that there is a significant risk of negative outcomes associated with disproportionately selecting White participants over Black participants when the cognitive ability is the primary selection tool. In addition, a literature review of studies provides evidence that job-related predictors often include measures of cognitive ability that produce mean group score differences, potentially contributing to adverse outcomes (Odendaal, 2015).

According to Roodt and La Grange (2001), research were conducted using the Verbal Evaluation Test 3 (VCC3), which measures a person's capability to comprehend and assess the logic of more complex written arguments. The test instrument has a reported reliability (Cronbach Alpha of 0.85), as reported by the User's manual for a sample size of 700 that was used. The study using multiple regressions found that certain personality dimensions significantly predicted job performance and that verbal reasoning had no significant predictive power of job performance. However, the results of this study cannot be generalised to other industries or roles due to the limitation that the population for the study consists of insurance salespeople of a large South African corporation.

A study conducted in Zimbabwe that used the DAT Test (Differential Aptitude Test) data from companies based there aimed to understand the relationship between individual ability test results and actual job performance. It was observed that some people perform well in their job even if they have not passed the psychometric tests. The main finding of this study was that the three subtests of DAT Test: verbal reasoning, numerical reasoning, and abstract reasoning, were positively correlated with job performance. This finding significantly gave insight to the Zimbabwean companies that participated in understanding the relationship between psychometric tests and job performance and could be used to inform their recruitment and selection policies (Munhamo, 2014).

An Indonesian study by Setiawati (2020), found that the development of the psychological dimensions influences the development of the quality of the tests used. The DAT Test was used in this study and stated that passing the aptitude tests must be proven by following the aptitude test. It investigated the predictive validity of the differential aptitude tests (DAT) to predict academic success by tracking the performance of 148 students majoring in psychology at a public university. The study concluded that the verbal and numerical subtest of the DAT Test could best predict academic success, and the DAT Test could predict success for the students enrolled in the psychology program.

A study conducted by Anazia (2019) investigated the predictive power of quantitative and verbal skills on economic performance in high school students studying Economics in Nigeria. Three test instruments were used in this study: the Economics Achievement Test, the Quantitative Aptitude Test, and the Verbal Aptitude Test. As a result, both the Quantitative Aptitude and Verbal Aptitude Tests have predictive power, but the Quantitative Aptitude Test had greater predictive power than the Verbal Aptitude Test.

Typical standardised tests of cognitive skills are used in South Africa to assess verbal skills, numeric skills, and deductive reasoning. Selection of candidates for development programs is an important point, as candidates selected from the development pool are critical to succeed in their chosen program and career. Efforts to get this right can contribute to transformational efforts in all sectors for skills development and employment equity. It can be seen from the research in this section that different organisations and institutions used various selection processes and tests, therefore having clear selection policies per industry or sector may be needed to streamline this process more, not only to ensure that similar test instruments are used, but also that whatever process is used is fair and equal to everyone considered for selection.

2.7 CONCLUSION

In this section, a literature review was conducted, which looked at the research conducted in this area. It aimed to provide the theoretical background needed to understand the current study. The chapter examined aptitude testing and how it has progressed over time. It also discussed psychometric tests and their properties. It looked at the South African landscape with a focus on the legal aspects related to South Africa. It also looked at selection practices, as the fair selection of employees for a position is important for business and South African legislation.

Using valid and reliable assessment tools to make selection decisions are critical for successful selection procedures as it can have legal implications in South Africa. Throughout the chapter comparatives were done with the international context and that of South Africa. The next chapter will examine the research methods utilised for this study. It will provide an exchange of the design of the study, examine the sample and population of the study and the procedures that are employed for this research and investigate the test instrument used. It will describe the collection of the data and the statistical techniques that will be used to test the hypotheses.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 INTRODUCTION

The purpose of the study was to determine if there were significant differences across the mean test scores obtained on the Critical Reasoning Test Battery related to differences in groups. This was done to inform us about the effectiveness and consistency of the CRTB2 psychometric test and allowed the interrogation of the data recorded from the employees for selection into management development programs in a retail company based in the Western Cape. The aim was to understand and determine whether any disparities exist among the results achieved by the participants. The previous chapter discussed the literature regarding the selection and the use of the psychometric test in the South African landscape. This chapter will describe the study methodology, study design, and participant respondents used in this particular survey. The instruments used for data collection and the statistical methods used in this study will also be presented. It will also present the ethical considerations applied in this study.

3.2 RESEARCH DESIGN

A quantitative research approach was adopted, because it is the best way to achieve the research objectives. Data was generated by the retail company that conducted CRTB2 assessments on candidates considered for management development. The CRTB2 is one of the psychometric test batteries used as part of the selection process, therefore, one of the tools that determine which candidates were firstly short-listed and finally selected to be part of the management development pool.

According to Bryman and Bell (2014), quantitative research deals with the quantitative measurement of data collection and analysis. A logical approach that emphasises detailed planning before data collection and analysis, which uses a deductive approach to test theories while investigating the relationship across theory and research.

The research design used is exploratory, as this type of design is suitable for cases where little is known about a research question and aims to provide insight and understanding when using literature reviews and secondary data (Sallis, Gripsrud, Henning Olsson, & Silkoset, 2021).

Exploratory research is often done to find something new but the results of the research determine whether something new is added by the research. It is done to gain background information about a phenomenon or issue and can be used to simplify research questions and hypotheses and to set research priorities. It is often used as a research question when the researcher does not have prior information or when there are limited studies to refer to (Thomas & Lawal, 2020). This study has been done in this design because no published research was found on the CRTB2 test in South Africa, and this test is used by various businesses as a selection tool.

The main information source was secondary data collected through the psychometric testing conducted by the retail company. This data was made available for this research and was drawn from the database of the retail company under the supervision of a registered industrial psychologist to ensure the anonymity of the participants.

According to Odendaal (2015), using secondary data is appropriate, and to compensate for the effect of sample selection, using appropriate independent variables can be useful to observe the presence of bias when evaluating an instrument. In this study, race, age and gender were used as independent variables because it is particularly important concerning diversity in South Africa.

Data generated by other researchers may be made available to the general research community for reuse. Using secondary data has the benefit of it being available at low or no cost and allows the researcher to explore questions that may not have been previously asked but can be answered by the data available to the researcher. Because this study is conducted as part of the fulfillment of the researchers' Masters degree at a university, the research is archived in the university library repository. The research can thus be accessed in the future and be used to do comparative research or to replicate the original research. The research may also be used to re-analyse the data and answer questions that may not have been asked by this research study. The results of this study can contribute to the existing knowledge framework and used for teaching and learning (Hox & Boeije, 2005; Steyn & De Bruin, 2020).

A limitation of using secondary analysis is that the data may not have all the key variables present, making it difficult for the researchers to draw unambiguous conclusions (Bryman & Bell, 2014). Other limitations to using secondary data relate to the inability of the researcher to control data collection errors that were recorded in the system. There is no control over the choice of samples and comparison groups, as well as the quality of the sampling frame, which could affect the generalisability of the results (Odendaal, 2015).

Using the secondary data of the retail company had the following limitations to this research study:

- The researcher had no control over the data did not indicate the educational level and gender representation of the company-provided samples
- The researcher had no control over the group categorisation of the data as these were following the reporting requirements of the Labour law of the South African government
- The researcher had no control for incomplete captured data points or missing values and used the data as provided by the organisation
- The study's data represent the total baseline score obtained for each test and do not have the individual item scores

3.3 POPULATION AND SAMPLE

A population is an entire collection of people, events, or things studied in a research investigation (Sallis et al., 2021). The population in this study was the workers of the retail company based in the wholesale and retail sector, who completed the Critical Reasoning Test Battery between the years 2010 and 2017 to be placed on the company's management development program. Following the requirements of the Employment Equity Law, the following categories were used to classify the race groups: Blacks, Coloureds, Asian and White. Sample studies allow the researchers to draw general decisions from the population (Sallis et al., 2021).

Convenience sampling was used as the research used the secondary data available from the retail organisation. The biggest benefit of convenience sampling is that sample selection is easily accessible and free, therefore cost-effective, and saving time because the results were readily available. A further advantage is that because the data source is internal to the company and confidential, it is not readily available to competitors. The limitations are that the available data sets did not have complete sets available for all the parameters, therefore only the completed sets of data were used. The generalisation of results is limited because of the sampling technique used. In addition, the age of the data is the data which were available for the participants that were tested by the company, therefore the gender, age, and race groups include the information that were available for the participants tested and not necessarily representative of the sector and region the research was conducted in (Sallis et al., 2021).

For this research results of the selected group of the retail company in the Western Cape (N=202) were used to conduct research analysis on the secondary data. Initially, a total of 215 data points were available for which 13 points did not have the scores available for both numeric and verbal subtests and were excluded. For racial and gender categories (n=202), as those were the number of data points for which race and gender were recorded. There were (n=57) respondents that had their age included and completed both the numeric and verbal tests, therefore these were included for age categories. The age group of the individuals was between 26 and 55 years. Education was not a strict criterion for admission, but in some cases, the position held and the grade of the individual was available; therefore, these parameters were not included in this research. Biographical information will be discussed in more detail in the next chapter.

The study adopted non-probability sampling, which means that because convenience sampling was used, the samples were selected from the population only because they were conveniently available (Bryman & Bell, 2014).

3.4 PROCEDURE FOR DATA GATHERING

The process for the gathering of the data for this study involved obtaining permission from the retail company based in the Western Cape to utilise the data obtained and gathered for the 7-year period as data points. During this process, ethical considerations about the confidentiality of the information, the purpose of the study, and the agreement of individuals' test results were discussed in detail. When undergoing psychometric testing at the company, the individuals signed consent that the data gathered by the organisation belongs to the company and the company has permission to use the data for research or analytical purposes.

The organisation also requested a formal written request from both the University and the researcher and on receipt of the researcher's research proposal, the researcher was given written permission to use the secondary information under the supervision of a registered Industrial Psychologist to ensure that correct data handling and confidentiality protocols were observed at all times. Once permission was obtained from the manager of the Talent Department of the company, a contract was signed to set suitable dates for the data to be drawn from the database by a registered Industrial Psychologist. This was done for two reasons. The first is to avoid interruption of office work, and the second was to give the researcher access to the required information under the supervision of the registered Industrial Psychologist to ensure the anonymity of participants.

When the CRTB2 tests were conducted, between 2010 -2017 the organisation employed a few registered Industrial Psychologists to administer the test. These were administered according to the test instructions provided in the test manual. It was required by each respondent to sign a consent form before taking part in the test and gave consent that results were the property of the organisation and were treated as confidential and used for research purposes. The results were captured in the test provider's database. Then these results were drawn from the database, given to the researcher and included only the gender, race, and age where this information was available plus the actual results to ensure anonymity. Statistical analysis was performed on the data and the results of these will be shown in Chapter 4 of this research.

The research proposal was ethically approved by the University of the Western Cape's Ethics Committee before proceeding with the study. It was stipulated that the information used was the property of the retail company and that anonymity was upheld by not disclosing the name of the organisation, and that the data was only used for this specific research study.

3.5 MEASURING INSTRUMENTS

The Critical Reasoning Test Battery, also known as the CRTB2, was reviewed and then classified as a psychometric test by the Psychometrics Committee of the Professional Board of Psychology in July 2005. The technical manual does not specify the date of development or the period for which the standardisation data were collected. The test battery consists of two instruments that measure a candidate's verbal reasoning ability and numerical critical reasoning ability. The data collected has also been graphically presented to simplify the reading and the interpretation of the biographical information (PsytechSA, 2017).

