THE RELATIONSHIP BETWEEN CTB2 NUMERACY SCORES AND WORK PERFORMANCE OF CALL CENTRE AGENTS IN A SOUTH AFRICAN RETAIL COMPANY

by

Peggy Happy Maphanga

Submitted in partial fulfillment of the requirements for the degree:
Master of Arts (Research Psychology)
in the

Department of Psychology
Faculty of Humanities
University of Pretoria

April 2011

Supervisor: Mrs. Rebone Gcabo
DECLARATION

I declare that this research report entitled *The relationship between CTB2 numeracy scores and work performance of call centre agents in a South African retail company* is my own work. All the sources used and quoted in this research have been indicated and reflected by means of a complete reference.

Signed: Ms Peggy Maphanga
Researcher

Date: 02/02/2011
Acknowledgements

My thanks go to the Almighty God who has made it possible for me to complete this project. I would also like to acknowledge and thank the following people for their contributions and support:

- My supervisor Rebone Gcabo for her guidance and patience; I thank you a lot.
- Tim Joubert for language editing, and Sue Randall for further help with editing and the methodology sections.
- My internship supervisors Andrew Johnson and Maria Sebate for their support and guidance.
- The assessments department of the company under study, for allowing me to access the database.
- My dearest parents for their support and encouragement throughout my studies. You sacrificed your happiness and material possessions in order to make me a better person in society, and for that I shall forever remain grateful. I love you so much.
- My two sisters, my brother, nephews and nieces: thank you for the support and love you gave me.
- Sabelo Nobangule, my partner; for your support and encouragement. I wouldn’t have made it without you.
Abstract

The call centre industry is gaining momentum in the retail sector, which can be attributed to the fact that many companies use call centre agents to communicate with the end users of their products. A major challenge facing retail companies is finding the right employees for their call centres, so that these people are able to perform according to their natural abilities and acquired skills.

The South African retail company under study uses the Clerical Test Battery (CTB2) to screen candidates for its call centre. The aim of this study was to investigate the relationship between the CTB2 numeracy subtest scores and work performance scores of employees in this call centre. The literature review examines the use of psychometric instruments in general, and then focuses on the use of the CTB2 in South African contexts. The objective of the current study was to evaluate the validity of the numeracy subtest when used as part of the selection process for call centre candidates.

A quantitative research approach was employed to analyse the raw data, which was already available in the company’s database. The two datasets used were the CTB2 numeracy subtest scores and the work performance scores of 300 call centre agents. The correlation between these two sets of scores was analysed using Statistical Package for Social Sciences (SPSS) software (version 17). No statistically significant relationship emerged between the two research variables. Therefore, it may not be appropriate to use the CTB2 numeracy subtest to screen call centre candidates.

The findings of this study suggest that the CTB2 should not be used in isolation when assessing call centre candidates. Face-to-face interviews should form an important part of the decision-making process, and the use of additional instruments to measure general cognitive aptitude and skills such as perceptual speed and accuracy is recommended. The dissertation concludes by outlining the limitations of the study and giving some suggestions for future research in this field.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPCAT</td>
<td>Learning Potential Computerized Adaptive Test</td>
</tr>
<tr>
<td>CTB2</td>
<td>Clerical Test Battery</td>
</tr>
<tr>
<td>H₀</td>
<td>Null Hypothesis</td>
</tr>
<tr>
<td>H₁</td>
<td>Alternative Hypothesis</td>
</tr>
<tr>
<td>HPCSA</td>
<td>Health Profession Council of South Africa</td>
</tr>
<tr>
<td>SA</td>
<td>South Africa</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for Social Sciences</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

## CHAPTER 1: INTRODUCTION

1.1 Introduction .................................................................................................................. 1

1.2 Background to the current study ................................................................................... 2

1.3 Rationale for the study ................................................................................................. 5

1.3.1 Overview of call centres in the retail company under study ................................. 6

1.3.2 Performance ratings for call centre agents ......................................................... 6

1.4 Research problem ....................................................................................................... 9

1.5 Research question ....................................................................................................... 9

1.6 Purpose of the study ................................................................................................... 10

1.7 Aim of the study ......................................................................................................... 10

1.8 Research hypothesis ................................................................................................... 11

1.9 Significance of the study ............................................................................................ 12

1.10 Clarification of terms ................................................................................................. 13

1.10.1 Reliability ............................................................................................................. 13

1.10.2 Validity ................................................................................................................. 13

1.10.3 Correlation ........................................................................................................... 13

1.10.4 Positive relationship ............................................................................................ 14

1.10.5 Statistical significance ......................................................................................... 14

1.10.6 Job category and band ......................................................................................... 14

1.10.7 Performance ratings ............................................................................................ 14

1.10.8 Performance scores ............................................................................................. 15

1.10.9 Call centre ........................................................................................................... 15

1.10.10 Call centre Agent .............................................................................................. 15

1.10.11 Clerical Test Battery (CTB2) ........................................................................... 15

1.10.12 Numeracy subtest ............................................................................................. 15

1.11 Assumptions .............................................................................................................. 16

1.12 Overview of the dissertation ...................................................................................... 16

1.13 Conclusion ................................................................................................................. 17
## CHAPTER 2: LITERATURE REVIEW

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Introduction</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Psychometrics: origins and background</td>
<td>18</td>
</tr>
<tr>
<td>2.3 Challenges to the field</td>
<td>23</td>
</tr>
<tr>
<td>2.4 The misuse of psychometric tests</td>
<td>24</td>
</tr>
<tr>
<td>2.5 Contexts in which psychometric tests are used</td>
<td>25</td>
</tr>
<tr>
<td>2.6 Characteristics of assessment instruments</td>
<td>26</td>
</tr>
<tr>
<td>2.6.1 Reliability of instruments</td>
<td>27</td>
</tr>
<tr>
<td>2.6.2 Validity of instruments</td>
<td>27</td>
</tr>
<tr>
<td>2.7 Assessments in the workplace</td>
<td>28</td>
</tr>
<tr>
<td>2.8 Types of workplace assessment</td>
<td>29</td>
</tr>
<tr>
<td>2.9 Psychometric assessments in South Africa</td>
<td>30</td>
</tr>
<tr>
<td>2.9.1 Classification of psychological tests in South Africa</td>
<td>31</td>
</tr>
<tr>
<td>2.9.2 Psychometric assessments pre-1994</td>
<td>32</td>
</tr>
<tr>
<td>2.9.3 Psychometric assessments post-1994</td>
<td>35</td>
</tr>
<tr>
<td>2.9.4 Legislative changes</td>
<td>37</td>
</tr>
<tr>
<td>2.10 Use of psychometric tests at the company under study</td>
<td>38</td>
</tr>
<tr>
<td>2.11 The use of the Clerical Test Battery (CTB) in South Africa</td>
<td>41</td>
</tr>
<tr>
<td>2.11.1 History of the CTB2 in South Africa</td>
<td>41</td>
</tr>
<tr>
<td>2.11.2 Level of education required for completing the CTB2</td>
<td>42</td>
</tr>
<tr>
<td>2.11.3 Standardisation of the CTB2 in South African retail companies</td>
<td>42</td>
</tr>
<tr>
<td>2.11.4 Design of the Clerical Test Battery</td>
<td>43</td>
</tr>
<tr>
<td>2.12 Theoretical model of competencies</td>
<td>44</td>
</tr>
<tr>
<td>2.13 Reliability of CTB2 in the South African context</td>
<td>48</td>
</tr>
<tr>
<td>2.13.1 The CTB2 in a retail environment</td>
<td>49</td>
</tr>
<tr>
<td>2.13.2 The CTB2 in a call centre context</td>
<td>51</td>
</tr>
<tr>
<td>2.13.3 Correlation between CTB2 and Occupational Personality Profile scores</td>
<td>53</td>
</tr>
<tr>
<td>2.14 Validity of CTB2 in the South African context</td>
<td>54</td>
</tr>
<tr>
<td>2.14.1 Correlation between CTB2 and LPCAT</td>
<td>54</td>
</tr>
<tr>
<td>2.14.2 Correlation between CTB2 subtests</td>
<td>55</td>
</tr>
<tr>
<td>2.15 Personality and creativity: a psychometric study</td>
<td>56</td>
</tr>
<tr>
<td>2.16 A review of South African research into dynamic assessments</td>
<td>57</td>
</tr>
<tr>
<td>2.17 Conclusion</td>
<td>58</td>
</tr>
</tbody>
</table>
CHAPTER 1
Introduction

1.1 Introduction

In today’s global marketplace there is a large demand for highly skilled individuals, and it is of great value for organisations to have the best recruitment methods possible. It is vital for businesses to recognise that attracting and retaining the right people for the job allows them to stay ahead of competitors (Nicholls, 2006). As South Africa is still a developing country, there are not always enough people with the appropriate skills to render good customer services to new clients while still retaining existing clients (Nicholls, 2006). Organisations are forced to work with fewer personnel, yet it is still expected of them to function optimally with this limited capacity. Every organisation strives to have the best employees for a job, hence they look for candidates who have optimal ability to absorb and apply knowledge.

Psychometric tests are widely used in selection processes for the workplace, and their use has increased over the years in business and industry (Menday, 1996). Psychometric tests aim to measure people’s attributes as objectively or quantitatively (mathematically) as possible, and include both ability and personality tests. Ability tests assess a person’s current ability in a specific field or their potential for learning to apply skills in their actual job, and for future achievement, while personality tests describe the personality traits that remain stable throughout a person’s adult lifetime (Black, 2000).

Van der Merwe (2005) states that in the early 1990s the outcomes of psychometric tests were somewhat disappointing because they were not standardised across racial groups. As a result they tended to discriminate
against some population groups while favouring others, simply by the nature of their design (Owen, 1992). For example, individuals with a low socio-economic status are often unfamiliar with psychometric tests and written tests in general, and they may not feel that such tests are important; these attitudes can negatively affect their scores. Children from impoverished communities often complete intelligence tests more quickly and choose more random answers, and their scores tend to be lower than the scores of children from more privileged backgrounds. This may mean that the instrument tends to measure such children’s educational background or their familiarity with written tests rather than actually measuring their intelligence (Owen, 1992).

Because of this problem, many tests in South Africa are now standardised and checked for accuracy with indigenous population groups before they are used in the workplace or educational settings. For example, the Learning Potential Computerised Adaptive Test (LPCAT) was validated at the retail company under study to verify, amongst other things, the test’s applicability and usability specifically in that company’s environment in 2006 (de Beer, 2007). The more developed, validated and reliable the psychometric tests are, the more successful their implementation in industries and businesses.

1.2 Background to the current study

The retail company which was examined in this research project is one of South Africa’s leading clothing, footwear and textiles retail groups. Henceforth in this report it will be referred to simply as “the company”. The company trades through a range of retail formats including clothing, stationery and home décor shops. The first store was opened on 6 September 1929 and was located in Joubert Street, Johannesburg.

Since then the company has grown to include ten retail brands trading in over 900 stores in South Africa, Botswana, Namibia, Swaziland and Lesotho. The company is a multi-brand retailer with a dominant sub-Saharan footprint, with customers across the spectrum of the southern part of the African continent. It
claims to have more than 4 million customers who trade in their stores during daytime shopping hours (the company’s database, 2006; the company’s website, 2010).

The company’s vision is to be the consumer’s store of choice, both for the products sold and because of the attractive environment in which trade takes place. Furthermore, the company strives to create wealth for the benefit of all its stakeholders (the company database, 2006).

Job categories at the company are divided according to the Hay grading system. The Hay grading system is a method used by many corporations and organisations to map out job roles in the context of the organisational structure. It is meant to enable organisations to map all the relevant roles in a manner that the deliverables can be recognized, improved upon, and used to inform more useful and focused job descriptions (Niles & Harris-Bowlsbey, 2008).

Although the Hay grading system is mainly used for the evaluation of managerial jobs, an increasing number of companies use it to achieve equitable salary standards for all job levels (Scottsdale, 1998). The company under study has developed a band system which is used together with the Hay grading system to assist in talent management, and to grade its employees’ progress in their work deliverables. According to this system, Bands A and B are on an operational or entry level, and comprise human resource assistants and consultant and assessment practitioners. Band C includes junior management, specialists and HR managers. Senior specialists and divisional managers fall under Band D, while senior management and HR executives are classified as Bands E and F. Band G consists of top leadership such as the chief executive in all business units.

In the retail industry, organisations like the one under study are faced with the challenge of selecting people who can meet and maintain high performance levels in demanding jobs. A major concern facing businesses is how to align the right person with the right job specification, where they will be able to perform according to their abilities. According to Muller, Buhner and Ellgring
(2003), good recruitment methods contribute 60% towards ensuring that individuals are placed in areas where they can perform to their best. Sound recruitment and selection strategies result in improved organisational outcomes. Van der Merwe (2005) states that the more effective organisations are at recruiting and selection, the more likely they are to hire and retain people who are competent to perform their role at an optimal level. For companies to meet this challenge, they generally use psychometric testing as one of the ways in which candidates are assessed.

Assessments allow human resource managers to identify people with skills and abilities that match the job requirements (Muller et al., 2003). Companies including the one under study use psychometric tests as a method of measuring the human attributes that are necessary to carry out the job. The goal of measuring these attributes in the workplace is thus to identify individual potential, and, accordingly, place people in the right jobs (Muller et al., 2003).

Van der Merwe (2005) states that selection is the process whereby an organisation identifies people who match specified job requirements and will be able to perform optimally at the specified job level. For example, in the company under study, departmental planners are required to have a high degree of mathematical ability in order to carry out their work.

The researcher was employed for some time at the company under study and she became aware of important questions surrounding the use of psychometric tests in the selection process. This led to an interest in examining in more detail the nature of those issues and attempting to clarify the findings, for the benefit of both the company itself and for students of human resources and business management in the South African context.
1.3 Rationale for the study

Many assumptions are made regarding any organisation’s human capital selection process. In the multicultural context of South Africa, the use of psychometric tools in the labour sphere has rightly been questioned (Muller et al, 2003). Recurring issues about the validity and reliability of the measuring instruments arise whenever results are glaringly inconsistent or not clearly explainable.

In the company under study, the Clerical Test Battery Report Version 2 (CBT2) was used as part of the selection process. The numeracy subtest of the CTB2 was routinely given to most people who applied for positions as call centre agents and this practice was being criticised by several business units, specifically call centre managers. The main concern of the managers was that the numeracy subtest had little or no significance and relevance to the job of a call centre operator. Many candidates who did fairly well on the overall test battery scored very poorly on the numeracy component, while scoring better on other subtests.

Additionally, candidates complained that they could not meet the test’s cut-off time (of one hour for completing the entire battery) because of the delay they encountered while trying to work through the numerical subtest. It was therefore being questioned whether this particular subtest was valid for the specified job category. The numeracy subtest might be valid for other job categories but in this case its relevance was questioned. It became important to determine whether the managers’ and candidates’ complaints were valid, and therefore to make recommendations about the use of the numeracy subtest in assessing applicants for call centre positions.

Before examining the research implications of this situation further, details are given in the following sections about the company’s structure and ways of operating, especially with regard to the group under study, namely call centre staff.
1.3.1 Overview of call centres in the retail company under study

The company’s call centre operated two services at the time of this study, namely an inbound service and an outbound service. The inbound service refers to instances where customers call in either to lodge a complaint or to ask for information (queries). In contrast, at the outbound service, agents call prospective customers in order to sell and market the company’s products (White, 2003). Accounts are also handled at the outbound centre. Call centre agents are employees whose main role in the business is to handle incoming and outgoing customer calls so as to meet clients’ requests and follow up on account payments (White, 2003).

Read (2001) defines call centres as places where activities such as sales, promotions, customer service and debt collections are performed, without any physical contact between agent and customer ever taking place. Due to the call centre’s inherent flexibility, the costs of meeting customer needs and expectations are reduced.

1.3.2 Performance ratings for call centre agents

Call centre agents are increasingly taking on important responsibilities in their role as contact points between the customer and business. Thus, when managers have to rate call centre agents, they should carefully consider the agents’ attributes and performances (Read, 2001). Call centre environments are dynamic in nature, and it is therefore important that the performance scores of call centre agents are accurately generated in a consistent and reliable way. Line managers are faced with the challenge of identifying important aspects of operator and call centre performance (Nicholls, 2006).

The company under study has adopted its performance measures from the theory developed by Nicholls (2006). According to Nicholls there are a number of performance approaches that can be used to evaluate an agent’s overall performance. These include productivity measures, adherence
measures and qualitative measures. At the company under study, the productivity measure was determined by the number of phone calls made and received by the agent during the day. The adherence measure evaluates the agent's ability to make calls and render services in adherence to the designated time schedule: for every call made, a certain period of time is allocated. In order to reach their targeted timeline for the week, agents have to make certain that they do not exceed that time limit. The third measure, the qualitative measure, measures the agents' knowledge of the specific product or service they render to the customer (White, 2003).

Line managers work with assessment managers to review their employees' performance and to discuss whether performance has declined or increased over a given period of time. For example, fewer calls translate into poor service, while good service implies the making of more calls. Employees' assessment scores are carefully reviewed and compared with their job performance scores in order to see whether there is consistency between the two. The ensuing results determine the overall performance scores of every call centre employee.

Els and de Villiers (2006) state that it is often very difficult for call centre agents to strike a balance between the quality and quantity of phone calls they make. Agents are constantly under pressure to complete a certain number of calls while at the same time maintaining a high quality and professional approach to handling and making calls. Agents facing this challenge may fall into the trap of managing only one aspect, and in neglecting others, it negatively impacts on overall performance results.

Read (2001, pg. 47) notes that "The quality of a service delivered in a customer focused environment cannot be separated from the competency or dedication of persons delivering the service. Organisations therefore need to devise and implement an approach that assists agents in balancing the quality and quantity of their workload. One possible approach is to assign agents realistic timelines, while at the same time underscoring the importance of delivering outstanding service through training. Call centre agents should
further, and on a continual basis, be made aware of the fact that they are business representatives: that is, the service they deliver to customers reflects back on the business image as a whole (Read, 2001).

Numerical Reasoning is that part of the Clerical Test Battery which is used to analyse candidates’ potential to reason and think logically. It is the only component that will be discussed in this study, as it was the only contentious subtest in the CTB2 as used to assess potential call centre agents. Numerical reasoning is a measure of numerical ability, which has been isolated as a component of general reasoning. In the numeracy subtest candidates have to draw inferences from the numerical information presented to them in tabular form. The questions have six possible answers, and candidates are required to pick the correct answer (Psytech International, 1991; Psytech-SA, 2010).

In 2007 call centre managers and assessment managers questioned if the numeracy subtest was valid for candidates who applied for positions as call centre agents. Most of the candidates were not able to complete their assessment test and complained that they spent most of their allocated time on the numeracy subtest. All CTB2 subtests were completed electronically using a mouse to select the correct answer, and each subtest was supposed to take no longer than 8 minutes (Psytech-SA, 2010). The three subtests used for call centre candidates were verbal reasoning, numeracy and spelling check.

It must be acknowledged that the numeracy subtest might be valid for other job categories, such as administrators and program assistants, but its validity in this instance was questioned and it was imperative to attend to the issue. The challenges of the numeracy subtest currently in use have been pointed out and discussed. The following section will give a brief explanation of the research problem and the aim of the study.
1.4 Research problem

The reliability and validity of the CTB2 has been tested in various South African contexts such as in other retail industries, the police service and fast moving consumer goods (FMCG) environment. In order to test the reliability of the test battery, data were collected on a sample group of 126 clerical staff Band A attending an in-service training course. The results revealed that the verbal reasoning, numerical ability, clerical checking (verbal) and clerical checking (numerical) subtests all have a high level of reliability (Psytech International, 1991; Psytech-SA, 2010).

However, in the company under study there was a low face validity of the CTB2 as evaluated subjectively by some of the call centre managers. Face validity refers to whether the instrument appears, on the surface of things and without statistical investigation, to measure what it purports to measure (Rosenthal & Rosnow, 1991). Call centre agents’ subsequent work performance results indicated that the predictive validity of the numeracy subtest needed to be investigated. It was therefore important to establish whether or not there was a statistical relationship between, on the one hand, candidates’ scores on the numerical skills subtest of the CTB2 and, on the other hand, their later work performance scores once they had started working as call centre agents.

1.5 Research question

The following specific research question was formulated:

*Does a statistically significant positive relationship emerge between individuals’ scores on the numeracy subtest of the CTB2 and their later work performance scores?*

The population under study was limited to the specific company context, and the research subjects were candidates who had passed through the initial
assessment process and had gone on to become call centre agents. Each subject’s CTB2 numeracy scores would be compared to their own later work performance scores.

1.6 Purpose of the study

The purpose of this study was to test the relationship between CTB2 numeracy subtest scores and later job performance, as measured in call centre candidates and staff at a South African retail company. The results of this investigation would be compared with other data obtained for the CTB2 in various South African contexts.

The purpose of this study was therefore twofold, firstly to increase the understanding of what was happening in the context of the specific company under study. It was anticipated that the study results would inform Assessment and Call Centre managers about the adequacy and accuracy of the CTB2 subtest as used for Band A\(^1\) call centre agents. Secondly the findings would contribute to a more general knowledge about the use of CTB2 within South African populations.

1.7 Aim of the study

The specific aims of this study were to:

- Ascertain the CTB2 numeracy scores for a sample of candidates who had been assessed in the past
- Ascertain the work performance ratings of the same individuals at a later point in time, after they had been employed by the company
- Establish whether or not a statistical relationship could be found between these two sets of scores

\(^1\) See section 1.2 for a definition of Band A
• Determine whether any such relationship was positive (i.e. higher numeracy scores are associated with higher work performance ratings)
• Determine whether any such relationship was statistically significant, or was likely to have been caused by chance or measurement errors.

Based on these findings it would be evaluated whether the numeracy subtest of the CTB2 was a valid and useful instrument in the selection of call centre agents, or not.

1.8 Research hypothesis

A hypothesis is a logical supposition that gives a tentative explanation for a phenomenon under investigation (Leedy & Ormrod, 2001). The different types of hypotheses are as follows:

• The null hypothesis \( (H_0) \) is the hypothesis that is tentatively held to be true, and which states that there is no statistically significant relationship between the variables under study
• The alternative hypothesis \( (H_1) \) (also known as the experimental hypothesis) states that there is a statistically significant relationship between the variables
• A directional hypothesis describes the expected nature of the relationship between variables
• A non-directional hypothesis does not predict the nature of any relationship which may be found between the variables.


Research hypotheses were formulated for the current study to determine whether a relationship exists between the CTB2 subtest scores and later performance ratings. The hypothesis was stated in non-directional terms for reasons which are discussed in the chapter on methodology (see section 3.8, p. 73). The hypotheses were formulated as follows:
\( H_0 \) There is no significant correlation between individuals’ scores on the numeracy subtest of the CTB2 and their later work performance ratings \((r = 0)\)

\( H_1 \) There is a significant correlation between individuals’ scores on the numeracy subtest of the CTB2 and their later work performance ratings \((r \neq 0)\).