In the Verbal Critical Reasoning Test Battery Second Edition a correction for guessing has been incorporated. It is based on the number of items a respondent got wrong on the test since the number of times a respondent got the answer right can be estimated by the number of incorrect answers they gave. The registered Industrial Psychologist of the retail company, who provided the data, however confirmed (via the personal information provided to the researcher in person) that the correction was not applied to any of the data points supplied for this study (PsytechSA, 2017).

The critical reasoning tests were standardised for use in South Africa by using a mixed sample of graduates, managerial and professional groups. Norm tables are made available in the test manual. The validity and reliability data provide comparative data for gender differences and educational and socioeconomic status (PsytechSA, 2017).

Two subtests of the CRTB2 are designed to assess the verbal critical reasoning and numerical or quantitative reasoning skills of individuals in scientific, engineering, financial, and professional roles and individuals who must make strategic business decisions based on written or quantitative information data. Students and full-time staff subgroups attended the first pool of the standardisation process. The test manual provides norm tables for the various subgroups, which are undergraduate students, professionals, managerial and technical groups (PsytechSA, 2017).

Both these second-edition batteries were developed using data from graduate students and is intended for individuals with above-average intelligence who may hold senior management positions (PsytechSA, 2017). This test battery is used to assess critical reasoning, which is both logical and deductive, rather than simply checking verbal and numerical ability. It assesses an individual's ability to reason, think critically and draw logical conclusions from verbal and numerical information without identifying factual errors and inconsistencies (PsytechSA, 2017).

Personal communication from the registered Industrial Psychologist explained that the norm tables provided in the test manual have various categories and cut-off scores, which are used as part of the process when interpreting the scores of the respondents. Cut-off scores are norms that are set and used to interpret test scores, which usually compares the individual's achievement to that of a group (Foxcroft & Roodt, 2018). The Industrial Psychologist further explained that the company uses the recommended cut-off score for the category of professional, managerial, and technical occupations when evaluating how well the respondents scored on both the verbal and numeric tests. Their selection policy advocates that the results should be used in conjunction with other assessments, such as interview scores, to make an overall decision. However, the placement manager can enquire about how the respondent scored on the test.

3.5.1 PSYCHOMETRIC PROPERTIES

As indicated, the Critical Reasoning Test Battery consists of two subtests, namely numeric reasoning and verbal reasoning. The Psytech SA (2017) test batteries are multiple-aptitude batteries designed with minimum recommended education and/ or appropriate employment level in mind. The verbal and numeric reasoning tests are among the tests that are identified as suitable ability tests for directors and senior managers (Foxcroft & Roodt, 2018).

The CRTB2 was standardised on a sample of 4625 working-age adults in various occupational and managerial occupations. The mean age of the standardised sample was 37.1 years, of which 30.1% were female. 27% of the sample identified as non-White (European). Of the total sample, 12.2% identified as Asian and 6.1% as Black (eg, Afro-Caribbean, Black African, etc.). 3.1% are Pakistani, 1% are Bangladeshi, and the remaining 5.6% are of various ethnicities (eg, Pacific Islander, Maori, etc.) (PsytechSA, 2017).

Reliability indicates the test is free from errors. A commonly used measure of the reliability scale is internal consistency, the degree to which all the items that make up the scale measure the same underlying property. The most frequently used statistic is Cronbach's alpha coefficient which indicates the average correlation between the items that make up the scale (Sallis et al., 2021). Values range from 0 to 1, with higher values here indicating higher reliability, and a minimum Cronbach alpha value of 0.70 is recommended. The alpha coefficient of the Verbal Critical Reasoning Test for professionals or managers is 0.87 and the Numeric Critical reasoning alpha-co-efficient is 0.81. The correlation between the Verbal and Numeric reasoning subtest is 0.54 for the professional and managerial categories (PsytechSA, 2017).

Babbie (2021), describes reliability as a matter of repeated use of a particular method. It gives the same result every time. A scale that ranges from (r= 0.78) or higher is reported as having a good test-retest reliability.

3.6 STATISTICAL TECHNIQUES

The statistical procedures used in the study were descriptive and inferential statistical methods. There are two descriptive statistics which describe the phenomenon of interest. These are the measure of central tendency that allows the investigation of the central points of the data-sets and the measure of dispersion, which allows to investigate the variability of the data-sets. It allows researchers to conclude from a population sample (Foxcroft & Roodt, 2018).

Inferential statistics included T-tests, Pearson's product-moment correlation coefficient, and analysis of variance (ANOVA). By ordering and manipulating the raw data collected, the researcher developed descriptive statistics that were presented as tables or histograms, or bar charts. Correlations were calculated to determine how one variable relates to the other, and inferences were made based on the values determined. A T-test was performed to determine whether groups were different and inferences were drawn from these (Bryman & Bell, 2014).

According to Bryman and Bell (2014), descriptive statistics include measures of central tendency to determine the average typical value and variability. It shows how strong the deviation from the central tendency is. This helps researchers to organise, summarise and simplify their findings. The mean is a measure of central tendency that provides a complete picture of the data without unnecessarily suppressing every observation in the data set. Data is usually presented in the form of a histogram or a bar graph (Sallis et al., 2021).

Inferential statistics allow you to conclude from the data that can determine the relationship between two variables and how the other independent variable might explain the variance of the dependent variable. On the other hand, it shows whether differences exist between variables, and between different subgroups. The T-test is used to test for significant differences between different types of means. In this study, the T-test was used to determine whether there was a significant difference in numeric reasoning ability and verbal reasoning ability between gender groups (Witte & Witte, 2017).

A Pearson's product-moment correlation is a numerical summary that indicates the strength or magnitude of an observed relationship. It also indicates the direction of the relationship (Bryman & Bell, 2014). For this study, the correlation coefficient for this test was used to understand how significant the correlation of the data was, as well as whether age had an influence on the scores achieved by the participants and whether any of these were statistically significant values.

According to Van De Vijver and Rothmann (2004), a normal distribution in the population is assumed by psychological constructs as the individual's observed score on a test reflects the actual state associated with the component that they are being measured on with an allowed degree of random measurement error. Therefore, when investigating to study the function of tests in different cultural groups, various methods such as means difference analysis and correlations have been used. In this study, the mean differences and standard deviations were examined, using the White group as the reference group to compare the other groups and to determine differences between the groups (Steyn & De Bruin, 2020).

The individual scores achieved on the multiple-aptitude battery used the actual raw score that was achieved by the test taker as this was available from the data provided. The resulting scores for the individuals were compared to others in the norm group so that their strengths or weaknesses could be identified. However, the level of skill an individual can achieve depends on his/her general intellectual abilities, interests, personality traits, attitudes, motivations, and knowledge (Foxcroft & Roodt, 2018).

3.7 ETHICIAL CONSIDERATIONS

Before conducting this research, written approval was elicited from the retail company. The research process necessitates ethical considerations about the confidentiality of the information, informed consent, anonymity, beneficence and non-maleficence, related to the purpose of the study, and the agreement that individual test results be meticulously safe-guarded. Detailed discussions were held with both the retail company and the University to ensure that good ethical conduct was always observed by all the stakeholders. These discussions carefully considered all the stakeholders needs, including the rights of the participants and their consent given to the organistion about the usage of their information and results.

The organisation requested a formal written request to conduct the research from the University and the researcher. On receipt of the research proposal, the researcher received written consent to use the secondary data under the supervision of a registered Industrial Psychologist. It formalised the correct etiquette for data handling and confidentiality.

All participants signed consent that the data gathered by the organisation belongs to the company and granted their permission to use the data for research or analytical purposes and ensure confidentiality. This is termed informed consent and all research undertaken must secure this consent from the participants. In this case, the organisation ensured that this was done before conducting the test with each participant.

The data collected by the organistion are captured on the database of the orgaisation and only accsible by authorised personell. Confidentiality was maintained so that it would guard the business interests and well-being through the protection of their identity from unauthorised parties. Anonymity involves the ethical protection that the organisation remains nameless so that its identity is not disclosed and remains unknown (Bryman & Bell, 2014). Similarly, the confidentiality of the participants was the company's and the researcher's main priority.Therefore,the identification of the individuals were not provided by the organisation and the researcher does not disclose the identity of the organisation.

Non-maleficence requires that the researcher cause no harm against any of the research participants as a direct or indirect consequence of the research (Bryman & Bell, 2014). The current study did not foresee or anticipate any harm. Beneficence implies that the researcher demonstrates compassion and takes positive action to help others, with the general desire and intention to do well to others (Bryman & Bell, 2014). The results of this research provided insigthts for the organisation and enabled them to evaluate the usage of the CRTB2 test for their selection process of their management development program, which may benefit previously excluded selection groups.

The research received approval from the University of the Western Cape's Ethics Committee before proceeding with the study. It was stipulated that the information used was the property of the retail company and that anonymity was upheld by not disclosing the name of the organisation.

3.8 CONCLUSION

This chapter described in detail the research methods and sampling method used. The target group and measurement instrument used (eg, the CRTB2 test battery) were discussed along with its psychometric properties. The chapter concluded with a discussion of the statistical methods used for hypotheses testing (eg, T-tests, Pearson's product-moment correlation) and the ethical considerations. In the next chapter, the results will be presented and interpreted using the Statistical Package for the Social Sciences (SPSS), 2016 edition.

CHAPTER FOUR

PRESENTATION OF THE RESULTS

4.1. INTRODUCTION

This chapter will present the results of the research so that a discussion about the data analysis process and the interpretation of the results can be done during the last chapter. The main objective of the study was to evaluate secondary data from the retail company according to one of their selection test batteries and to understand whether there were any differences between the test scores achieved by the different gender, age and race groups that have completed the CRTB2 test battery. The findings from the secondary data will be analysed using statistical tools. First, the demographic information about the participants will be presented, and then the results of the data analysis will be conducted.

The aim of presenting the data is to allow a logical flow of the results achieved and to understand and determine whether there are differences in the means for various groupings of the data tested. It is to determine if any outcomes are significant or not so that the various hypotheses formulated are tested and will assist with understanding the usefulness of this psychometric/ assessment tool for selection. Thereafter, the results will be presented using inferential statistical analysis such as ANOVAs to test for differences and T-tests to determine the correlation for relationships.

4.2 DATA ANALYSIS

4.2.1 DESCRIPTIVE STATISTICS

Descriptive statistics are used to provide descriptive data for the overall sample and the groups. Graphs are developed to graphically display this information. The data are statistically analysed using the Statistical Package for the Social Sciences (SPSS) version 16.

Before running the statistical analysis, all data with missing values for race, age, and gender for both verbal and numeric reasoning test scores are discarded and therefore not included for analysis, which results in the total sample size (n) being 202 data points. The final raw scores of each of the verbal reasoning and numeric reasoning subtests are used as the scores per item were not available, and to reduce any calculation errors that may occur to the final scores.

Gender and race are categorical variables; therefore, the data points are coded to run the statistical analysis on these. Categorical variables are items that have the same meaning and the same number, but in reality, there are several possibilities. An example of this is, that a human being is either a male or a female or in South Africa people are classified by race groups namely, White, Black, Coloured, and Asian (Witte & Witte, 2017). To statistically analyse this information, these variables are coded as numeric numbers, such as the number 1 representing males and the number 2 representing females.