1.9 Significance of the study

This research was important because it set out to determine whether a positive relationship exists between CTB2 numeracy subtest scores and the work performance scores of Band A call centre employees in the company under study. That is, the study investigated whether numerical ability could accurately predict work performance for call centre agents. Investigating the relationship between these two variables statistically would show whether the subtest scores have any relationship to the subsequent performances of call centre agents.

Insight into this relationship or lack thereof would help to determine whether the numeracy subtest of the CTB2 should be phased out or sustained in the company’s assessment department. If no statistically significant correlation was found to exist between the two variables, the company’s head office should consider ceasing to use this subtest when screening call centre candidates. A greater understanding of the selection methods used during recruitment would help the company’s human resource practitioners to select appropriate tools when testing the ability of applicants in lower grades such as Band A.

On a broader level, this study can contribute to knowledge about the reliability and predictive validity of the CTB2 numeracy subtest when it is used with a specific South African population. This may help the test developers to ensure that their instrument is designed in such a way that it measures what it is
intended to measure, and that performance on the test items remains consistent over time. The test developers also have a responsibility to inform organisations about the educational experience which applicants will need in order to take the test effectively.

Furthermore, all assessment instruments should be selected on the basis of an analysis of the job for which the candidates are applying. In the case of the company’s call centre, it has not been adequately demonstrated or justified why it is important to assess a candidate’s numerical ability, other than to say that this taps an aspect of their ability to think logically and employ reasoning skills.

1.10 Clarification of terms

The following terms are used throughout this report, and are clarified in this section for ease of reference:

1.10.1 Reliability
The extent to which a measuring instrument assigns the same score to the same characteristic each time the measurement is repeated; that is, its stability or consistency over time (under conditions that are essentially the same).

1.10.2 Validity
The extent to which a measuring procedure accurately measures the variable or characteristic it is supposed to measure (rather than accidentally measuring an extraneous variable).

1.10.3 Correlation
The degree of relationship or co-variance between two variables.
1.10.4 Positive relationship
A relationship of co-variance between two variables such that an increase in variable A is associated with an increase in variable B, while a decrease in A is associated with a decrease in B. This co-variance does not necessarily imply causality. In a negative or inverse relationship, an increase in A would be associated with a *decrease* in B.

1.10.5 Statistical significance
A mathematically calculated assessment of whether the research findings might be attributed purely to chance or not; this also determines whether or not to accept or reject the null or alternative hypotheses.


1.10.6 Job category and band
†Job category† refers to the classification, for purposes of analysis or official reporting, of the company†s employees according to job requirements or job descriptions. As mentioned earlier in this report (see section 1.2), different job levels in the company are referred to as bands. Bands A and B are the operational level or entry level. Call centre agents are classified as Band A, while in the human resources field Band A includes HR assistants and administrators.

1.10.7 Performance ratings
A performance rating is a procedure used in the workplace to determine the value for a specific factor, which in turn is used to measure an employee†s observed task performance against the task time normally allowed for the task. This procedure assists managers to ascertain if their staff can perform at the normal pace of the job requirement (Wilson, 2001).
1.10.8 Performance scores
This refers to the total number of points attributed to an employee according to the level of effectiveness with which he or she carries out work duties after taking up a position in the company (Wilson, 2001).

1.10.9 Call centre
A call centre is a place where calls are placed or received in high volume for the purposes of sales, marketing, customer service, telemarketing, technical support or other specialised business activity. (Bodin & Dawson, 1999, p. 45).

1.10.10 Call centre Agent
A call centre agent is a person who handles incoming or outgoing customer calls for business purposes. Depending on the agent’s department, a call centre agent might handle account inquiries, customer complaints or support issues (SearchCRM.com, 2009; Callcentrekit.com).

1.10.11 Clerical Test Battery (CTB2)
The Clerical Test Battery is a test battery which was developed to assess the aptitudes that are predictive of performance in a wide range of clerical jobs (Psytech International, 1991; Psytech-SA, 2010).

1.10.12 Numeracy subtest
The numeracy subtest of the CTB2 measures the ability of a candidate to use numbers efficiently in clerical and administrative contexts. This test assesses the ability to perform tasks such as calculating travelling expenses and working out the unit pricing of goods (Psytech International, 1991; Psytech-SA, 2010).
1.11 Assumptions

The following assumptions apply to this study:

- It was assumed that the sample’s data had the characteristics of normal distributions, and that the use of statistical tests for normal distributions would therefore be appropriate.
- This study is exploratory in nature and its findings should be considered descriptive rather than definitive. Findings are limited to the specific context and time under study (i.e. the company’s call centre staff prior to or during 2006), and should not be generalised to other groups in the absence of further studies.
- It was assumed that accurate records had been kept by the company of candidates’ scores on the CTB2 numeracy subtest, and of their later job performance scores.
- A correlation between numeracy subtest scores and job performance was the only relationship to be investigated. Other CTB2 subtests were not investigated as these were not perceived to be problematic, and the researcher was given permission to access only the numeracy subtest scores.
- The company’s human resources department would make the results of the study available to all staff who could benefit from it.

1.12 Overview of the dissertation

This dissertation consists of five chapters. This chapter (Chapter 1) has provided the introduction and background to the study and the rationale for carrying out the research, as well as listing the specific aim, objectives and significance of the research. Key terms used in this report and the study’s underlying assumptions have been clarified.
In Chapter 2 a literature review is presented, which includes a brief survey of some of the existing literature related to psychometric testing and assessment tests in general. The process of investigating the reliability and validity of such instruments is outlined, in preparation for the methodology chapter which will follow. This is followed by an overview of the specific test battery (CTB2) under study, with special attention paid to the numeracy subtest.

Chapter 3 discusses the research design and the methodology used for gathering and analysing the data used in this study. The steps taken to conduct the analysis are described and explained. Chapter 4 presents the actual findings of the research in tabular format as well as text, and discusses these results in some depth.

Chapter 5 provides a conclusion to this research, and clarifies the limitations of the study. The implications of the findings are discussed, and recommendations are made for future research in similar areas.

1.13 Conclusion

In this chapter, the introduction and background of the study is discussed. The aim, purpose, objectives and significance of the study is also highlighted. Chapter 2 will discuss psychometrics, its definition, origins and assessment trends. The use of assessment tests in the South African context will also be discussed as it is reported in the relevant literature.
CHAPTER 2
Literature review

2.1 Introduction

In this chapter, some of the existing literature on psychometric instruments is reviewed. The origins of psychometrics and trends within the field of assessment will be examined. The use of psychometric tests in the South African context will also be critically discussed.

2.2 Psychometrics: origins and background

Psychometrics can be defined as the field of study concerned with the theory and practice of educational and psychological measurement (Anastasi, 1988). This includes measuring the knowledge, abilities, attitudes and personality traits of people. The psychometric field is concerned with the construction, study and refinement of measurement instruments such as questionnaires and tests. A psychometrist (also referred to as a psychometrician) must understand the principles and procedures involved in psychological measurement, and the development of theoretical approaches to such measurement (Anastasi & Urbina 1997). Another focal point of modern psychometrics is the development of new methods of statistical analysis or the refinement of older techniques, which has been made easier by the advancement of computer technology (Rosenthal & Rosnow, 1991).

The field of psychometrics has been in existence since the early 1800s and it initially focused mainly on measuring intelligence. Foxcroft, Roodt and Astbury (2001) state that psychological assessments started growing as a science due
to the development of scientific methods, and that the primary principle underlying the scientific method of assessment is measurement. During the early stage of psychometrics, many psychologists began to accept that such assessments could be useful (Weiten, 1989; Sundberg, Taplin & Tyler, 1983; Indiana University).

In the early 1900s there was a rapid and vast development of specific assessment instruments. This progress was due to the advancements of human behavioural theories such as classical conditioning theory, and the statistical methods that aided the analysis of data obtained from measures to determine job performance. Soon assessments were being applied in clinical, educational, industrial and military contexts (Foxcroft et al., 2001; Weiten, 1989).

Assessments were also in demand to resolve the need to treat mentally disturbed and disabled people in a more humanitarian way. Due to the development of assessment tests, mental disorders and deficiencies could be better assessed and classified. This in turn called for new methods of differentiating between people who were intellectually challenged and those who suffered from emotional disorders (Foxcroft et al., 2001; Weiten, 1989; Sundberg et al., 1983).

Much can be understood about psychometrics by studying the history of intelligence tests and related instruments that were developed around the same time. The remainder of this section therefore summarises some of the key events in this area, after giving an initial explanation of the nature of intelligence.

Intelligence is a construct, which is an abstract idea used to explain an observed phenomenon (Rosenthal & Rosnow, 1991). For example, a child who scores highly on both mathematics and English tests is usually inferred to have a high intelligence. But intelligence is not directly observable, as are physical qualities such as height or weight. Therefore what is being measured is not necessarily the construct in question (Rosenthal & Rosnow, 1991;
Weiten, 1989; Kerlinger, 1986). It is possible that the test is measuring the child’s education rather than intelligence, or a combination of both variables (Rosenthal & Rosnow, 1991; Kerlinger, 1986).

Sir Francis Galton (born in 1822 and deceased in 1911) was the first person to suggest that intelligence is hereditary. He was born into a wealthy family and Charles Darwin was one of his cousins (Weiten, 1989; Bulmer, 2003). Galton found that social and professional success seemed to run in certain families. In 1869 Galton published the theory that success was due to genetic superiority, but he did not acknowledge that these families were usually the wealthiest ones, where people were well-educated and had a lot of social influence (Weiten, 1989).

Galton also theorised that higher intelligence was associated with faster sensory reaction. He carried out research to test this hypothesis and saw it was not true. But through his experiments Galton began to develop the method of statistical correlation, which showed that it might be possible to measure mental ability (Weiten, 1989; Andrich & Luo, 1993).

Alfred Binet furthered Galton’s work with intelligence testing, and in 1904 Binet was commissioned to create a test to identify intellectually handicapped children who would not benefit from mainstream education (Weiten, 1989; Indiana University; Foxcroft et al., 2001). Binet thought that verbal reasoning ability and the ability to manipulate data would be the best markers of intelligence. In 1905 Binet and Simon published an intelligence test (the Binet-Simon test) which demonstrated both practical value and good validity, in that it predicted school success quite well. It was soon in use across Europe and America. Binet revised the test twice before his death in 1911 (Weiten, 1989; Indiana University; Sundberg et al., 1983; Foxcroft et al., 2001).

Other psychologists continued this work, notably Terman and Wechsler. Wechsler developed adult intelligence tests, while Terman’s team at Stanford University modified Binet’s test and in 1916 the Stanford-Binet Intelligence
Scale was published. It yielded an intelligence quotient or IQ, a measurement that is still in use today (Weiten, 1989; Kerlinger, 1986; Jones & Thissen, 2007). The Stanford-Binet test became the golden standard for assessing any new intelligence tests (Weiten, 1989; Sundberg et al., 1983).

Moving away from intelligence testing, psychiatric screening instruments were next devised for use in the military during World War One. In 1917 Robert Yerkes, an American psychologist, helped to develop screening instruments to assess soldiers, and Terman was also involved in this (Jones & Thissen, 2007). The Army Alpha was based on Binet's tests but modified for adults, while the Army Beta was a non-verbal test used for illiterate or non-English speakers (Sundberg et al., 1983). The final versions of these two tests were published in 1919 and by the end of the war they had been used to assess about 2 million men (Indiana University).

The Army Alpha and Beta tests were a milestone as this was the first time group intelligence tests had been used. The general public became aware of psychometric testing and it became more accepted, although there was also a growing awareness that different racial groups tended to score differently and the reasons for this were not clear (Indiana University). Personality assessments, and another psychiatric screening test for army recruits developed by Woodworth, were also created around this time (Weiten, 1989; Sundberg et al., 1983; Kerlinger, 1986).

Charles Spearman was the next influential intelligence researcher. His theory allowed for specific domains of intelligence as well as a general or core ability. He named the specific domains S₁, S₂, S₃ and so on, and he called general ability řgò (Weiten, 1989; Andrich & Luo, 1993). The relevance of this approach to the current research is evident in that the CTB2 battery has different subtests, which are roughly comparable to Spearman’s specific domains. Spearman made use of factor analysis, a statistical technique which analyses the relationships between variables and the patterns by which some scores cluster together (Rosenthal & Rosnow, 1991; Kerlinger, 1986; Jones &
Thissen, 2007). The factor analytical approach is still in wide use today (Black, 2000; Jones & Thissen, 2007).

In the 1940s, during the World War Two era, Thurstone argued that Spearman had put too much emphasis on Ŕg. He suggested instead that there were seven main mental abilities which might be very unevenly distributed within one individual (Weiten, 1989; Jones & Thissen, 2007). In contrast, R.L. Thorndike, another prominent psychometrist, was moving away from trying to identify and measure specific abilities towards a more global approach (Indiana University). Once again the relevance to the current research lies in the fact that ability in one specific area, such as numeracy, may not translate to general competence, especially in a job that may deal with issues not related to numeracy. Guilford followed Thurstone's line of thinking and did away with the concept of Ŕg altogether, dividing intelligence into 150 separate abilities (Weiten, 1989; Jones & Thissen, 2007).

Currently, intelligence research focuses more on cognitive processes rather than on supposedly fixed traits. Robert Sternberg has proposed a popular cognitive model of intelligence in which knowledge acquisition, rather than inherent ability, is seen as being most important (Weiten, 1989; Jones & Thissen, 2007). Less emphasis is placed on IQ tests today, and achievement or outcome-based tests are often preferred as a method of assessment. Modern psychometric testing is used in many contexts including retail companies and tertiary institutions, which use psychometric tests for recruitment and enrolment purposes (Murphy, 2002).

With regard to personality testing, there have been a number of theoretical approaches to measuring personality traits, with one of the best known currently being the Five Factor Model (also referred to as ŔBig 5Ό. Some of the better known instruments include the Minnesota Multiphasic Personality Inventory (MMPI), the Personality and Preference Inventory (PAPI) and the Myers-Briggs Type Indicator (Andrich & Luo, 1993; Rosenthal & Rosnow, 1991).
People’s attitudes have been extensively studied in the field of psychometrics, with these types of instruments often making use of a Likert scale. A Likert scale presents the reader with various statements and asks them to rate their agreement or disagreement on a scale ranging from “strongly disagree” to “strongly agree” (Rosenthal & Rosnow, 1991).

2.3 Challenges to the field

In spite of the developments and progress that assessments were making over the years, there were still challenges that fuelled criticism. The limitations and weaknesses of measures were criticized, specifically for being too dependent on language and verbal skills and discriminatory towards illiterate people (Perone, 1991; Foxcroft et al., 2001; Weiten, 1989; Owen, 1991). A need arose for the development of more advanced, equitable and valid measures.

In response to the wide criticism, Wechsler developed performance tests that did not require verbal responses, and published his first version of the Wechsler intelligence scales in 1937. The Wechsler adult intelligence scale (WAIS) is an individually-administered measure of intelligence that is suitable for adults aged 16 to 89 (Weiten, 1989; Foxcroft et al., 2001). As Foxcroft et al. (2001, p. 13) state, “Wechsler intelligence scales yielded a variety of summative scores from which a more detailed analysis of an individual’s pattern of performance could be made.”

In the late 1900s the need to acknowledge multiculturalism arose in many countries, and this issue has continued to dominate the psychometric field in the early twenty-first century (Perone, 1991; Foxcroft et al., 2001; Foxcroft, Roodt & Abrahams, 2006; Rao & Sinharay, 2007; Van der Merwe, 2005). It became evident that most of the instruments designed in Europe or the United States were more appropriate for Westernized people and were normed on those groups. From the 1980s and 1990s the focus of
psychological testing shifted to cross-cultural test adaptation, and the new trend was cognisant that tests used internationally should be developed with a multicultural perspective in mind (Foxcroft et al., 2001).

Psychometrists have tried to develop culture-free tests such as the Culture-Free Intelligence Test (1980) (Foxcroft et al., 2001; Rao & Sinharay, 2007). However, after various attempts, it became clear that it is not possible to develop a totally culture-free test, and most test developers have therefore focused more on "culture-reduced" tests. These tests aim at removing as much of the cultural bias as possible by including only behaviour perceived to be common across all cultures (Foxcroft et al., 2001).

2.4 The misuse of psychometric tests

During the mid-1900s it was recognised that standardised tests for school children could assist in making sure that all children were receiving equal education. The tests' results were later used to underscore the need for school reform (Black & William, 1998). Haladyna, Haas and Allison (1998) state that earlier forms of the tests (used at the turn of the century) had become a form of discrimination that negatively impacted upon many ethnic groups which had recently immigrated to the United States. In the 1970s and 1980s this became a major point of contention in educational circles (Haladyna et al., 1998).

Since then, standardised testing has continued to come under attack by academics, parents, teachers and politicians. For a large portion of society, however, standardised testing remains a vague subject that is little understood (Cunningham, Brandon & Frydenberg, 1999). Perone (1991, p. 2) further points out the possible negative impact of standardised tests: "As a result of test scores, children have been labelled unready or slow learners and experts report that children in special education and lower level tracks come from lower socioeconomic populations, including large numbers of minorities who presumably test poorly in the standardised tests."
The relevance of these issues to the current research is that some of the candidates for call centre positions may not have received a solid education in numeracy. These individuals can be expected to score poorly on the CTB2 numeracy subtest, which may negatively affect their chances of employment in the call centre. If numeracy skills are not, in fact, all that important in carrying out the job of a Band A call centre agent, this would amount to an unfair assessment practice which is potentially harmful not only to the individual candidates but also to the company, because it may wrongfully screen out applicants who are well suited to the job.

2.5 Contexts in which psychometric tests are used

Psychometric tests comprise a variety of instruments. These instruments are designed to measure various aspects of an individual’s intellectual and emotional functioning, including personality traits, motivation, attitudes, abilities and intelligence. Two tests commonly used by organisations as part of their personnel selection process are personality tests and cognitive ability tests, as they both assist employers to see whether an applicant’s attributes are congruent with the organisation’s culture (Earl, 2003; Rao & Sinharay, 2007; Michell, 1997).

Personality tests aim to describe aspects of a person’s character that remain stable throughout the adult lifetime, namely the individual’s characteristic patterns of behaviour, thoughts and feelings. Personality tests are used to examine whether a person will be able to adapt to a job environment by considering these attributes, and they further examine how well the attributes will enable the individual to fit into a specific work context. Thus, by matching the right personality with right job context, institutions and businesses can achieve a better synergy and avoid the pitfalls of high employee turnover and low job satisfaction (Black, 2000; Michell, 1997; Nicholls, 2006).
Ability tests are similarly used to assess a person’s potential for learning and for their likely future achievement in a working environment. These kinds of tests are often referred to as intelligence tests and are intended to measure the overall capacity of an individual to cope with the intellectual demands made on them in a working environment. Ability tests are being increasingly used as part of the selection process, particularly for graduate positions such as internship programmes and learnerships (Black, 2000).

2.6 Characteristics of assessment instruments

An assessment measure refers to an objective standardized measure that is used to gather data for a specific purpose (Foxcroft et al., 2001, p. 4).

Assessment measures include a variety of psychological, occupational and educational assessments. The measure can be administered to an individual, a group or an organisation. Through the use of assessment measures, personality and intellectual inferences can be made about normal and abnormal behaviour. Furthermore, assessments should be administered under cautious and controlled conditions, and systematic methods should be applied to evaluate the assessment protocols.

Guidelines are available to aid the interpretation of assessment measure scores, and can easily allow for the comparison of an individual’s performance to the given norm group.

Assessment measures may vary in the manner in which they are administered: for example, the measure may vary the time limits imposed on certain test items based on the item’s level of difficulty. Depending on the qualitative and quantitative data, assessment measures may also differ in the way that they are scored. Further, they are normed differently according to the different cultural groups. Lastly, assessment measures are intended for different purposes: for example, screening job applicants as well as diagnosing clinical disorders (Fadel, Honey & Pasnik, 2007).
2.6.1 Reliability of instruments

In statistical terms, reliability refers to the consistency in the results which an assessment instrument yields over repeated trials (Yu, 2005; Kerlinger, 1986; Rosenthal & Rosnow, 1991). A reliable assessment is viewed as one which consistently achieves similar results within the same group of subjects, and it should make no difference who administers the test (Anastasi, 1988). There are various factors that may affect reliability, such as ambiguous questions or too many answer options (Yu, 2005; Owen, 1991 & 1992). Yu (2005) states that the reliability of an instrument is traditionally based on the following pointers:

- **Temporal stability**: an individual’s performance on a test should be comparable on two or more occasions.
- **Internal consistency**: an individual’s responses on a test should be consistent across questions.

2.6.2 Validity of instruments

Once the reliability of an instrument has been established, its validity must also be examined. A valid instrument is one which successfully measures that which it is designed to measure (Kerlinger, 1986; Rosenthal & Rosnow, 1991). Many of the issues which affect reliability may affect validity too, such as ambiguous wording or the use of culture-bound questions (Rosenthal & Rosnow, 1991; Owen, 1992; Anastasi, 1988).

There are different types of validity. The validity of an instrument is generally assessed through an examination of evidence in the following categories:

- **Content validity** means that the test items measure the stated objectives and do not measure irrelevant things. For example, in
personnel assessments the content of such tests must relate to an analysis of the job in question (Rao & Sinharay, 2007; Earl, 2003).

- **Criterion validity**, also referred to as **predictive validity**, means that the scores yielded by the instrument show a correlation with an outside reference. These may be scores obtained on an older and more established test designed to assess the same variable or trait, or a life outcome such as success at university or in the workplace (Rosenthal & Rosnow, 1991; Anastasi, 1988).

- **Construct validity** means that scores yielded by the instrument show that the hypothesised construct seems to be a sound concept, and a measurable one (Moskal & Leydens, 2000; Rosenthal & Rosnow, 1991).

Moskal and Leydens (2000) assert that a good assessment has both reliability and validity, although in practice an assessment instrument is never found to be totally valid or reliable. Yet these two aspects are typically viewed as essential elements for determining the quality of any test (Anastasi, 1988; Moskal & Leydens, 2000; Babbie, 2007).

### 2.7 Assessments in the workplace

In the workplace, assessment entails interviewing an individual and asking him or her to complete personality and ability tests. The process can be viewed as holistic since it involves gathering and synthesising data from different sources (Foxcroft et al., 2006). A rich and wide array of assessment information can yield holistic sampling of behaviour or functioning, as it does not focus only on the outcome of one test but across all different measures administered. The assessment battery items must be consistent with the group or organisation’s needs, as well as with the purpose of the assessment (Foxcroft et al., 2006).

Bagnato and Neisworth (1991) view also assessment as a process that should be multidimensional in nature, as it involves gathering and interpreting
data so as to understand and describe a person's overall functioning. This insight can further inform future decision-making or intervention. For example, it would not be valid to assess job skills through written tests alone; the assessment would have to incorporate other elements such as the verbal sharing of work experience or practical performance tasks (such as an "in-trayōtest) administered to the applicant (Moskal & Leydens, 2000).