4.2.1.1 GENDER OF RESPONDENTS

To clearly understand the data, graphs and tables are done in SPSS to depict the demographics of the data in terms of the gender of the sample.

Table 4.1

Gender of respondents

	Frequency	Percent	Valid Percent	Cumulative percent
Female	74	36.6	36.6	36.6
Male	128	63.4	63.4	100.0
Total	202	100.0	100.0	

The sample (n= 202) in Table 4.1, shows that there were 74 female respondents and 128 male respondents, which can also be expressed as 36. 6% of the respondents were female, whilst 63. 4% were males. It means that the majority of the respondents were male.

4.2.1.2 RACE OF RESPONDENTS

Similar to the previous section, graphs and tables are done in SPSS to depict the demographics of the data in terms of the race groups of the sample.

Table 4.2

Race of respondents

	Frequency	Percent	Valid Percent	Cumulative percent
Black	23	11.4	11.4	11.4
Coloured	51	25.2	25.2	36.6
Asian	10	5.0	5.0	41.6
White	118	58.4	58.4	100.0
Total	202	100.0	100.0	

Table 4.2 presents the respondents which were White (118), Black (23), Coloured (51) and Asian (10). That equates to White (58.4%), Black (11.4%), Coloured (25.2 %) and Asian (5%). The racial categories of the study are aligned to the reporting categories that are used by South African organisations to report on their employment equity statistics to the Department of Labour.

To demonstrate that the data of this study are consistent with these assumptions and to assist the reader with the understanding of the basic descriptive statistics values, Table 4.3 and Figure 4.1 is included.

Figure 4.1

Histogram of Verbal Reasoning Subtest for Male participants

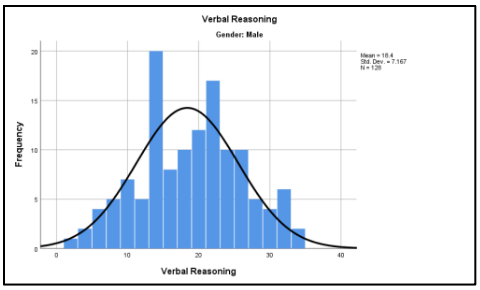


Figure 4.1 is a histogram of the data of the male participants' verbal reasoning test scores, which graphically depicts the data distribution. A bell-shaped curve that can be observed depicts that the data has a normal distribution (Sallis et al., 2021). Furthermore, the scores used are the actual raw score achieved by each test taker, and because the scores were not adjusted, the data complied in terms of the variances and independence of the data.

Table 4.3

Statistics of Verbal Reasoning Subtest for Male participants

Statistics[a]			
		Gender	Verbal
N	Valid	128	128
	Missing	0	0
Mean			18.40
Std. Error of Mean			.633
Median			19.00
Mode			21
Std. Deviation			7.167
Variance			51.360
Skewness			-.012
Std. Error of Skewness			.214
Kurtosis			-.577
Std. Error of Kurtosis			.425
Range			31
Minimum			2
Maximum			33
Percentiles	25		13.00
	50		19.00
	75		23.00

Table 4.3 displays the mean, median, mode, standard deviation, variance, and range of the test scores on the verbal reasoning subtest of the CRTB2 test battery among the male gender group, which was a total of 128 male participants. The sample mean is the average score that was achieved by the respondents and for males, the average score achieved for verbal reasoning was $\bar{x}$ =18.40. The median is the middle value that was scored by the population and in the case of the male participants, this value is 19. The mode is the most frequent value or the score achieved most often by males on the verbal reasoning subtest, which is 21 (Foxcroft & Roodt, 2018).

The standard deviation is σ = 7.167 and shows how much dispersion or variation exists from the average and how well it represents a measure or the average 'fit' of the data. Percentiles split the data into 100 equal parts. From the data it can be seen that the lowest 25% of the scores for males were 13 or below; 50% of the scores were 19 or below and the top 25%, which is the 75 percentile, scored 23 or above (Sallis et al., 2021).

Table 4.4

Normality Test for gender scores on Verbal Reasoning subtest

Tests of Normality							
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Gender		Statistic	Df	Sig.	Statistic	Df	Sig.
Female	Verbal Reasoning	.069	74	.200*	.982	74	.375
Male	Verbal Reasoning	.074	128	.082	.986	128	.219

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4.4 shows the Kolmogorov-Smirnov statistic for gender on the verbal reasoning subtest. This assesses the normality of the score distribution. Insignificant results (Sig. values greater than 0.05) indicate the normality of the data. In this case, the Sig. value of 0.300 indicates that the assumption of normality is met (Pallant, 2011). To conduct this study and to meet the requirements of scientific research, the researcher formulated the hypotheses, which are tested and discussed next as well as the inferential statistical analyses that are performed.

4.3 INFERENTIAL STATISTICS

Bonamente (2013) explains that the theory of probability is intuitive and can assign formulas to the possibility or chance that an event may happen during sampling. Therefore, it is often necessary to guess about the outcome of an event to make a decision, and to do this, the researcher creates hypotheses to express these assumptions. Therefore, hypotheses are developed so that the researcher can draw accurate conclusions about the reality of the sample that they study. Also, the probability is associated with how certain we can be of the outcomes of a study. Assigning probability formulas to these hypotheses assist the researcher when establishing the rules of whether or not to accept these hypotheses (Bonamente, 2013).

In this study, we examine the difference between the means of the populations of the gender, race groups, and age groups. The chance that the means are different between the populations is not dependent on the mean of the other population, which means that the probability is independent and the following formula can be assigned for this (Bonamente, 2013):

$$P\ (A\ \text{and}\ B) = P\ (A)\ P\ (B) \ldots\ldots\ldots\ldots (1.1)$$

Descriptive statistics are used for sample and data description and to examine the information that we can observe. Inferential statistics are used when we want to study or analyse things that are unobserved in the wider population. Therefore, we make generalisations about the sample data, by formulating hypotheses and using statistical analysis such as the T-Test, correlation studies, and analysis of variance (ANOVA) to decide whether any variances are present in the information (Howell, 2013).

Furthermore, we can use these statistical calculations and confidence intervals to draw conclusions about the data and understand the significance level of these conclusions and statistical inferences that are made. Inferential statistics estimates and calculates the uncertainty or variation of the different samples and allows us to understand how much our results differ or how uncertain our findings are. These should then be considered when we draw our conclusions as we analyse the possible ranges of values that we calculate statistically that are representative of the data (Howell, 2013).

It is important to distinguish between theoretical and statistical significance, since this research adds to the present knowledge bank and can be viewed as new knowledge. Even though the hypotheses of a study provide evidence that the information is statistically noteworthy, it does not guarantee that there is academic proof or confirmation that exists to support the results. This is also why the power of the test and a large sample size is important, so that the alternative hypotheses can be supported and the null hypotheses can be rejected. The significance value is stated as 1.00 if there is absolute or 100% confidence that the null hypothesis is correct. Because we can never be absolutely sure it is the truth, the most commonly used significance level in social science is 5%, which means that we can be 95% certain that the alternative hypothesis is correct and not just be a coincidence or accidental (Sallis et al., 2021).

4.3.1 CRTB2 TEST SCALE AND STATISTICAL CONCEPTS

Statistical analysis uses models that assume different things to reflect reality accurately, therefore, for the research to be valid and reliable, these assumptions that are drawn by the researcher need to be true. Four rudimentary rules must be satisfied for the tests to be precise. Therefore, the foundation behind hypothesis testing depends on having data that satisfies a bell-shaped curve to meet this logic of hypothesis testing. Furthermore, when testing multiple sets of groups, it is presumed that each of these samples comes from populations with similar inconsistencies throughout the data. The information from different groups should be autonomous, which means that the performance of one group does not affect the performance of others (Field, 2009; Sallis et al., 2021).

4.3.2 HYPOTHESIS 1

In this research, we explore the relationship among variables and examine the association power between the variables, as well as exploring the differences between groups to determine if any differences exist through formulating hypotheses and statistically analysing these.

H$_1$ propose that: There is a noteworthy difference in the average test totals achieved among the diverse gender groups for verbal reasoning ability.

By conducting the student T-test, we could decide if a statistically significant variance exists in the average scores for the two groups, in other words, whether males and females vary meaningfully in terms of their verbal reasoning ability (Pallant, 2011). The findings are reported next.

Table 4.5

T-test for Gender differences on Verbal Reasoning subtest

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
VERBAL	1(Males)	128	18.40	7.167	.633
	2(Females)	74	18.46	7.242	.842

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VERBAL	Equal variances assumed	.136	.712	-.058	200	.954	-.061	1.051	-2.133	2.011
	Equal variances not assumed			-.058	151.202	.954	-.061	1.054	-2.143	2.021

Independent Samples Effect Sizes

		Standardiser[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
VERBAL	Cohen's d	7.194	-.008	-.295	.278
	Hedges' correction	7.221	-.008	-.294	.277
	Glass's delta	7.242	-.008	-.295	.278

Table 4.5, presents the outcome of the student T-test conducted to equate the verbal reasoning ability results for females and males. Statistical significance is absent for males scores (M = 18.40, SD = 7.167) and females (M = 18.46, SD = 7.242; t (200) = -0.058, p = 0.954, two-tailed). The size of the variances in the averages (mean difference = 7.19% CI: -0.295 to 0.278) was negligible (Cohen's d = -0.008).

Based on the outcomes of the data, this hypothesis is rejected, as there is no significant proof that there are changes in the averages of the different genders groups for verbal reasoning ability.

4.3.3 HYPOTHESIS 2

H2 propose that: There is a significant difference in the mean test scores achieved between different gender groups for numeric reasoning ability. The findings are reported below.

Table 4.6

T-test for Gender differences on Numeric Reasoning subtest

Group Statistics

	GENDER	N	Mean	Std. Deviation	Std. Error Mean
NUMERIC	1(Males)	128	13.31	5.434	.480
	2(Females)	74	12.58	4.737	.551

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
NUMERIC	Equal variances assumed	2.877	.091	.965	200	.336	.731	.758	-.763	2.226
	Equal variances not assumed			1.001	169.817	.318	.731	.731	-.711	2.174

Independent Samples Effect Sizes

		Standardiser[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
NUMERIC	Cohen's d	5.190	.141	-.146	.427
	Hedges' correction	5.210	.140	-.145	.426
	Glass's delta	4.737	.154	-.133	.441

Table 4.6 above presents the results of a student T-test conducted to equate the numeric reasoning ability totals for men and women. There were no noteworthy variances in totals for men ($M = 13.31$, $SD = 5.434$) and females ($M = 12.58$, $SD = 4.737$; $t\ (200) = 0.965$, $p = 0.336$, two-tailed). The size of the variance in the averages (mean difference = 5.19% CI: -0.146 to 0.427) was negligible (Cohen's d = -0.141).

Based on the outcomes of the information, this hypothesis is rejected, as there is no noteworthy proof that there are inconsistencies in the averages of the different genders for numeric reasoning ability.

4.3.4 HYPOTHESIS 3

H_3 propose that: There is a noteworthy variance in the average test totals achieved between diverse ethnic groups for verbal reasoning ability.