2.8 Types of workplace assessment

Diagnostic assessment can be used in the workplace to establish a baseline for future assessments, or to track employees' progress (The Scottish Council for Research in Education, 1995). An initial diagnostic assessment measures an applicant's current knowledge and skills for the purposes of identifying suitable job roles for him or her (Caretta & Ree, 1997). Subsequent diagnostic assessments involve judging how well an employee is performing according to a predetermined set of criteria. The assessment must be linked to further intervention with the employee in order to address any problems that may be identified (Caretta & Ree, 1997; The Scottish Council for Research in Education, 1995).

Diagnostic tests, like any other assessment method, are not 100% accurate. Practitioners must strive to make use of multicultural and objective diagnostic assessments procedures (The Scottish Council for Research in Education, 1995; Claasen, 1997; Foxcroft et al., 2006).

Performance-based assessments are similar to summative assessments as they focus on an applicant's score achievements, and therefore indicate whether the applicant possesses the attributes required for the job (Earl, 2003).

In broad terms, assessments are categorized as being either objective or subjective. Objective assessments present the candidate with questions which have a single correct answer, such as true or false, multiple choice or
matching tasks. Subjective assessments contain questions which do not have any particular correct answer, such as essays and extended response questions (Joint Information Systems Committee, 2009).

However, the Joint Information Systems Committee (2009) indicates that the distinction between objective and subjective assessments is not entirely accurate, since it is impossible to design a totally objective assessment instrument. The Committee argues that all assessments are created with inherent biases reflected in decisions about relevant subject matter, as well as cultural, ethnic and gender biases.

Assessments are also classified as formal or informal (Valencia, 1997; Fadel et al., 2007). Formal assessments include written documents such as test results or a formally drafted report. Most formal assessments are given numerical scores, whereas informal assessments usually occur in a more causal manner. Informal assessments may include portfolio assessments and self evaluations (Valencia, 1997).

According to Fadel, Honey and Pasnik (2007), it is important to note that traditional assessment practices focus largely on the individual, and fail to account for a person’s knowledge-building and learning context—that is, the influence of the workplace itself.

2.9 Psychometric assessments in South Africa

Foxcroft, Roodt and Abrahams (2006) define psychological assessments as a process-oriented activity that is aimed at gathering a broad collection of information, by using questionnaire tests and information from other sources such as interviews. Assessment also involves the use of tests to measure actual behaviour. Psychometric assessments are tests and exercises that are designed to measure cognitive, behavioural and personality constructs of an individual (Brown, 2003).
Foxcroft et al. (2006, p. 12) state that in South Africa, a test is classified as a psychological test when the purpose of the test results in the performance of a psychological act. The Health Professions Act 56 of 1974 defines a psychological act in this context as “The use of measures to assess mental, cognitive, or behavioural processes and functioning, intellectual or cognitive ability or functioning, aptitude, interest, emotions, personality, psychophysiological functioning, or psychopathology” (Foxcroft et al., 2006, p. 114).

Tests that measure psychological constructs should be administered only by trained, qualified and registered psychologists, psychometrists or other qualified professionals (Caretta & Ree, 1997). Psychometric testing in South Africa is defined as a psychological intervention which must be performed by a psychological professional registered with the Health Professions Council of South Africa (HPCSA) (Foxcroft & Paterson, 2004). Any recruitment officer who is unregistered and performs a psychometric test is committing an illegal act. Most psychometric assessments are administered either on computer or manually through the usage of test booklets and answer sheets. Psychometric tests in many corporate environments are used to measure aptitude, personality and vocational interest (Hardy, 2009).

2.9.1 Classification of psychological tests in South Africa

Foxcroft and Paterson (2004, p. 15) state that

It is the responsibility of the developer of the test to apply to the HPCSA Board for classification, and it is the responsibility of the student in psychology, psychotechnician, psychometrist, intern psychologist, registered counsellor or psychologist to ensure that any test he/she intends to use has been classified by the Professional Board and that such a test is accompanied by a classification certificate issued by the Board.

---

2 Section 37 (2) clauses a) to e)
Although assessment tests have been done in South Africa for years, a current problem is that general tests are being used to screen people for a variety of different jobs. The tests are thus neither designed nor validated for use in assessing people’s ability to do these particular jobs (Foxcroft & Paterson, 2004).

Contemporary South African assessment tests are governed by the legislative guideline given in Section 8 of the Employment Equity Act No. 55 of 1998. This clearly states that “Only assessment tests that have provided evidence in support of validity and reliability can be used.” Therefore tests must have demonstrated that they are not racially biased but are fair, showing consideration for all individuals and racial groups (Professional Board of HPCSA, 1998). Furthermore, the HPSCA is the only legal entity in South Africa that can classify and register psychological tests, as discussed by Foxcroft and Paterson (2004, p. 15):

The HPSCA is the only national statutory body responsible for inter alia classifying, registering and reviewing the use of psychometric and psychological tests, questionnaires, apparatus and instruments used in order to determine intellectual ability, aptitude, personality make-up, personality functioning, psycho-physiological functioning and psychopathology.

According to van der Merwe (2005), all psychometric tests are developed to make presumptions about people, and validity plays a role in assuring the accuracy of these presumptions. Beyond merely making presumptions, test results are also used to inform important decisions about the employment of an individual. It is thus imperative to ascertain the validity of the test items (van der Merwe, 2005).

2.9.2 Psychometric assessments pre-1994
In 1948, the policy of Apartheid was put into practice in South Africa, dictating where people could live, work, learn and stay. Assessment measures used under the Apartheid regime impacted very negatively on test development and usage in South Africa, partly because separate tests were designed and used for candidates from different racial categories. In addition, the practice also arose of using tests developed specifically for a Westernized population on people from other cultural groups, thus putting those other cultural groups at a disadvantage because the instruments had not been normed on them (Brown, 2003; Owen, 1992).

South Africa was originally a British colony, and the use of psychometric assessments may have stemmed from this colonial heritage (Claasen, 1997). The development of psychometric tests in South Africa has largely been characterized by the unequal distribution of resources based on racial classification; that is, the development of psychometric instruments reflected the racially segregated society in which they evolved (Claasen, 1997). The earliest psychometric measures in South Africa were standardised for white people only, and were used by the Department of Education and the Human Sciences Research Council (HSRC) to identify white school pupils who needed special education. These measures could be traced back to foreign instruments such as the Stanford-Binet; the South African revision was called the Fick Scale (Claasen, 1997).

Claasen (1997) asserts that the use of assessment tests in South Africa cannot be separated from the country’s political and economic history. The use of psychometrics was driven partly by political ideology in an attempt to show that white people were superior to other racial groups (Nzimande, 1995). According to Christie (1991, p. 9),

The extension of the University Education Act, Act 45 of 1959, put an end to black students attending white universities (mainly the universities of Cape Town and Witwatersrand). Separating tertiary institutions according to race, this Act set up separate "tribal colleges" for black university students. The so-called "bush Universities such as
Fort Hare, Vista, Venda, and Western Cape were formed. Blacks could no longer freely attend white universities.

Measures standardised for white people created or maintained the perception that African students should be educated in a way appropriate to their black culture. Assessments were thus misused as yet another way of preventing African students from receiving an education that would fulfil their aspirations and allow them to take up the careers they wanted to follow. Instead, African children were to receive an education designed to provide them with skills to serve their own people in the Bantustan "homelands" or to work in manual labour jobs under white control (Christie, 1991).

The use of psychometric measurement in South African industry gained momentum in the post-World War Two era, after the National Party came into power in 1948. This was largely in response to an urgent need to identify the occupational suitability of mine workers, of whom many were Africans who had received little formal education. One of the measures used at the time was the General Adaptability Battery (GAB), which comprised practice sessions during which test-takers were familiarized with the concepts required to solve the test problems. The test was predominantly used on preliterate black populations (Foxcroft et al., 2006).

The white population working in industry was tested using the Otis Mental Ability Test, which had been developed in the United States of America and was consequently biased towards American norms. Only white South Africans were assessed using this test, since whites were the only racial group allowed to apply for positions which required higher education skills (Claasen, 1997). Owen (1991, p. 112) notes that in the racially segregated South Africa of the Apartheid era, it was almost inevitable that psychological measures would be developed along cultural racial lines as there was little specific need for common tests because various groups did not compete with each other.ö
2.9.3 Psychometric assessments post-1994

During the 1980s and early 1990s, and especially after the first democratic election in 1994, the political landscape began to change and discriminatory laws were repealed. People from different racial groups began competing for the same jobs and the use of separate measures came under close scrutiny. In an attempt to solve the problem, instruments normed on more than one racial group were developed. Such tests include the General Scholastic Aptitude Test (GSAT), the Ability Processing of Information Learning Battery (APIL-B), and the Paper and Pencil Games (PPG) which was the first test to be available in all 11 official South African languages (Foxcroft et al., 2006).

The political changes in post-Apartheid South Africa (i.e. over the last 16 years) have been accompanied by drastic reforms in the education system. The most noticeable changes include the desegregation of schools through the development of the National Qualifications Framework and Outcomes Based Education, and the adoption of new language policies (Department of Education, 1996).

Foxcroft et al. (2001, p. 20) point out that the HPCSA is mandated to protect the public and to guide the profession of psychology. Since 1994, the board of psychology at the HPCSA has become increasingly aware of the ways in which psychometric tests have been misused in South Africa historically. Nonetheless they recognise the important role of this field within professional psychological practice, and for the purposes of research (Foxcroft et al., 2001). Psychometric assessments post-1994 have been gaining momentum in the South African recruitment industry, with recruiters and employers using psychometric testing during candidate selection, organisational development and promotional processes (Hardy, 2009).

The HPCSA has encouraged test developers to acknowledge that many underprivileged South Africans cannot compete with the privileged sectors of society in accessing positions in organisations that use psychometric assessments as part of their recruitment process (HPCSA, 2002). South
Africa is still in the process of transforming and standardising psychometric tests across all racial groups; other current problems include language barriers. Assessment practitioners should consider these factors when conducting psychometric assessments (HPCSA, 2002).

The Employment Equity Act No. 55 of 1998 (Section 8) offers the strongest stance against the improper use of assessment measures. Foxcroft et al. (2001, p. 18) state that "[U]nder the adoption of the new Constitution and the Labour Relations Act in 1996, workers unions and individuals now have the support of legislation that specifically forbids any discriminatory practices in the workplace." The wording of the relevant part of the Employment Equity Act is as follows:

Psychological testing and other similar forms of assessments of an employee are prohibited unless the test or assessment being used has been scientifically shown to be valid and reliable... [that it can] be applied fairly to all employees and it is not biased against any employee or group (cited in Foxcroft et al., 2001, p. 19).

The Act has implications for assessments in South Africa as many of the instruments currently used here are imported from the United States or Europe. A few instruments have been developed locally, but the imported tests are often not checked properly for test bias or cross-cultural validity (Foxcroft et al., 2001).

However, according to Foxcroft et al. (2001) the Employment Equity Act itself has several flaws, and consequently many psychometrists (both test developers and administrators) have been left uncertain about the Act’s effectiveness. One of the challenges is that test developers and publishers are supposed to certify their measures as being Employment Equity Act Compliant (Foxcroft et al., 2001). This step is intended to help practitioners know which measures are compliant and which are not, so that they can make informed choices. The Act, however, fails to recognize that some publishers assert that they have certified their tests when in fact they have
not. Furthermore, even if the instruments have indeed been certified, this does not guarantee that test results will not be misused when human resource personnel decide whom to employ (Foxcroft et al., 2001).

Hardy (2009) states that globally psychometric assessments are becoming very useful as part of recruitment and development. However, problems emerge when tests are used for incorrect purposes, such as using tests that are not relevant for the recruitment group being assessed. This problem is common when the test administrators have little knowledge about psychometric assessments. In order for people to administer tests accurately, they need to be very clear about the test’s intended purpose (Hardy, 2009).

In South Africa, psychometric tests are widely applied in educational and employment assessments to measure abilities in domains such as reading, writing, creative thinking, critical reasoning and numerical abilities such as the clerical and numerical ability tests.

2.9.4 Legislative changes

The adoption of the new Constitution enhanced equality and human dignity among all people in South Africa. The rights contained in Section 9 of the Bill include the following subsections:

(3) The state may not unfairly discriminate directly or indirectly against anyone on one or more grounds, including race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth.

(4) No person may unfairly discriminate directly or indirectly against anyone on one or more grounds in terms of subsection (3).
(5) Discrimination on one or more of the grounds listed in subsection
(3) is unfair unless it is established that the discrimination is fair.³

(Employment Equity Act, No. 55 of 1998, pg. 6)

The South African Employment Equity Act states that every employer must take steps to promote equal opportunity in the workplace by eliminating unfair discrimination in any employment policy or practice (Employment Equity Act, No. 55 of 1998, p. 6). This clause has direct relevance for psychometrists and human resource personnel. The purpose of the Employment Equity Act is to ensure equity in the workplace by:

(a) promoting equal opportunity and fair treatment in employment through the elimination of unfair discrimination

(b) implementing affirmative action measures to redress the disadvantages in employment experienced by designated groups in order to ensure their equitable representation in all occupational categories and levels of the workforce.

(Employment Equity Act, No. 55 of 1998, p. 7)

2.10 Use of psychometric tests at the company under study

The use of psychometric assessments is well established in large organisations in South Africa, but has generally been disregarded in smaller organisations, where such tests are often viewed as being expensive and meaningless (Hardy, 2009).

The company under study is one of the biggest organisations in southern Africa; as noted earlier, since 2001 it has been using psychometric assessments for recruitment and staff development programmes. Candidates go through one-on-one and panel interviews, in addition to completing ability

³ Such as legislative or administrative steps taken to redress historical imbalances between the race groups (Section 9, subsection 2).
and personality tests. The tests are used to draw conclusions about whether the candidate can adapt to the company culture and carry out the duties required. Assessments are also used in the company for talent forums. In this regard they are used to help evaluate the promotional eligibility of staff members, to indicate whether an employee still has interest in a particular work area, or to assess whether an employee would like to expand into other work areas that are more challenging, thereby encouraging career growth (Hardy, 2009).

There are challenges associated with internal assessments, such as the complaint of some employees that the test results do not give a true reflection of their potential. Line managers at the company have also questioned the interpretation of assessment results, specifically the results of the ability test (Cognitive Process Profile). However, assessment test results are not used in isolation to draw conclusions about an employee's career growth, but are used in conjunction with other measurement tools. The concerns of line managers fuelled discussions about the possibility of using more than one test in assessing candidates, as some tests might be seen as being culture sensitive, thereby disadvantaging some applicants and internal employees (Hardy, 2009).

Most corporate sectors, including the company under study, establish their own assessment centres where batteries of psychometric assessments are given to candidates, along with competency-based tasks which are observed by an assessor. The aim of the assessment centre is to uncover which applicant has the most suitable personal attributes, problem solving ability and general aptitude for a specified post. The assessments also aim to determine which candidate would be most suitable from an organisational culture and structural point of view (Ngidi, 2009).

Most corporate organisations also use interviews, which allow the assessor to gain a large amount of relevant information in a relatively short time. The interview and assessment scores together give the assessor a holistic view of the potential job applicant, which ensures that the most suitable candidate is
chosen for the job. Hardy (2009) asserts that psychometric assessments add considerable value to the processes of candidate selection and employee development. Assessments should be viewed as an opportunity to learn more about employees and the jobs they are either employed to do, or positions they are applying for currently.

Assessments were introduced at the company under study in the period 2000 to 2002 and were initially confined to recruitment. At that time a high need for call centre agents was identified, and candidates were assessed using the CTB2 to try and improve the chances of recruiting the right person for the job. The application of assessments has since been broadened in the company to cover talent management, development centres for training, and personal development for purposes such as team building. In all divisions, internal and external job applicants at the company are assessed. The assessment tools are generic across positions and are administered according to job levels (the company’s database, 2006). Though many different assessment tests are used by the company, this study focused only on the clerical test battery (CTB2) numeric subtest, which at the time of the research was being used to assess call centre candidates.

Though many different assessment tests are used by the company, only the Clerical Test Battery (CTB2) numeric subtest was examined in the current research. It was important to ascertain the validity and reliability of the CTB2 when used with call centre agents, to try and establish whether call centre applicants were being tested using an instrument that measures skills that are relevant to call centre job requirements. The test content needed to verified as being in line with the call centre job functions, and the current correlational study would shed light on this question.
2.11 The use of the Clerical Test Battery (CTB) in South Africa

2.11.1 History of the CTB2 in South Africa

The CTB2 was developed by Psytech International in 1991 and later adopted by Psytech-South Africa in 1995, and was introduced to the South Africa public that year. Initially its use was very limited as the companies which introduced the test generally stipulated their own norms of psychometric assessment (Psytech-SA, 2002).

One of the main reasons for the usage of the CTB2 in South Africa was the implementation of affirmative action as stated in the Employment Equity Act 55 of 1998. The test battery was submitted to the Psychometrics Committee of the Health Professions Council of South Africa in April 2001. The process of developing, evaluating and classifying a test battery is a lengthy one. Feedback reports received from the test reviewers indicated that the CTB2 met the requirements for classification as a psychological battery, and as such was valid for use with South Africans (HPCSA, 2002). Currently, many employers in South Africa use the CTB2 in conjunction with affirmative action recruitment and selection practices (Psytech-SA, 2002). Psytech-SA offers users help with establishing their own South African norms where necessary, and it also encourages users to submit their findings to Psytech so as to add to the data pool (Psytech-SA, 2010).

The Clerical Test Battery must be used under the supervision of a psychologist. When purchasing test materials or scoring services related to the CTB2, the signature and registration number of a psychologist or registered counsellor is required (HPCSA, 2002).
2.11.2 Level of education required for completing the CTB2

The CTB2 is intended for respondents with a Grade 12 education level. However, it must be kept in mind that educational standards differ considerably in South Africa. People from impoverished backgrounds have not had the same standard of education as those from privileged backgrounds. Psychologists doing assessments should evaluate the situation with care, taking into account the respondent’s background, fluency in English, as well as other relevant factors that could influence his or her test performance (Psytech-SA, 2002; 2010).

According to Psytech-SA (2002, p. 8), “The CTB2 provides a short yet comprehensive assessment of a broad range of core clerical aptitudes and abilities as it is appropriate for all people who have achieved a basic level of education and are being considered for clerical and junior administrative posts. Although it is quick to administer, the CTB2 is an essential battery for ensuring that clerical and administrative staff has the core skills and aptitudes that are required for successful performance.” (Psytech-SA, 2002, p. 8; see also Psytech-SA, 2010). The battery is a useful tool for assessing staff for recruitment, promotion and training (Psytech-SA, 2002).

2.11.3 Standardisation of the CTB2 in South African retail companies

The CTB2 has been tested and standardised in the South African context, with both reliability and validity studies having been undertaken (Psytech-SA, 2010). The South African norm groups that have been researched include call outbound centre staff and applicants, retail and administrative staff, a general population sample, clerical applicants (black, white, Asian and coloured) to the police service, and different racial groups of candidates applying for retail positions (black, Asian, white and Coloured) (Psytech-SA, 2010).
2.11.4 Design of the Clerical Test Battery

The battery consists of four tests bound into one booklet, with responses being recorded either on computer or on a written answer sheet (Psytech-SA, 2002). The battery assesses a range of clerical aptitudes and skills, namely verbal reasoning, numerical ability, clerical checking, and spelling. The CTB2 is designed for general clerical and administrative positions—the tests can either be administered on their own to assess one specific aptitude, or the whole battery can be administered to produce a candidate profile.\(\text{(Psytech-SA, 2002, p. 8)}\). It is important to note this flexibility in the context of the current research, given that the use of the numeracy subtest specifically was being questioned.

Psytech-SA has made feedback and norms about the CTB2 available to anyone who uses the tests, and decision-maker and candidate feedback reports are provided as part of the test. In the report intended for decision-makers, profiles about the candidate's present raw score, sten and percentile ranks against clerical norm groups\(\text{(Psytech-SA, 2002, p. 9)}\).

Psytech-SA (2002; 2010) explains the subtests of the CTB2 as follows:

- **Verbal Reasoning** measures basic vocabulary, verbal fluency and the ability to reason using words. This test is appropriate for all clerical and administrative jobs that require a basic level of verbal ability.

- **Numerical Ability** measures the ability to use numbers efficiently in clerical and administrative contexts. This test assesses the ability to perform tasks such as calculating travelling expenses and working out the unit pricing of goods. Thus it is useful for assessing existing and prospective clerical staff.

- **Clerical Checking** assesses the ability to quickly and accurately check verbal and numerical information such as names, addresses, code numbers and telephone numbers against a target. A classic precision
test assesses the ability to quickly and accurately code data— a skill that is essential for most clerical positions.

- **Spelling** assesses the ability to correctly spell commonly misspelt words.

- **Typing** is an on-screen application only. The typing test presents three user definable passages of text to be typed by the respondent. The test measures speed and identifies the number and nature of errors made.

- **Filing** is an on-screen application only. The filing test assesses the ability to classify names quickly and accurately into an existing electronic alphabetical filing system.

Numerical ability is the only component that will be discussed in this study. It measures numerical reasoning, which has been identified as a component of general reasoning. In this test candidates have to draw inferences from numerical information which is presented in tabular form. Each question has six possible answers and candidates have to pick the correct answer (Psytech-SA, 2002).

### 2.12 Theoretical model of competencies

Many tools and models exist to help organisations recruit the right talent and retain well-performing employees. When theories are implemented correctly, they can enhance an organisation's performance by tapping into employees' skills, abilities, wants and needs (Competency Model Clearinghouse, 2008, p. 1).

The research paradigm employed for this study was positivism. The positivist approach assumes that only observable behaviour can be studied and interpreted (Meyer, Moore & Viljoen, 1989). In the current study, this
observable behaviour refers to job performance within the retail sector, as well as performance on the CTB2. The positivist approach was integrated with the competency theory as both theories break down the behaviour, attitudes and skills of an individual into measurable pieces. Both theories give focus to ways of combining these pieces together to create a picture of an individual who can demonstrably improve his or her performance standards (Mascarenhas, Baveja & Jamil, 1998).

Competency theory asserts that a collection of competencies work in combination to create successful performances in a particular work environment. Competencies play a vital role in the recruitment, hiring, training and development of employees. Organisations develop competency-specific jobs and job groups (Competency Model Clearinghouse, 2008).

According to Murphy (2002), knowledge and skills are surface competencies that can be easily measured and learnt, while psychological attributes (attitudes, values, capabilities and behaviour) are considered to be hidden and more difficult to develop. Psychometric tools are used to measure these attributes (Psytech, 2002). The company uses the CTB2 and various other tests to understand more about the hidden psychological attributes of candidates (Psytech, 2002).