For this hypothesis, we compare the average totals for multiple groups, because the data has four race categories, therefore we use the analysis of variance (ANOVA). The one-way analysis of variance comprises an independent variable with different levels matching different groups or conditions. In this research, we compare the different race groups' achievements on the verbal reasoning ability subtest, which have one factor (race) with four categories (White, Black, Coloured and Asian). ANOVA compares the variability in test scores between the diverse groups with the inconsistencies within each of the groups. An F ratio is calculated and indicates if there are more inconsistencies among the groups than there are amid each group, but does not indicate which groups differs. If the result is significant, the null hypothesis can be rejected (Pallant, 2011). The findings are reported next.

Table 4.7

ANOVA analysis for Verbal Reasoning subtest

Descriptives

VERBAL

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum	Between-Component Variance
1 (Black)		23	13.43	7.525	1.569	10.18	16.69	3	30	
2 (Coloured)		51	17.96	6.412	.898	16.16	19.76	6	34	
3 (Whites)		118	19.54	7.011	.645	18.26	20.82	3	34	
4 (Asians)		10	19.00	7.916	2.503	13.34	24.66	2	31	
Total		202	18.42	7.176	.505	17.43	19.42	2	34	
Model	Fixed Effects			6.969	.490	17.45	19.39			
	Random Effects				1.534	13.54	23.30			5.028

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
VERBAL	Based on Mean	.173	3	198	.915
	Based on Median	.216	3	198	.885
	Based on Median and with adjusted df	.216	3	192.030	.885
	Based on trimmed mean	.178	3	198	.911

ANOVA

VERBAL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	734.371	3	244.790	5.040	.002
Within Groups	9616.862	198	48.570		
Total	10351.233	201			

Robust Tests of Equality of Means

VERBAL

	Statistic[a]	df1	df2	Sig.
Welch	4.302	3	33.012	.011
Brown-Forsythe	4.564	3	48.755	.007

a. Asymptotically F distributed.

$$\text{Eta squared} = \frac{\text{Sum of squares between groups}}{\text{Total sum of squares}}$$

$$= \frac{(734{,}371)}{(9616,\ 862)}$$

$$= 0{,}076$$

Table 4.7 shows the results of a one-way between-groups ANOVA performed to examine the influence of race on the achievement of verbal reasoning ability, as reported by the CRTB2 Test. Members were divided into four groups in keeping with their race (Group1: Black; Group 2: Coloured; Group 3: White and Group 4: Asian). There was a statistically significant difference at the $p < 0.05$ level amid verbal reasoning scores for the four race groups: $F (3, 198) = 5.040$, $p = 0.002$. Regardless of attaining statistical significance, the real difference in average scores between the groups was not large. The impact of the sample size was calculated utilizing the eta squared, which was 0.076. Post-hoc evaluations using the Tukey HSD test showed that the average score for Group 1 (M = 13.43, SD = 7.525) was meaningfully different from Group 3 (M = 19.54, SD = 7.011). Group 2 (M = 17.96, SD= 6.412) and Group 4 (M = 19.00, SD= 7.916) did not differ meaningfully from either Group 1 or Group 3.

Based on the findings of the data, this hypothesis is accepted.

4.3.5 HYPOTHESIS 4

H_4 suggests that: A noteworthy difference in the average test scores achieved among the different race groups for numeric reasoning ability. The findings are reported next.

Tables 4.8

ANOVA analysis for Numeric Reasoning subtest

Descriptives

NUMERIC

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum	Between-Component Variance
1 (Black)		23	9.70	5.243	1.093	7.43	11.96	2	20	
2 (Coloured)		51	12.20	4.617	.647	10.90	13.49	2	20	
3 (White)		118	13.88	5.020	.462	12.97	14.80	3	25	
4 (Asian)		10	15.20	6.408	2.026	10.62	19.78	5	23	
Total		202	13.04	5.189	.365	12.32	13.76	2	25	
Model	Fixed Effects			5.020	.353	12.35	13.74			
	Random Effects				1.173	9.31	16.78			2.974

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
NUMERIC	Based on Mean	.894	3	198	.445
	Based on Median	.779	3	198	.507
	Based on Median and with adjusted df	.779	3	189.675	.507
	Based on trimmed mean	.878	3	198	.453

ANOVA

NUMERIC

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	423.751	3	141.250	5.606	.001
Within Groups	4988.848	198	25.196		
Total	5412.599	201			

Robust Tests of Equality of Means

NUMERIC

	Statistic[a]	df1	df2	Sig.
Welch	4.963	3	32.740	.006
Brown-Forsythe	4.718	3	39.876	.007

a. Asymptotically F distributed.

$$\text{Eta squared} = \frac{\text{Sum of squares between groups}}{\text{Total sum of squares}}$$

$$= \frac{(423{,}751)}{(4988,\ 848)}$$

$$= 0{,}085$$

Table 4.8 shows the results of a one-way between-groups ANOVA performed to examine the influence of race on the achievement of verbal reasoning ability, as reported by the CRTB2 Test. Members were divided into four groups in keeping with their race (Group1: Black; Group 2: Coloured; Group 3: White and Group 4: Asian). Statistical significance at the $p < 0.05$ level was found in verbal reasoning totals for the four race groups: $F (3, 198) = 5.606$, $p = 0.001$. Regardless of the statistical significance, the real difference in average scores across the groups was not large. The power of the sample size was calculated using eta squared, which was 0.085. Post-hoc evaluations using the Tukey HSD test showed that the average score for Group 1 (M = 9.70, SD = 5.243) was meaningfully different from Group 3 (M = 13.88, SD = 5.020). Group 2 (M = 12.20, SD= 4.617) and Group 4 (M = 15.20, SD= 6.408) did not differ meaningfully from either Group 1 or Group 3.

Based on the findings of the data, this hypothesis is accepted.

4.3.6 HYPOTHESIS 5

H_5 suggests that: A relationship between age and both verbal and numeric reasoning ability exists.

Although the data of this research did not have age included for all the respondents, 57 respondents included their age and completed both the numeric and verbal subtests, therefore a correlation study is included to establish whether a relationship exists between the age of the respondents and the verbal and numeric reasoning ability test scores achieved. The Correlation coefficient (r) obtained by conducting the Pearson's product-moment correlation provides a quantitative synopsis of the trend and the power of the linear relationship between two variables. Correlations range from –1 to +1. The mathematical symbol in front indicates whether there is a positive relationship (an increase in one variable increases the other) or an opposing relationship (an increase in one variable decreases the other). The magnitude of the absolute value, irrespective of the symbol, gives insight into how strong the relationship is. A perfect correlation of 1 or –1 shows that the value of one variable can accurately be calculated by knowing the value of the other variable, whereas a correlation of 0 shows that variables are unrelated. The statistical significance of r is also provided (Pallant, 2011). The findings are reported next.

Table 4.9

Correlation analysis for Numeric and Verbal Reasoning subtest

Correlations

		NUMERIC	VERBAL	AGE
NUMERIC	Pearson Correlation	1	.537**	-.215
	Sig. (2-tailed)		.000	.108
	N	202	202	57
VERBAL	Pearson Correlation	.537**	1	-.114
	Sig. (2-tailed)	.000		.398
	N	202	202	57
AGE	Pearson Correlation	-.215	-.114	1
	Sig. (2-tailed)	.108	.398	
	N	57	57	57

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.9 presents the outcomes of the relationship between the age of the members and the reported totals of both the verbal and numeric reasoning subtests, as calculated by the CRTB2 test that was examined using the Pearson's product-moment correlation coefficient. Initial investigations were conducted to guarantee no violation existed regarding the rules of normality, linearity, and the homogeneity of variance. There was a strong, positive relationship between verbal reasoning ability, r = 0.537, and numeric reasoning ability, r = 0.537, n = 57, p < .0005, with older age groups scoring lower on the verbal and numeric subtests. Both correlation coefficients are significant at the 90 percent confidence level.

The results, therefore, indicate that hypothesis can be accepted. There is enough proof to propose that a relationship exist and that older people scored lower on both verbal and numeric reasoning test.

4.4 CONCLUSION

This chapter reported the outcomes of processing the data. The results were displayed in tables and graphically and briefly explained what the different values were. The results reported statistically significant findings for the race, gender, and age groups that scored differently on the CRTB2 test battery. In Chapter five, the researcher will discuss the results and the outcome of the study and the confines of the research. Recommendations about the study will be given for upcoming studies and to the organisation and will be explored during the final section.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 INTRODUCTION

The primary object of this research is to analyse the secondary data from a retail company and explore the various relationships to facilitate understanding if any variances exist among the groups that completed the CRTB2 test battery. In this section of the research, the findings are discussed by unpacking the findings of the results of the previous chapter and comparing and relating the views of other existing research findings and theoretical concepts. This chapter concludes with recommendations for future research and for the organisation.

5.2. DISCUSSION OF RESULTS

5.2.1 DESCRIPTIVE STATISTICS

The results of this research study reported the descriptive statistics first, and a summary of the biographical information is provided. When analysing and comparing the data and the graphs for the male and female participants on the verbal reasoning subtest, it can be seen that the mean for both the males and females are very similar, with females having a slightly higher mean than that of the males, and that both genders' means are very close to the population mean. However, when comparing the average totals of the various race groups on verbal reasoning, a much lower score is achieved by the Black race group, which is also lower than the population average. A similar pattern is noticed with the numeric reasoning subtest where the mean scores for males and females are reasonably similar, however, differences were noticed across the racial groups, with the Black group scoring the lowest amongst the race groups and lower than the population mean score.

From the data it can be observed that midpoints of the gender groups are similar to each other for both the verbal and numeric reasoning subtests and that there is not as much difference between the midpoints of the race groups, but the Black race group has achieved markedly lower midpoints for both the verbal and numeric reasoning subtests than the other race groups.

For both the verbal reasoning and numeric reasoning subtests, the mode for females was much lower than that of males. From the values reported, it can be seen that there is markedly lower scoring achieved by the Black race group when compared to the other race groups for both the numeric and verbal reasoning subtests.

The Kolmogorov-Smirnov statistic allows for the evaluation of the regularity of the spread of test totals to provide a more accurate interpretation of whether the data is normally distributed or dispersed (Pallant, 2011). The Kolmogorov-Smirnov statistic for the different gender, age and race groups are more than the 0.05 statistical significant level, which indicates that the distributions are normal. The bell-shaped curve of the histogram reflects this. These tests were conducted to inspect the variables for any irregularities in the assumptions that underpin the statistical techniques that were used and to answer the research questions.

The next subsections will address the results generated by the inferential statistical investigation that was conducted regarding the hypotheses, before discussing the limitations of the results, recommendations, and conclusion.

5.2.2 INFERENTIAL STATISTICS

5.2.2.1 GENDER DIFFERENCE EFFECT ON VERBAL AND NUMERIC REASONING SUBTESTS

The first two hypotheses proposed that there is a noteworthy difference in the average test totals achieved across different gender groups for verbal and numeric reasoning ability. This research study used T-tests to equate the average scores of the gender groups, which consisted of 128 males and 74 females on each subtest. The findings confirmed no noteworthy differences in the average totals for men and women on either verbal or numeric subtests and the size of the variances in the averages, as seen by Cohen's d statistic, confirmed that the difference was negligible. Based on these findings, the hypotheses were rejected, as no significant differences were found.

Research conducted by Spelke (2005) found that research offers indications that mathematical and scientific skills develop from a collection of biologically constructed reasoning capabilities, are shared by men and women. These capabilities allow men and women to cultivate an equal flair for mathematics and science. This study, therefore, concurs with the findings of this research.