Competencies often serve as the basis for skill standards that specify the level of knowledge and skills required for success in a workplace. Competencies also serve as the potential measurement criteria for assessing competency attainment (Dalton, 1997).

Competency is therefore seen as fundamental knowledge, ability or expertise in a specific subject area or skill set. When conducting assessments, it is important to distinguish between individual competencies and core competencies. Core competencies are more than the traits of individuals. They are conceptualised as being the aggregates of capabilities, where synergies are created that have sustainable values and broad applicability. Competency in organisations such as the company under study is viewed as
an employee’s capacity to apply his or her knowledge, skills and abilities successfully in the performance of critical work functions (Gary & Prahalad, 1990; Dalton, 1997).

The company under study employs the following competency model:

The first (base) block of the pyramid consists of personal competencies, which differ between individuals. People use their personal competencies in all areas of their lives; for example in their role as a sibling, parent, spouse, member of a community, student and an employee of a certain organisation. These competencies are learnt at home, in society or at a previous workplace. They contribute to the success of a person’s career or role in the workplace. Employers value these personal competencies as they provide an idea of the
type of person one is, and whether or not one would fit into or even thrive within the organisation (Competency Model Clearinghouse, 2008).

Personal competencies are closely followed by academic competencies which are learnt in academic settings. Academic competencies assist the employer to understand the employee's thinking styles and cognitive functioning (Competency Model Clearinghouse, 2008). Academic competencies are often viewed as the foundation for industry-specific outcomes, as they are the most important and decisive factor to an employer hiring employees. Employers want to hire people with the ability to read and write, and in many cases employers need candidates who specialise in a certain field of work such as accounting, marketing, or the sciences.

Even though most applicants possess the right academic and personal competencies, employers are next obliged to ascertain their workplace competencies, as these are the competencies that enable individuals to function effectively in an organisational setting. Workplace competencies are generally applicable in occupations and industries; employers expect an applicant to focus on his or her job, work well with team members, be able to adapt easily to different settings, and have a sharp customer focus (Competency Model Clearinghouse, 2008).

Organisations also put strong emphasis on technical competencies as they represent the knowledge, skills and abilities that are needed from employees. These competencies differ across industries: for example, in the company under study they include supply chain management, quality assurance, and style and design skills (Competency Model Clearinghouse, 2008).

Industries specialise in different merchandise or products, therefore an occupation-specific knowledge competency is essential to the growth of organisations. Employees are required to have a specific knowledge base which can be used in detailed knowledge areas required for a specific job. These competencies tie in with the occupation-specific technical competencies which define job descriptions and expected deliverables. These
competencies are mostly articulated by shareholders and management (Competency Model Clearinghouse, 2008).

The occupation-specific requirement indicates that an employee must possess sufficient job experience, degree of education and training. These competencies differ according to different occupational level requirements.

Management competency is the highest in the hierarchy, and comprises supervisory and managerial occupations. Employees on this level are expected to delegate, network, lead staffing procedures, monitor work, direct strategic planning and so on (Competency Model Clearinghouse, 2008).

The assessment department at the company does not view individual competencies in isolation. Competencies are instead viewed as combinations of complementary skills and knowledge bases embedded in a group, resulting in the group’s ability to carry out one or more of the critical processes in business. When interpreting assessment results, it is always important to measure people’s ability to function optimally. The skills, knowledge and psychological attributes (e.g. behaviour, attitudes and values) of candidates must be complementary; working together they should provide superior competency (Mascarenhas et al., 1998).

**2.13 Reliability of CTB2 in the South African context**

The next chapter (Chapter 3) provides an outline of reliability and validity studies, and the requirements for carrying out such research. That information will not be repeated here and the reader is requested to refer ahead for a description of the methodology of these kinds of studies, and why they are necessary. In this section the literature relating to reliability or validity studies on the CTB2 in a South African context is reviewed, and the results of each study are summarised.
2.13.1 *The CTB2 in a retail environment*

This research is reported on the Psytech-SA (2010) website. The original study was conducted in 2002 in the South African retail context at a major clothing retail group in Kwazulu-Natal and the Western Cape (Psytech, 2002; 2010). The aim of the research was to check the internal consistency reliability of the CTB2 subtests, and to establish norms for this type of South African population.

The results of a sample of 2,429 job applicants were analysed, with different races, genders, ages and levels of work experience being included in the sample (Psytech SA, 2010; case studies N6 and R4). The majority of respondents were African (52%), with the rest of the sample including Coloured (20%), white (2%) and Asian (21%) respondents, with the remainder of the respondents not being clearly classified racially. The sample was mostly female (69%) with a relatively small number of males (31%). Level of formal education was not consistently reported, with 65% of participants having an unknown level of schooling. The sample's mean age was 25.9 years. All language groups were represented but inadequate data was obtained for this demographic variable (missing in 72% of cases).

The study was carried out by Psytech and was motivated by the agency's need to standardise the test result across different population groups. Thus the validity and reliability of the instrument had to be studied in a retail context consisting of different races in order to include all norm groups. To ascertain whether the battery discriminated against any racial group, the CTB2 results of participants from all racial groups were compared. A descriptive quantitative research design was used. The data analysis involved running correlational tests to establish if any relationships existed between the different subtest scores or racial groups (Psytech, 2002).

The internal consistency reliabilities for the CTB2 subtests is shown in Table 1 on the next page.
A Cronbach’s alpha of 0.70 or more indicates good internal consistency reliability, while an $\alpha$ of 0.85 to 0.90 is excellent (Rosenthal & Rosnow, 1991). In the reported study, all subtests showed good or very good internal reliability, with the numerical ability subtest having an $\alpha$ of 0.73 (Psytech, 2010).

However, the above results were obtained from analysing the data for all combined race groups. The internal reliability of the subtests was also assessed for each racial group separately, and the results were not as impressive. Although $\alpha$ was calculated for all subtests, only the numerical ability subtest results are shown below. All data is from the Psytech South Africa website (2010) and the findings relate to the same 2002 study of candidates in the retail industry.

<table>
<thead>
<tr>
<th>Racial group</th>
<th>N</th>
<th>Mean age</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>1253</td>
<td>26.3</td>
<td>.61</td>
</tr>
<tr>
<td></td>
<td>(M= 394; F = 866 Unknown = 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>502</td>
<td>25.2</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>(M=163; F=343)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>51</td>
<td>30.2</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>(gender breakdown not available)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coloured</td>
<td>478</td>
<td>24.8</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>(M=135; F=344)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Cronbach’s alpha for racial sample groups (Numeracy subtest only)
(source: Psytech website, CTB2 reports R5; R6; R7; R8).
It can be seen from Table 2 above that while the numeracy subscale $\alpha$ was highly acceptable when the test was administered to white respondents, the $\alpha$ was unacceptably low when the test was used with black respondents. The Asian subgroup results yielded an $\alpha$ of 0.75, which is also acceptable, but the results for the Coloured group were again unacceptable at 0.66. This study conducted by Psytech in 2002 showed that internal reliability of the CTB2 numeracy subtest was adequate when the respondents were white or Asian but not when they were black or Coloured respondents.

### 2.13.2 The CTB2 in a call centre context

The Psytech website (2010) gives the data from a study that was done with 184 applicants for outbound call centre sales positions in 2001. The company dealt with insurance and the minimum level of education that was required was Grade 12. Most of the respondents were white or Coloured (54%), with a substantial number of blacks (36%) and some Asians (6%). Racial information was missing for 4% of respondents (Psytech-SA, 2010). Males comprised 37% of the sample and females 59%, with 4% of respondents missing gender classification. The overall sample’s mean age was 25.8 years.

The internal consistency reliabilities for the subtests are shown in Table 3 below.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Cronbach alpha ($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>.83</td>
</tr>
<tr>
<td>Numerical ability</td>
<td>.79</td>
</tr>
<tr>
<td>Clerical Checking verbal</td>
<td>.76</td>
</tr>
<tr>
<td>Clerical Checking numerical</td>
<td>.78</td>
</tr>
<tr>
<td>Clerical Checking total</td>
<td>.88</td>
</tr>
<tr>
<td>Spelling</td>
<td>.81</td>
</tr>
<tr>
<td>Mean alpha</td>
<td>.81</td>
</tr>
</tbody>
</table>

**Table 3: Cronbach alpha for the CTB2 with a South African call centre sample; N = 184**

(Source: Psytech, 2010; case study R1)
All of the reported alphas reached or exceeded the required 0.70, and the numerical subtest had an $\alpha$ of 0.79, which is perfectly acceptable.

Similar results were found in a second study conducted on 192 applicants and incumbents at an outbound call centre in the insurance industry, also in 2001. In this case the sample was 38% white and Coloured, 20% Asian and 41% black (data missing for 1%). Females made up 49% of the sample and males 50% (data missing for 2%). Mean age was 25.8 years. The internal reliability scores were as follows:

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Cronbach alpha ($\alpha$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>.79</td>
</tr>
<tr>
<td>Numerical ability</td>
<td>.87</td>
</tr>
<tr>
<td>Clerical Checking verbal</td>
<td>.90</td>
</tr>
<tr>
<td>Clerical Checking numerical</td>
<td>.91</td>
</tr>
<tr>
<td>Clerical Checking total</td>
<td>.95</td>
</tr>
<tr>
<td>Spelling</td>
<td>.85</td>
</tr>
<tr>
<td>Mean alpha</td>
<td>.88</td>
</tr>
</tbody>
</table>

Table 4: Cronbach alpha for the CTB2 with a South African call centre sample; N = 192
(Source: Psytech, 2010; case study R2)

It can be seen from Table 4 above that all of the reported alphas were highly acceptable. This indicates that the entire CTB2 and (for the purposes of the current research) specifically the numerical subscale, could be used without any modifications with this kind of population. It is noted that in both studies cited in this section, the samples were composed of a good mix of people from different racial groups. However, no data was available for the separate race groups, which is unfortunate given that the good results obtained with one race group might mask the poor results found with a different race group, as in the previous example (see 2.13.1 above).

It is also noted that the results of these two studies conducted in the insurance sector do not quite correspond with the findings obtained in the
retail sector (see 2.13.1). The retail sample yielded a numeracy subscale $\alpha$ that was inadequate in the case of black and Coloured respondents. Further investigation into this area was therefore required, and the aim of the present study was to carry out such research using a different sample that was drawn from a retail call centre.

2.13.3 Correlation between CTB2 and Occupational Personality Profile scores

A correlational study was conducted by Psytech-SA in 2002 with a group of MBA students at a postgraduate business school in Gauteng. The students’ scores on the CTB2 and a separate instrument, the Occupational Personality Profile (OPP), were correlated in an attempt to explore the construct and content validity of the CTB2.

The students were selected based on their academic performances, interviews and psychometric test results (Psytech-SA, 2002). The majority of the students in this study were white (82%), while African (9%), Indian (8%) and Coloured people (1%) made up the remainder. Most participants (70%) were Afrikaans; 43% were male and 57% were female. The CTB2 subscales used for data analysis were the Critical Reasoning Test, consisting of verbal and numeric reasoning subtests.

Findings showed that the OPP and numeric reasoning subtest scores both correlated strongly with the students’ academic scores (Psytech-SA, 2002). However, the fact that students were pre-selected for their academic achievement and work experience may have reduced the variance for ability measures (Psytech-SA, 2002).
2.14 Validity of CTB2 in the South African context

2.14.1 Correlation between CTB2 and LPCAT

A validity study was conducted in 2002 with the CTB2 across different race and gender groups. The sample included 892 respondents, with the racial composition being 34% Asians, 4% Coloured, 55% African, 2% European and 6% unclassified. Males comprised 30% of the sample, and females 70% (Psytech-SA, 2010). Mean age was 25.7 years. Respondents were all applicants for various jobs in the retail clothing industry, including temporary and casual work, in KwaZulu-Natal in 2002 (Psytech-SA, 2010).