Ramírez-Uclés and Ramírez-Uclés (2020) conducted a study on students that were chosen to join a mathematical talent stimulus project after completing a difficult problem-solving test. They administered two instruments: the Differential Aptitude Tests-Space Relations (DAT-SR) and the Primary Mental Abilities-Space Relations (PMA-SR) test. The findings were that no measure considered showed differences across the males and females in either test. Similarly, this study also concurs with the findings of this research.

A validation study that was conducted on the GRT2 test in South Africa by Abrahams et al., (2012) found significant differences across gender groups, with men achieving higher scores, and that the test appears to exhibit a predictive bias for language, race, and gender, specifically for the numeric reasoning subtest. It is recommended that care is taken when interpreting test results due to the differences across gender groups and language groups. These findings are thus contrary to the findings of this research in terms of gender differences.

Steegh, Höffler, Keller, and Parchmann (2019) led the research which evaluated whether mathematics and science competitions assist with the impartial promotion of female and male interests. Achievement by men and women at fairs were comparable, but men outperformed women candidates at the Olympiads with the least variances seen in the biology Olympiad.

Arendse (2021) conducted a validation study of the newly developed English Comprehension Test battery to investigate gender discrimination in cognitive testing. Not only did the findings of this research contradict the belief held that females are better at the verbal task than males, but in addition, it highlighted that language caused differential performance in assessments in the multi-cultural setting of South Africa.

The study conducted on the achievement of Grade Twelve learners from the Western Cape Province by Letsoalo (2018), found that female learners scored significantly higher, at least 0.45% more than male learners. The findings from the study therefore suggests that females achieve higher academic results in the Western Cape Province than their male counterparts. According to this study even though the differences are not large and significant, females scored higher than males.

A meta-analysis conducted by Voyer and Voyer (2014) on gender differences in scholastic achievement produced findings that a female advantage exists, and was most pronounced for language courses and least for math courses. This study showed the existence of a steady female advantage in scholastic achievement.

A study conducted on Grade 9 learners at a private school in the Philippines by Mingoa and Abocejo (2021), found a robust relationship between reading understanding and science achievement of females and males with reasonable direct correlations noted among the learners' academic aptitude related to verbal skills, mathematical skills, logical-reasoning skills, and visual manipulative skills. Science achievement statistically differed between males and females and the female students' achievements were superior to their male colleagues in science.

Stoet and Geary (2013) analysed one decade of data that included the maths and reading accomplishments of almost 1.5 million 15-year-olds across 75 countries. The results were collected by the Programme for International Student Assessment (PISA). Their study report that across nations, boys fared better than girls in maths but were worse than girls in reading. The results between the different gender groups in reading were three times larger for mathematics. Furthermore, results between the different groups in mathematics were steady and intensely inversely correlated with the reading results of the different genders.

Petersen (2018) conducted a meta-analysis of gender differences in verbal achievement on the United States State Performance Assessments. The data represents more than 10 million United States scholars between Grades 3 and 11. The outcomes show a minimal gender difference that favors girls for verbal achievement overall (d = 0.29).

A published doctoral thesis by Reilly (2019) conducted a meta-analysis across a broad range of gender differences in cognitive ability. The findings of research in similar areas of gender differences on aptitude test varies, with some research finding that females score higher, other research findings holding that males score higher, and research suggesting that the gap is narrowing. It is notable that the generalisability of most of the research is limited due to sample size and concludes that eliminating gender differences still has a long way to go. More studies are required to understand the things that contribute to these differences.

Even though no significant differences are noted between genders from this research, what is demonstrated by all these studies is that often, either some differences are found and that either a male or female advantage often exists. Despite the findings of this research, the researcher agrees with the sentiment of Reily (2019), which is that whether gender scores are higher or lower on a test should not be the focus. The focus should be that if these differences do exist, the need for more generalisable research is amplified in this area to address and eliminate the causes of these differences.

5.2.2.2 RACE DIFFERENCE EFFECT ON VERBAL AND NUMERIC REASONING SUBTESTS

The next two hypotheses proposed that a noteworthy variance in the average test totals is achieved between the different race groups for verbal and numeric reasoning ability. This study used the one-way between-groups ANOVA to compare the influence of race on the performance of both numeric and verbal reasoning ability subtests and conducts the comparison for more than 2 groups. The findings were that there was a statistically significant difference at the $p < 0.05$ level for race groups on both the verbal and numeric subtests, that the average totals were meaningfully different between the Black and White groups, and the other race groups showed negligible differences on both the verbal and numeric subtests.

We, therefore, accept the hypothesis, which is that there are noteworthy differences in the average test totals for different race groups on both the verbal and numerical reasoning subtests.

The finding concurs with similar research that was conducted in South Africa for an aptitude test. A study conducted by Abrahams et al., (2012) found using the General Reasoning Test Battery subtests that English-speaking persons had a lead over non-English-speaking persons. Women were disadvantaged concerning their numerical reasoning test totals and lastly, that evident mean score variances were observed between the Black and other groups. However, due to limitations such as sample size, a definite sanction about the bias and fairness of the GRT2 was not given for that specific study.

Rushton and Jensen (2005) conducted a literature study that covered 30 years of ethnic group differences in cognitive skills. It found that a 1,1 standard deviation difference in the “IQ” scores achieved between Black and White Americans are not in empirical dispute. The question remains whether these inconsistencies in "IQ" scores are linked to social, economic, and cultural factors, or other factors that contribute to these differences. Since the question is so complex, it is unlikely that one study will be able to answer the question but instead review the independent evidence produced by the various studies conducted (Rushton & Jensen, 2005).

Another recent study by Cockcroft et al., (2015) compared the performance of multilingual South African university students, which were classified in the low socioeconomic class, to the performance of British university students, who were monolingual and classified in the high socioeconomic class on the WAIS-III test battery. The WAIS-III test battery is the most frequently used intelligence test in South Africa. This study was explicitly conducted because the socioeconomic conditions in South Africa have changed over the past 20 years after the abolishment of apartheid. The primary outcome of this study was that most subsets of this test battery hold cross-cultural biases and challenge the opinion that the non-verbal test is culturally fairer than the verbal test. This study further demonstrates the disparities in South Africa are not only attributed to race, but also other conditions, such as socioeconomic conditions.

In South Africa, race is one of the variables deemed to affect performance on cognitive ability tests, but variables such as language, culture, and socioeconomic conditions are also deemed to contribute to these differences. A recent study by Cockcroft, Bloch, and Moolla (2016) assessed the verbal functioning of school-beginning children in South Africa from diverse socioeconomic backgrounds. Even though this research study does not look at the adult population of South Africa, an outcome of this study was that socioeconomic status accounts for considerable variance in the vocabulary measures amongst the different race groups.

South African studies conducted using various psychometric tests by Lynn and Owen (1994); Lynn and Vanhanen (2002); Owen (1992); Skuy, Schutte, Fridjhon, and O'Carroll (2001) found that the mean scores of these studies showed lower mean scores than the studies that were done in the United States of America. It observed that the size of this difference was between 1 and 2 standard deviations. Furthermore, it found that studies investigating the difference of "***g***" between groups found a 2 standard deviation difference between Blacks and Whites (Rushton & Jensen, 2005).

The unanswered questions and debates of many studies of whether these differences in "IQ" scores or intelligence between different race groups are hereditary or related to genetics, social, economic, and cultural factors remain a debate in social science. Some opinions ascribe these differences to the differences in opportunity structures that some are afforded and others are excluded from. Access to these opportunity structures does not guarantee upward-mobility but increases the likelihood that pathways may be presented. People from wealthier backgrounds are more likely to be afforded educational opportunities of their choice than low-income groups. In South Africa, the past disparities and unequal opportunities, such as the quality of education in lower-income areas, different languages, and disparities could all likely contribute to these differences existing. Most of these research studies conducted in South Africa do not conclusively attribute a specific factor that accounts for these differences between race groups or gender groups due to the complexity of isolating specific factors (Arendse, 2021; Cockcroft et al., 2015; Cockcroft et al., 2016; Krieg, 2014).

5.2.2.3. AGE DIFFERENCE EFFECT ON NUMERIC AND VERBAL REASONING SUBTESTS

The final hypothesis proposed that there is a relationship between age and verbal and numeric reasoning ability. This study conducted a Pearson's product-moment correlation analysis and found a robust, positive correlation between verbal reasoning ($r = 0.537$) and numeric reasoning ($r = 0.537$). The older age groups scored lower on the verbal and numeric subtests. Both correlation coefficients are significant at the 90 percent confidence level ($p < 0.005$). However, the sample size for this test was limited ($n=57$) because not all the ages of the participants were captured.

The positive, mathematical symbol of the correlation coefficient refers only to the trend of the relationship, not the power of the relationship. The power of the correlation is determined by how close the correlation value is to 1. This study suggests that a powerful relationship exists between age and the mean scores of both the verbal and numeric ability subtests, with older people scoring lower on both subtests. Furthermore, the level of statistical significance does not indicate how strongly the two variables are associated, but rather it indicates the level of confidence we could have in the results, which is 90% in this case (Pallant, 2011). For these reasons, we, therefore, accepted this hypothesis as the results indicate that older people scored lower on both verbal and numeric reasoning subtests.

A few studies have been conducted in academic settings such as at universities or high schools to look at the effective predictability of various aptitude test used for selection, but not necessarily explored the impact that age have. A study by Ebenuwa-Okoh (2010) examined the influence of age on academic performance among undergraduates but determined that age was not a noteworthy forecaster of academic success.

A study conducted by Munukka, Koivunen, von Bonsdorff, Sipila, Portegijs, Ruoppila, and Rantanen (2021) monitored differences in the cognitive performance of two cohorts over 28 years. The study reported statistically significant differences in the cognitive ability tests and that the later-born cohort performed superior in all the tested outcomes. They concluded that their study provides robust proof that cognitive performance is superior in the more recent cohorts of older people than their counterparts that were tested 28 years earlier.

There is not enough research that specifically investigates the effect of age on cognitive test performance. Even though this scarceness makes it challenging to support the results of the current study, it is worthwhile to note that similarities exist between the findings of this research and that of the cohort study. This study also found that older people scored lower, even though the context of these two studies differs. Much of the age-related cognitive studies were conducted in the medical sector, but those studies predominantly relate to neuroscience and will not necessarily illustrate or negate whether age differences have an impact on cognitive test performance. Much more research is required in this area.

5.2.2.4 SUMMARY OF RESULTS

This research presented the results of a series of statistical analyses that were conducted on the mean scores that were achieved on both the verbal and numerical reasoning subtests of the CRTB2 test battery. Other research could not be sourced that examined the statistical differences of the mean scores on the CRTB2 test battery; therefore it remained unexplored until now. No significant differences were found in the mean scores of the gender groups on either the verbal or numeric subtests. Probably the most significant contribution of this research is that a statistically significant difference was found between the race groups on both the verbal and numeric subtests, with the Black group scoring significantly lower than the White and other race groups on both the verbal and numeric subtests. A strong correlation was found between the different age groups, with the older age groups scoring significantly lower on the verbal and numeric subtests.

The outcomes of this study have repercussions for both research and practice of personnel selection and psychometric testing. This research implies that personnel practitioners should examine the selection procedures that they use, particularly, scrutinizing the validity, reliability, and bias of the test instruments, which they opt to include in their recruitment strategies. The outcomes of the research indicate that the CRTB2 is not a useful selection tool in the Retail industry, because of the significant differences in the average scores of the different races and age groups. The current study also highlighted the tension in South Africa between achieving equity in the workplace and increasing productivity. This debate has importance when it comes to the use of psychological tests as part of selection and recruitment strategies, as well as using and creating appropriate norms groups, to obtain a shortlist of candidates that are considered for development opportunities.

Despite the limitations of this research, which will be discussed next with recommendations that are given; it is important to consider the results obtained in this study. Therefore, they should be regarded as provisional, until additional studies can be added. In closing, the researcher will also outline key areas where more empirical research is needed, as this research highlights and provides a compelling example of much more research needed. To better understand and disrupt inequality in psychological testing, and scrutinizing the various selection strategies used, and covers a broad stakeholder group. It also amplifies the demand for more consistent recruitment standards.

5.3. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

5.3.1 LIMITATIONS OF THIS RESEARCH STUDY

Although this research achieved its goals and objectives and will add to the existing knowledge and information in this research area, a few limitations were encountered.

The sample consisted of the total raw scores of the secondary data and the size of this research depended on the secondary data that was available and was specific to the retailer based in

Western Cape. Therefore, the extent to which the findings of this study can be generalised should be limited to similar populations.

The geographic region of the sample is reported as Western Cape because the head office of the retail company is based in this province of South Africa. However, the individuals that participated in this psychometric testing may have hailed from other provinces, as the retail company operates across South Africa, but the physical residential addresses of the participants were not recorded as part of the database information. It restricts the generalisability of this research, even though a large sample size was used.

The research study used convenience sampling as this data was easily available to the researcher but not all the data points had age groups, therefore, the findings related to age should only be limited to this population or similar sample groups. Furthermore, due to the convenience sample used, the researcher had no control for the group categorisation of the data as these were inherent due to the reporting requirements in terms of the labour laws of South Africa.

A limitation of the data is that not all the job titles or job grades were available for all the data points as well as not all the age groups of the respondents were captured. Therefore, these parameters are excluded in the data analysis performed on the sample, and only information where respondents completed both the verbal and numeric reasoning subtests of the CRTB2 test battery are used to conduct the data analysis.

Due to available information provided by the company, the data for this study reported only the overall raw scores achieved on each of the verbal and numeric reasoning subtests and did not include the scores per item for both tests. Raw scores are the added total of the correctly answered responses that an individual achieved on the test before any statistical conversions are done. These scores are limited and do not provide an accurate comparative reflection between individuals because the level of difficulty for each item of the test may differ. Therefore, caution should be used and this should be considered when interpreting and generalising the results. To have a more generalisable view, it would require the use of each item score so that it can be seen if there are any significant differences on any specific items in each of the tests (Arendse, 2021; Salkind, 2010).

Due to the sample nature, the statistical analysis conducted in this study was limited to what the data allowed. More complex and other statistical analyses, such as factor analysis could not be conducted, as the raw scores available are the overall scores and not the score per item.

Other research could not be sourced that used the CRTB2 test instrument. Therefore, direct statistical comparisons for significant results were not directly comparable with other South African studies, but with it could at least be compared to those studies that used similar test instruments.

5.3.2 RECOMMENDATIONS FOR THE FUTURE

Notwithstanding the observed limitations of this research, it should not detract from the quality of these research findings and the contribution that this research is adding to the knowledge repository and this field. It is much needed not only in South Africa but can also assist international researchers in this area of research. It would add value if additional studies can be added that specifically investigates the CRTB2 test battery to determine whether a consensus of the findings and views can be achieved across different industries and regions.

Various reasons could contribute to the differences in mean scores among different groups. Several researchers have pondered the issue of the existence of a relationship between the quality of education and the level of achievement on cognitive tests. It appears that there is a necessity for further research on the fairness of local selection practices and the various factors that shape the differences in average scores. Other studies demonstrated that race, education, language, and understanding of English were the primary factors influencing test achievement (Arendse, 2021; Cockcroft et al., 2015; Cockcroft et al., 2016; Foxcroft & Aston, 2006; Owen, 1992; Skuy et al., 2001). The demand for more decisive advice regarding the fairness of using the GRT2 within the selection process should be investigated (Abrahams et al., 2012). The same should be done for the CRTB2 test to understand what causes these mean differences, and will simultaneously help to determine the fairness in local selection processes.

Due to the outcomes of this study, which indicate a significant difference, more research could explore the cause of these differences. It will be especially useful if these studies are conducted on South African audiences and include the age of the participants so that more generalisable evidence can be produced. This will be helpful to not only understand whether mean differences exist, but will be helpful to streamline as well as standardise selection policies and reconsider the recommended tests that could be used to remove any possible biases that may exist.

The researcher had to rely on the statistical information supplied in the test manual of the CRTB2 test. More published research that uses the CRTB2 test in South Africa will be welcome and will assist the literature reviews where similar test instruments were used. More sectors that are using the CRTB2 test for selection should publish research that may have been conducted, as it will be helpful to compare if similar findings or trends in the data can be observed.

Statistically significant results should give the organisation pause to interrogate the validation study that was undertaken by the test publisher, especially since the practice of using the recommended cut-off scores is used by the organisation as part of its selection strategy. Alternatively, a separate validation study should be undertaken to ensure that the test instrument complies with the labour law requirements of South Africa, which need tests to be valid, reliable, and fair. If possible, a factor analysis of the items of the CRTB2 test and a confirmatory reliability analysis of at least the race group differences should be conducted, and endorse whether these factor structures agree with the predominant factor structure and the repeatability of the CRTB2 test totals.

Using cut-off scores is similar to benchmarking the achievements of the candidates and comparing them to a control group. The concern of norming or re-norming psychological tests is a contentious issue because it may exclude or discriminate against certain groups. It necessitates vigilant deliberation of all the relevant factors and should consider including influences such as age, education, gender, and race. These should be done during and in conjunction with standardisation and validation studies during test development, and the outcomes should be encompassed in the test manual (Arendse, 2021; Foxcroft, 2011; Paterson & Uys, 2005).

Given the impact of development initiatives in South African organisations, selection and the range of functions linked to using psychometric tests can be useful if organisations can optimise the predictive worth of testing that is embraced (Carless, 2009).

This study concurs that it is important that tests are meticulously investigated for their cross-cultural relevance before using them in multicultural settings. Therefore, as a suggestion for future research, it is recommended that new studies should examine the usage and relevance of the type of selection test batteries that are used by different organisations, and comparisons are made across different industries so that it can assist with best practice recommendations for companies. Companies should also take accountability to review and decide on the acceptability of the selection methods that they use.

The selection policies of companies should be reviewed and emphasis must be placed on the inclusion of the results of research that the company has undertaken that focused on the perceived fairness of its selection tools that they elect to use. The apparent fairness of the selection decisions regarding equal opportunity and impartiality should also be pondered. In South Africa, this could easily be incorporated as part of current legislation requirements and address the need that all organisations should be able to provide evidence of fair selection practices. Employment equity reports could also include this as a requirement for reporting the organisation's equity results.

A larger sample that is accurately reflecting the geographic origin of the participants will allow for a larger geographical spread and could be more accurate and generalisable, especially for future research.

This research calls attention to the constant call for nationwide engagement about a possible standardised selection process and constant research to gauge real practices in industries (Arendse, 2021; Clay, 2017; Foxcroft, 2011; Van der Merwe, 2002). Difficult conversations will need all stakeholders to review current practices and objectives, which include the explicit process by which selection committees make their decisions to accomplish the best conceivable consequences for applicants of all groups.

Considering the restricted amount of studies, which show validity evidence for predicting success for the most recent version of the CRTB2 test instrument as a selection tool, there is still relevance for studies that can contribute extra insight into the predictive validity of the instrument. It may bolster the usefulness when undertaking selection decisions for development opportunities. Holistically, the importance of these findings deserves future research studies to clarify the utility of the CRTB2 instrument as a selection tool for the development of diverse groups.

5.4. CONCLUSION

Progress was made in decreasing inequality, which includes gender disparities and racial or ethnic disparities; yet these inequalities remain a concern due to the past legacy, especially in South Africa or countries where these still seem apparent. This is due to a legacy of discrimination against different groups and the consequences are far-reaching if this perception remains. Researchers and registered psychologists need to collectively commit to understanding and counteracting this perception and be open to embracing new ideas about making psychological testing desirable and more acceptable. It may need a necessary mindset change to mitigate the effects of this perception.

Summarizing the results of this research, it can be surmised that: (1) No noteworthy mean differences were observed across the different gender groups on either the verbal or numeric subtests of the CRTB2 test; (2) A statistically significant difference was proven for the average scores of the ethnic groups on both the verbal and numeric subtests of the CRTB2 test, with the Black group scoring significantly lower than the White and other race groups; (3) A strong correlation was established for the average scores of the different age groups, with the older people scoring meaningfully lower on the verbal and numeric subtest of the CRTB2 test.

The study is finalized by remarking that while more studies are crucial, the current proof of this study principally supports the suggestion that aptitude tests that are used for selection in South Africa should be reviewed for inclusion in selection test procedures across industries and regions. It will relay relevant concerns about choosing more diverse, deserving candidates. The report highlighted that it is inappropriate for those involved in selection decisions to use standardised tests in isolation from other selection aids and to ensure that assessment tools have proven valid predictors of the aptitudes that they assess.

REFERENCES

Abrahams, F. (1996). *The cross-cultural comparability of the 16 Personality Factor inventory (16 PF),* University of South Africa, Pretoria, <http://hdl.handle.net/10500/16756>

Abrahams, F., Friedrich, C., & Tredoux, N. (2012). Selection of postgraduate students in a South African management programme. *Industry and Higher Education, 26(2),* 139–149. https://doi.org/10.5367/ihe.2012.0086

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). *The Impact of the Cattell–Horn–Carroll Theory on Test Development and Interpretation of Cognitive and Academic Abilities.* Contemporary Intellectual Assessment, 2nd Edition: Theories, Tests, and Issues. Guilford Publications.

Anazia, I. U. (2019). Quantitative and Verbal Aptitudes as Predictors of Senior Secondary School Students' Performance in Economics. *IAFOR Journal of Education, 7(1),* 7–18. Advance online publication. https://doi.org/10.22492/ije.7.1.01

Arendse, D. E. (2021). Addressing gender discrimination in cognitive assessment using the English Comprehension Test. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 47(0),* a1776. https://doi.org/10.4102/sajip. v47i0.1776

Babbie, E. R. (2021). *The practice of Social Research* (15th ed.). Cengage Learning.

Baron, J., & Norman, F. (1992). *SATs Achievement Tests and High School Class Rank as Predictors of College Performance*. University of Pennsylvania. Educational and Psychological Measurement Inc. https://doi.org/10.1177/0013164492052004029

Bedell, B., Van Eeden, R., & Van Staden, F. (1999). Culture as moderator variable in psychological test performance: Issues and trends in South Africa. *SA Journal of Industrial Psychology, 25(3),* 1–7. https://doi.org/10.4102/sajip.v25i3.681

Bonamente, M. (2013). *Statistics and Analysis of Scientific Data*. Springer Science and Business Media, ISSN18684521. Advance online publication. https://doi.org/10.1007/978-1-4614-7984-0

Bryman, A., & Bell, E. (2014). *Research methodology: Business and management contexts*. Oxford University Press Southern Africa.

Carless, S. A. (2009). Psychological testing for selection purposes: A guide to evidence-based practice for human resource professionals. *International Journal of Human Resource Management, 20(12)*, 2517–2532. https://doi.org/10.1080/09585190903363821

Carroll, J. B. (2005). *The three-stratum theory of cognitive abilities*. (In Flanagan, D.P. & Harrison, P.L. eds. Contemporary intellectual assessment: theories, tests and issues. 2nd Edition. New York: Guilford Press. p.69 - 76.).

Clay, R. A. (2017). *Decolonizing psychology in South Africa.* Monitor on Psychology, 48(10). https://www.apa.org/monitor/2017/11/decolonizing-psychology

Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). *A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales* 3rd Edition (WAIS-III). Front. Psychol. 6:297.https://doi.org/10.3389/fpsyg.2015.00297

Cockcroft, K., Bloch, L., & Moolla, A. (2016). Assessing verbal functioning in South African school beginners from diverse socioeconomic backgrounds: A comparison between verbal working memory and vocabulary measures. *Education as Change, 20(1)*, 199–215. www.educationaschange.co.za. https://doi.org/10.17159/1947-9417/2016/559

Cormier, D. C., Bulut, O., Mc Grew, K. S., & Singh, D. (2017). Exploring the relations between Catell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement. *Applied Cognitive Psychology, 31*, 530–538.https://doi.org/10.1002/acp.3350

De Goede, J., & Theron, C. (2010). An investigation into the internal structure of the learning potential construct as measured by the APIL-B test battery. *Journal of the Southern African Institute for Management Scientists, 19(4)*, 30–55.

Department of Education. (2001). *Education White Paper 6. Special needs education: Building an inclusive education and training system*. Pretoria: Department of Education.

Donald, F., Thatcher, A., & Milner, K. (2014). Psychological assessment for redress in South African organisations: Is it just? *South African Journal of Psychology*. *Suid-Afrikaanse Tydskrif vir Sielkunde, 44(3)*, 333–349. sap.sagepub.com https://doi.org/10.1177/0081246314535685

Dupper, O., Bhoola, U., & Garbers, C. (2009). *Understanding the employment equity act.* Juta publication.

Ebenuwa-Okoh, E. E. (2010). Influence of age, financial status and gender on academic performance among undergraduates. *The Journal of Psychology, 1(2)*, 99–103. https://doi.org/10.1080/09764224.2010.11885451

Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). Sage Publications.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. Wiley.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2002). *The achievement test desk reference* (ADTR): Comprehensive assessment and learning disabilities. Boston: Allyn & Bacon.

Foxcroft, C. D. (2011). Ethical Issues Related to Psychological Testing in Africa: What I Have Learned (So Far). *Online Readings in Psychology and Culture,2(2).* Advance online publication. https://doi.org/10.9707/2307-0919.1022

Foxcroft, C. D., & Aston, S. (2006). Critically examining language bias in the South African adaptation of the WAIS-III. *SA Journal of Industrial Psychology, 32(4)*, 97–101. https://doi.org/10.4102/sajip.v32i4.243

Foxcroft, C. D., & Roodt, G. (2018). *An introduction to psychological assessment in the South African context* (5th ed.). Oxford University Press.

Gardner, H. (1993). Frames of mind: the theory of multiple intelligences (2nd ed.). Basic Books.

Gardner, H. (1999). Assessment in context. *Learners, learning and assessment*, *7*, 90-117.

Gardner, J. E. (2012). *Assessment and learning*. Sage Publications. https://doi.org/10.4135/9781446250808

Goldstein, G., Allen, D. N., & De Luca, J. (2019). *Handbook of Psychological Assessment* (4th ed.). Academic Press.

Goldstein, H. W., Pulakos, E. D., Passmore, J., & Semedo, C. (2017). *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention*. 1st Edition. Wiley, J., & Sons Ltd. https://doi.org/10.1002/9781118972472

Görgens-Ekermans, G., & Herbert, M. (2013). Psychological capital: Internal and external validity of the Psychological Capital Questionnaire (PCQ-24) on a South African sample. *SA Journal of Industrial Psychology, 39(2)*. Advance online publication. https://doi.org/10.4102/sajip.v39i2.1131

Health Professions Act, Act 56 of 1974. Health Professions Council of South Africa.

Herrnstein, R. J., & Murray, C. (1994). The bell curve. Free Press.

Howell, D. C. (2013). *Fundemental Statistics for the Behavioral Sciences* (8th ed.). Cengage Learning.

Hox, J. J., & Boeije, H. R. (2005). Data collection primary vs secondary. *Encyclopedia of social measurement, Volume 1*. Elsevier Publication. https://doi.org/10.1016/B0-12-369398-5/00041-4

Huysamen, G. K. (2002). The relevance of the new APA standards for educational and psychological testing for employment testing in South Africa. South African Journal of

Psychology. *Suid-Afrikaanse Tydskrif vir Sielkunde*, *32(2)*, 26–33. https://doi.org/10.1177/008124630203200203

James, L., Jacobs, K. E. & Roodenburg, J. (2015). *Adoption of the Cattell-Horn-Carroll model of cognitive abilities by Australian Psychologist*, 50, 194- 202. https://doi.org/10.1111/ap.12110

Katz, L. J., & Brown, F. C. (2019). *Aptitude and achievement testing*. Handbook of Psychological Assessment. https//doi.org/ https://doi.org/10.1016/B978-0-12-802203-0.00005-5

Kemp, R. (2000). *An investigation into the effects of socio-economic and education factors on WAIS III performance in a stratified South African sample*. Rhodes University Library, BRN 208842b1493003

Krieg, E. J. (2014). *Statistics and Data analysis for Social Science*. 1st Edition. Pearson New International Edition. Pearson Education Limited.

La Grange, L., & Roodt, G. (2001). Personality and cognitive ability as predictors of the job performance of insurance sales people. *SA Journal of Industrial Psychology, 27(3)*, 35–43. https://doi.org/10.4102/sajip.v27i3.23

Laher, S., & Cockcroft, K. (2014). *Psychological assessment in South Africa: Research and applications.* Wits University Press.

Letsoalo, M. E. (2018). Grade 12's overall performance in South Africa's Western Cape Province : An analysis using mixed-effects model. *Journal of Gender [JGIDA]. Information and Development in Africa, 7(2)*, 285–307. https://doi.org/10.31920/2050-4284/2018/v7n2a14

Long, W. (2017). Essence or experience? A new direction for African psychology. *Theory & Psychology*, *27(3)*, 293–312. https://doi.org/10.1177/0959354316686428

Lynn, R. (2006). *Race Differences in Intelligence: An Evolutionary Analysis*. Washington Summit Publishers., https://doi.org/10.1016/j.paid.2005.10.004, ISBN 1-59-368021.

Lynn, R., & Meisenberg, G. (2010). *The average IQ of sub-Saharan Africans: Comments on Wicherts, Dolan, and van der Maas.* Published by Elsevier Inc., https://doi.org/10.1016/j.intell.2009.09.009

Lynn, R., & Owen, K. (1994). Spearman's hypothesis and test score differences between whites, Indians, and blacks in South Africa. *The Journal of General Psychology, 121*, 27–36. https://doi.org/10.1080/00221309.1994.9711170 PMID:8021630

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Praeger.

Mahembe, B. (2014). *The development and empirical evaluation of an extended learning potential structural model*. Stellenbosch University. Retrieved, 20 April 2017, from http://scholar.sun.ac.za

Maloa, F., & Bux, C. (2015). *Psychometric testing as a predictor of academic performance – A South African Experience*. Fifth International Conference on engaged management scholoarship, Baltimore, Marland, September, 10 – 13, 2015.

Macklem, G. L. (1990). *Measuring aptitude. Practical Assessment, Research and Evaluation*. Vol. 2, No 5, http://PAREonline.net/getvn.asp?v=2&n=5.Page. ISSN=1531-7714.

Mangara, B. T., Chiyindiko, E. (2019). *Revisiting the concept of the gifted learners in the South African context.* ICERI 2019 Proceedings, 1638-1644. https://doi.org/10.21125/iceri.2019.0466

Matsepe, D., Cross, M., & Fenyane, S. (2020). Equity in Admissions Policies of Undergraduate Students in Post Democracy in Selected South African Universities. *International Journal of Criminology and Sociology, 9*, 437-445

McGill, R. J., & Dombrowski, S. C. (2019). Critically Reflecting on the Origins, Evolution, and Impact of the Cattell-Horn-Carroll (CHC) Model. *Applied Measurement in Education, 32(3)*, 216–231. https://doi.org/10.1080/08957347.2019.1619561

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference* (ITDR): Gf-Gc crossbattery assessment. Boston: Allyn & Bacon.

Mingoa, J. I., & Abocejo, F. T. (2021). Science performance and scholastic aptitude of grade 9 learners. *European Journal of Education Studies, 8(3)*. Advance online publication. https://doi.org/10.46827/ejes.v8i3.3660

Munhamo, B. T. (2014). *The relationship between individual aptitude test scores and job performance for employees in selected Harare companies*.

Muller, N. R. (1999). *An Investigation into the concurrent validity of selected competency-based assessment centre methods*.

Murphy, R., & Maree, D. J. F. (2006). A review of South African research in the field of dynamic assessment. *South African Journal of Psychology. Suid-Afrikaanse Tydskrif vir Sielkunde, 36(1)*, 168–191. https://doi.org/10.1177/008124630603600110

Myors, B., Lievens, F., Schollaert, E., Van Hoye, G., Cronshaw, S. F., Mladinic, A., Rodriguez, V., Aguinis, H., Steiner, D. D., Rolland, F., Schuller, H., Frintrup, A., Nikolaou, I., & Tomprou, M. (2008). International Perspectives on the Legal Environment for Selection. *Society for Industrial and Organisational Psychology, Vol. 1*, p 206–246.

Munukka, M., Koivunen, K., von Bonsdorff, M., Sipilä, S., Portegijs, E., Ruoppila, I., & Rantanen, T. (2021). *Birth cohort differences in cognitive performance in 75- and 80-year-olds: A comparison of two cohorts over 28 years.* Aging Clinical and Experimental Research, 33, 57–65. https://doi.org/10.1007/s40520-020-01702-0 PMID:32918698

Nieuwoudt, S., Dickie, K. E., Coetsee, C., Engelbrecht, L., & Terblanche, E. (2019). Age-and education-related effects on cognitive functioning in coloured South African women. Neuropsychology, Development, and Cognition. *Section B, Aging, Neuropsychology* *and Cognition, 27(6)*, 1–17. https://doi.org/10.1080/13825585.2019.1598538 PMID:30922209

Nugba, R. M., & Quansah, F. (2021). Standardized Achievement Testing, Aptitude Testing, and Attitude Testing: How Similar or Different are these Concepts in Educational Assessment? *Asian Journal of Education and Social Studies, 15(3)*: 42-54. ISSN: 2581-6268 https://doi.org/10.9734/ajess/2021/v15i330383

(2020*)* Statement of Retraction: Age- and education-related effects on cognitive functioning in Colored South African women. *Aging, Neuropsychology, and Cognition, 27:6*, 963, DOI: 10.1080/13825585.2019.1614759

Odendaal, A. (2015). Cross-cultural differences in social desirability scales: Influence of cognitive ability. *SA Journal of Industrial Psychology, 41(1)*. Advance online publication. https://doi.org/10.4102/sajip.v41i1.1259

Omoluabi, P. F. (2006). Development and standardization of Perafom Integrated Aptitude Test (PIAT). *Psychology in Africa, 14(1)*, 1–14. https://doi.org/10.4314/ifep.v14i1.23691

Oosthuizen, R. M., & Naidoo, V. (2010). Attitudes towards and experience of employment equity. *SA Journal of Industrial Psychology, 36(1)*. Advance online publication. https://doi.org/10.4102/sajip.v36i1.836

Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences, 13*, 149–159. https://doi.org/10.1016/0191-8869(92)90037-P

Owen, K. (1996). *Aptitude tests.* In K. Owen & J. J. Taljaard (Eds.), Handbook for the use of psychological and scholastic tests of the HSRC (pp. 191–270). HSRC.

Pallant, J. (2011). *SPSS Survival Manual: A step by step guide to data analysis using SPSS* (4th ed.). Allen & Unwin.

Pascoe, M., McLeod, S. (2016). *Cross-cultural adaptation of the Intelligibility in Context Scale for South Africa*. https://doi.org/10.1177/0265659016638395

Paterson, H., & Uys, K. (2005). Critical issues in Psychological test use in the South African workplace. *SA Journal of Industrial Psychology, 31(3),* 12–22. https://doi.org/10.4102/sajip.v31i3.204

Petersen, J. (2018). Gender Difference in Verbal Performance: A Meta-analysis of United States State Performance Assessments. *Educational Psychology Review, 30,* 1269–1281. https://doi.org/10.1007/s10648-018-9450-x

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans., 2nd ed.). International Universities Press. (Original work published 1936)

Piaget, J. (1950). *The psychology of intelligence* (M. Percy & D. E. Berlyne, Trans.). Routledge & Kegan Paul.

Potgieter, T. E., & Van Der Merwe, R. P. (2002). Assessment in the workplace: A competency based approach. *South African Journal of Psychology. Suid-Afrikaanse Tydskrif vir Sielkunde, 28(1),* 60–66.

Psytech, S. A. (2017). *Technical Manual for Critical Reasoning Battery*. Psytech South Africa catalogue: products, services and training offered by Psytech South Africa. Psytech SA.

Ramírez-Uclés, I. M., & Ramírez-Uclés, R. (2020). Gender Differences in Visuospatial Abilities and Complex Mathematical Problem Solving. *Frontiers in Psychology, 11(191),* 191. Advance online publication. https://doi.org/10.3389/fpsyg.2020.00191 PMID:32210859

Reilly, D. H. (2019). *Sex Differences in Cognitive Abilities and Educational Outcomes: Examining the Contribution of Sex-Role Identification*. School of Applied Psychology. Griffith Health. Griffith University, Gold Coast.

Republic of South Africa. (1995). *Labour Relations Act,* Act 66 of 1995. Government Printer.

Republic of South Africa. (1998). *Employment Equity Act*, Act 55 of 1998. Government Gazette, 400, No. 19370. Government Printer.

Republic of South Africa. (2013). *Broad-Based Black Economic Empowerment Amendment Act*, Act 46 of 2013. Government Gazette, No. 37271. Government Printer.

Republic of South Africa. (2019). *Quarterly Labour Force Survey, QLFS 4:2018*. Retrieved 17April 2019, http://www.statssa.gov.za/?p=11882

Reuben, S., & Bobat, S. (2014). Constructing racial hierarchies of skill – Experiencing affirmative action in a South African organisation: A qualitative review. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 40(1),* Art. #1158, 12 pages. https://doi.org/10.4102/sajip.v40i1.1158

Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). Mastering Modern Psychological Testing-Theory and Methods. 2nd Edition. Springer Press.Roman, L. J., &Mason, R. B. (2015). Employment equity in the South African Retail sector: Legal versus competence and business imperatives. *South African Journal of Labour Relations, 39(2)*, 84-104. https://doi.org/10.1007/978-3-030-59455-8

Roman, L. J., & Mason, R. B. (2015). Employment equity in the South African retail sector: Legal versus competence and business imperatives. *South African Journal of Labour Relations*, *39(2)*, 84–104

Roodt, G., & La Grange, L. (2001). Personality and cognitive ability as predictors of the job performance of insurance sales people. *SA Journal of Industrial Psychology, 27(3)*, 35–43.

Rushton, J. P., Skuy, M., & Fridjhon, P. (2002). *Jensen Effects among African, Indian, and White engineering students in South Africa on Raven's Standard Progressive Matrices.* Published by Elsevier Science Inc. PII: S 0 1 6 0 - 2 8 9 6 (0 2) 0 0 093-4 https://doi.org/10.1016/S0160-2896(02)00093-4

Rushton, J. P., & Jensen, A. R. (2005). *Thirty years of research on race differences in cognitive ability*. American Psychological Association. Psychology, Public Policy, and Law, 11(2), 235–294. https://doi.org/10.1037/1076-8971.11.2.235

Rushton, J. P. (2008). *Testing the genetic hypothesis of group mean IQ differences in South Africa: Racial admixture and cross-situational consistency*. Published by Elsevier Ltd., https://doi.org/10.1016/j.paid.2007.10.011

Ryan, A. M., McFarland, L., & Shl, H. B. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52(2),* 359–392. https://doi.org/10.1111/j.1744-6570.1999.tb00165.x

Salgado, J., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology, 12(1),* 1–17. https://doi.org/10.1080/13594320244000292

Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design (Vol. 1).* Sage Publication. https://doi.org/10.4135/9781412961288

Sallis, J. E., Gripsrud, G., Olsson, U. H., & Silkoset, R. (2021). *Research Methods and Data Analysis for Business Decisions: A Primer using SPSS*. Springer Nature Switzerland., 10.10079783030844219. https://doi.org/10.1007/978-3-030-84421-9

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Setiawati, F. A. (2020). Aptitude Test's Predictive Ability for Academic Success in Psychology Student. *Psychological Research and Intervention Journal, 3(1),* 1-12, ISSN 2614-7041 https://doi.org/10.21831/pri.v3i1.34731

Shuttleworth-Edwards, A. B. (2012). Guidelines for the use of the WAIS–IV with WAIS–111 cross-cultural normative indications. *SA Journal of Psychology, Vol. 42*(3), 399-410.

Skuy, M., Shutte, E., Fridjon, P., & O'Carroll, S. (2001). Suitability of neuropsychological test for students in South Africa. *Personality and Individual Differences, 30*, 1413–1425. https://doi.org/10.1016/S0191-8869(00)00123-9.; https://www.sciencedirect.com/science/article/pii/S0191886900001239

Society for Industrial and Organisational Psychology in South Africa. (2022). *SIOPSA Professional Practice Guidelines for Industrial Psycho-legal work.* Retrieved October 7, 2022; https://www.siopsa.org.za/11072022siopsa.pdf

SPSS Inc. (2016). *IBM SPSS Statistics*, Version 23, Release 23.0.0, Copyright IBM Corporation and its licensors. http://www-01.ibm.com/software/analytics/spss/ Date of access: 26 June 2018.

Spelke, E. S. (2005). Sex Differences in Intrinsic Aptitude for Mathematics and Science? A Critical Review. *American Psychological Association, 60(9,)*, 950–958. https://doi.org/10.1037/0003-066X.60.9.950

Steegh, A. M., Höffler, T. M., Keller, M. M., Parchmann, I. (2019). Gender differences in mathematics and science competitions: A systematic review. *Journal of Research in Science Teaching, 56*, 1431–1460.Wiley Periodicals, Inc. https://doi.org/10.1002/tea.21580

Steyn, R., & De Bruin, G. P. (2020). An investigation of gender-based differences in assessment instruments: A test of measurement invariance. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 46(0)*, a1699. https://doi.org/https://doi.org/10.4102/sajip.v46i0.1699

Stoet, G., & Geary, D. C. (2013). *Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data.* PLoS One, 8(3), e57988. https://doi.org/10.1371/journal.pone.0057988 PMID:23516422

te Nijenhuis, J. (2013). The Flynn effect, group differences, and g loadings. *Personality and Individual Differences, 55*, 224–228, Advance online publication. https://doi.org/10.1016/j.paid.2011.12.023

te Nijenhuis, J., Al-Shahomee, A., van den Hoek, M., Allik, J., Grigoriev, A., & Dragt, J. (2015). Spearman's hypothesis tested comparing Libyan secondary school children with various other groups of secondary school children on the items of the Standard Progressive Matrices. *Intelligence, 50*, 118–124. Advance online publication. https://doi.org/10.1016/j.intell.2015.03.002

te Nijenhuis, J., Choi, Y. Y., van den Hoek, M., Valueva, E., & Lee, K. H. (2019). *Spearman's hypothesis tested comparing Korean young adults with various other groups of young adults on the items of the Advanced Progressive Matrices*. Journal of Biosocial Science, https://doi.org/10.1017/S0021932019000026. Cambridge University Press.

Theron, C. (2009). The diversity-validity dilemma: In search of minimum adverse impact and maximum utility. *SA Journal of Industrial Psychology, 35(1)* 183 – 185, Art. #765, 13 pages. doi:10.4102/sajip.v35i1.76. Retrieved on 2 December 2021 from http//www.sajip.co.za

Theron, D., & Roodt, G. (2000). Mental models as moderating variable in 360 degree competency assessments. *SA Journal of Industrial Psychology, 26(2)*, 14–19. https://doi.org/10.4102/sajip.v26i2.702

Thomas, O. O., & Lawal, O. R. (2020). Exploratory Research Design in Management Sciences: An X-Ray of Literature. Annals of the University Dunarea de Jos of Galati: Fascicle: I, *Economics & Applied Informatics, 26(2).*

Trahan, L.H., Stuebing, K.K., Fletcher, J.M., Hiscock, M. (2014). The Flynn effect: a meta-analysis. *Psychology Bulletin Sep; 140(5):*1332-60. Doi: https://doi.org/10.1037/a0037173.

Van der Merwe, R. P. (2002). Psychometric testing and human resource management. *SA Journal of Industrial Psychology, 28(2)*, 77–86. https://doi.org/10.4102/sajip.v28i2.52

Van De Vijver, A. J. R., & Rothmann, S. (2004). Assessment in multicultural groups: The South African case. *SA Journal of Industrial Psychology, 30(4),* 1–7. https://doi.org/10.4102/sajip.v30i4.169

Venter, R., & Levy, A. (2009). *Labour relations in South Africa.* Oxford University Press.

Visser, D., & De Jong, A. (2001). Comparing fairness perceptions of personnel selection techniques of American, French and South African job applicants. *SA Journal of Industrial Psychology, 27(2),* 62–68. https://doi.org/10.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140(4),* 1174–1204. https://doi.org/10.1037/a0036620

Warne, R. (2020). *In the Know: Debunking 35 Myths about Human Intelligence*. Cambridge University Press., https://doi.org/10.1017/9781108593298

Wasserman, J. D. (2019). Deconstructing CHC. *Applied Measurement in Education, 32(3),* 249–268. https://doi.org/10.1080/08957347.2019.1619563

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2010). *Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn Effect.* Elsevier Publication., 10.1016j.lindif2009.12.001.

Witte, R. S., & Witte, J. S. (2017). *Statistics*. 11th Edition. Hoboken, NJ. John Wiley & Sons Inc. ISBN: 9781119254454(EVALC).

Zunker, V. G. (2016). *Career counselling: A holistic approach*. Cencage Learning