In an attempt to establish the content and construct validity of the CTB2 (namely whether it measures what it is designed to measure), CTB2 scores were correlated with participants’ scores on an independent test, the Learning Potential Computer Assisted Test (LPCAT). Since 2006 the retail company in question had been using the LPCAT in conjunction with other tests to assess candidates for job Bands A and B. By not using one test battery in isolation, the test results were more consistent and informative (de Beer, 2007).

Respondents were given clear instructions on how to complete both tests and the time-frame for each test. The results were as shown in Table 5 below.

<table>
<thead>
<tr>
<th>CTB2 Subtest</th>
<th>LPCAT composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checking numerical</td>
<td>.31 *</td>
</tr>
<tr>
<td>Checking verbal</td>
<td>.37 *</td>
</tr>
<tr>
<td>Checking total</td>
<td>.35 *</td>
</tr>
<tr>
<td>Numerical ability</td>
<td>.40 *</td>
</tr>
</tbody>
</table>

* = significant at p < 0.5

Table 5: Correlations between CTB2 subtests and LPCAT composite score; N = 892

(Source: Psytech, 2010; case study V1)

All correlations between the CTB2 subtests and the LPCAT composite scores were positive and statistically significant at p < 0.5. According to the
interpretation of the results as given by Psytech, this may reflect an underlying general ability, such as mental alertness or general intelligence (Psytech, 2010, case study V1).

2.14.2 Correlation between CTB2 subtests

Psytech-SA (2002; 2010) also validated the CTB2 in 2001 by calculating the inter-correlations of the subscales. The sample consisted of 192 respondents, all of whom were job applicants at a call centre selling insurance. The majority of the respondents were black (41%), with whites and Coloureds comprising 38%, and Asians 20%. The sample included 49% females and 50% males, with 1% being unknown. Mean age was 26 years. The results are shown in Table 6 below.

<table>
<thead>
<tr>
<th>Subtest</th>
<th>VR</th>
<th>NA</th>
<th>CCV</th>
<th>CCN</th>
<th>CCT</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Reasoning</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerical ability</td>
<td>.61</td>
<td>*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerical Checking verbal</td>
<td>.30</td>
<td>*</td>
<td>.43</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerical Checking numerical</td>
<td>.26</td>
<td>*</td>
<td>.44</td>
<td>.92</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Clerical Checking total</td>
<td>.28</td>
<td>*</td>
<td>.45</td>
<td>.98</td>
<td>.98</td>
<td>1.00</td>
</tr>
<tr>
<td>Spelling</td>
<td>.41</td>
<td>*</td>
<td>.28</td>
<td>.14</td>
<td>.11</td>
<td>.13</td>
</tr>
</tbody>
</table>

* = significant at $p < 0.5$

Table 6: CTB2 subscale correlations, South African call centre sample; $N = 192$
(Source: Psytech, 2010; case study V2)

The Psytech (2010) interpretation of the high correlation (.61) between verbal reasoning and numerical ability is that both tests rely on reasoning ability. The subscales of the checking test (i.e. CCV and CCN) correlate extremely highly (.92), indicating that they are measuring almost the same construct. The very high correlations between the two checking subscales and the checking total (.98 in each case) may be explained in the same way.

Spelling and verbal reasoning were moderately correlated (.41) which indicates that both tests may be tapping into the same underlying ability, probably verbal or language ability (Psytech, 2010). The correlation between
spelling and checking tests was weak (ranging from 0.11 to 0.14 ) thus indicating that spelling skills involve a different mental process from merely paying attention to detail (Psytech, 2010).

The subtest of particular interest in the current research is the numerical ability test. In the 2001 study this score showed a moderate correlation with the checking tests (ranging from 0.43 to 0.45), suggesting that the ability to pay attention to detail is important to both types of ability. The numerical scores correlated weakly (.28) with spelling, which indicates that the constructs of numerical ability and spelling ability are largely independent of each other.

2.15 Personality and creativity: a psychometric study

Naude (2005), at the University of Pretoria in South Africa, conducted a study in which the main aim was to test the relationship between personality and creativity. The psychometric tool used to test creativity in people’s performance was the 16 Personality Factor (16PF) South African version and the Abbreviated Torrance Test for Adults. The 16PF is used to assess normal personalities in multiple settings; it is often used for personnel selection, career planning and career development (Naude, 2005).

The majority of the respondents in this study were white (76.7%), while African (16%), Indian (3%) and Coloured people (0.02%) constituted the rest. The sample was mostly Afrikaans (60%), ad mostly female (82%). The ages of participants ranged from 21 to 23 years. At the time of the study they were all in their fourth year of study in the psychology department at the University of Pretoria (Naude, 2005).

Individuals with high creativity scores, as measured by the Abbreviated Torrance Test for Adults, indicated high creativity levels in most of the constructs measured in the test. It was concluded from the results of the study that 87% of creative individuals tend to perceive themselves to be significantly
more independent in thought, more open to experience, and more dominant, individualistic and competent than individuals with average and low creativity levels (Naude, 2005).

According to Feldhusen (1995), personality factors like self-confidence constitute a state within which creative behaviour can most readily take place, while it might even serve as a facilitator or stimulator of cognitive creative processing.

It was further concluded that creative people perceive themselves to be independent thinkers with a high competency level who are open to experience, thus explaining why they would perform well at work even though their psychometric (numeracy) results might often be low. Naude (2005) states that it often happens that people who perform poorly in psychometric tests are nonetheless highly creative, and in many instances become leaders in their workplaces. Therefore, according to Naude (2005), creativity must be seen as something which occurs in various contexts through the interaction of many important dimensions. The implications of this complex situation should not be overlooked when using the CTB2 to assess an individual’s ability to contribute to the workplace.

### 2.16 A review of South African research into dynamic assessments

The aim of a study conducted by Murphy (2002) from the University of Pretoria was to investigate psychometric tests used in the South African context for assessment purposes. One specific study reviewed by Murphy is discussed in this section. The study was conducted by Hoffenberg (1988) with a focus on gifted children from disadvantaged learning environments (Murphy, 2002). To carry out this study, control and experimental groups were used. The experimental group received mediation on four subtests of the Feuerstein Learning Potential Assessment Device (LPAD), while the control group was not exposed to mediation on the LPAD task.
Two hypotheses were stated, the first assuming that the group receiving guidance on the LPAD would perform significantly better than the control group; the second hypothesis assumed that the transfer measures of the experimental group would be higher than those of the control group. However, no statistically significant differences were found between the two groups’ results overall. Out of the four subtests completed, the experimental group outperformed the control group in only one of the subtests (Murphy, 2002).

Hoffenberg concluded that the results of the study were not encouraging in using the LPAD in the context of identifying disadvantaged gifted learners (Murphy, 2002). Murphy further concluded that it is important to use more than one battery to show how locally developed and normed tests can be applied in a variety of contexts. It was established from Hoffenberg’s results that assessment tests show more conclusive findings when more than one instrument is used (Murphy, 2002).

2.17 Conclusion

This chapter started by introducing the concept of psychometrics, and outlined the history of this field both globally and within a South African context. The development of intelligence tests and the broader category of job-related psychological assessments was explored. The implications of South Africa’s shift from Apartheid to the new democratic constitution, and the status of psychometrics under the regulations of the Health Professions Council, were also discussed. Some theories used within the field of assessments and their application in corporate organisations were presented.

The chapter further looked at the CTB2 as the focal point of the current research. Studies done by Psytech-SA (2002; 2010) regarding the reliability and validity of the CTB2 in the South African context were reviewed. The evidence suggested that while the instrument is generally reliable when used with South Africans, there may be problems when it is used with certain race groups, notably those who were historically disadvantaged and thus perhaps
had poor education. Psytech’s findings in this area have not been entirely consistent, with some reliability indices being acceptable while others were not. The current research aims to further the knowledge and data pool about this aspect of the CTB2, particularly with regard to the use of the numeracy subtest with call centre job applicants.

The use of the CTB2, along with other assessment instruments, is likely to impact extensively on the decisions a company or business makes regarding the appointment of employees. This chapter concluded with a reminder that psychometric tests carry more weight when several instruments are used in combination rather than a single test being used in isolation.
CHAPTER 3
Methodology

3.1 Introduction

This chapter will define and expand upon the rationale behind the choice of methodological approach used in the current study. The research procedure, population and sample are described, and the relevant steps in gathering and processing the data are discussed. An important aspect of explaining the methodology is to provide a description of reliability and validity studies in general, before focusing on how these techniques were applied in the current study.

3.2 Quantitative and qualitative research designs

There are two broad categories of research design which may be used to conduct research in the social sciences, namely a quantitative or a qualitative design (Kerlinger, 1986; Whitley, 2002). Quantitative research uses numeric data (quantities) that can be measured, counted or ranked. Numbers or scores tend to be the main type of data that researchers collect and analyse using this framework (Graham & Skinner, 1991). Quantitative researchers analyse their data in numerical form so that means, correlations and frequencies can be easily distinguished. Numerical data can illustrate a typical behaviour of people in a given population or situation (Whitley, 2002).

In contrast, qualitative research generally does not involve numeric data; instead, it tries to explain social phenomena by means of narratives about people’s life experiences (Graham & Skinner, 1991). Qualitative data is open-ended and describes the nature of behaviour through themes and narrative responses (Whitley, 2002). The main aim of qualitative research is to seek an
understanding of a given research problem from the perspectives of the population it involves. That is, a qualitative approach is effective in obtaining culturally specific information about values, opinions, behaviours, and social contexts of particular population groups (Mack, Woodsong, MacQueen, Guest & Namey, 2005).

Whitley (2002, p. 34) states that "All researchers, whether they deal with numbers or words in the laboratory or field must deal with issues of generalizability, validity, replicability, ethics, audience and their own subjectivity or bias. Both research designs involve people, and it is every researcher’s responsibility to pick the best methodology to yield optimum results for an intervention or study. The qualitative approach is perhaps best viewed more as a philosophy than a precise research method, while quantitative research tends to be guided by positivism (Whitley, 2002).

Table 7 below shows the main differences between quantitative and qualitative research.

<table>
<thead>
<tr>
<th>Paradigm dimension</th>
<th>Quantitative research</th>
<th>Qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underpinning philosophy</td>
<td>Rationalism: humans achieve knowledge through their capacity to learn (Bernard, 1994, p. 2)</td>
<td>Empiricism: The only knowledge that humans acquire is from experience (Bernard, 1994, p. 2)</td>
</tr>
<tr>
<td>Approach to Enquiry</td>
<td>Structured, rigid predetermined methodology</td>
<td>Unstructured, flexible and open methodology</td>
</tr>
<tr>
<td>Main purpose of investigation</td>
<td>To quantify extent of variation in a phenomenon or situation</td>
<td>To describe variation in a phenomenon or situation</td>
</tr>
<tr>
<td>Measurement of variables</td>
<td>Emphasis on some form of measurement or classification of variables</td>
<td>Emphasis on the description of variables</td>
</tr>
<tr>
<td>Sample size</td>
<td>Big samples necessary</td>
<td>Smaller sample size</td>
</tr>
</tbody>
</table>
Table 7: Qualitative versus quantitative research
(From Kumar, 2005, p. 17)

Quantitative research may focus on identifying cause-and-effect relationships among different variables, or on establishing correlations (i.e. not causality but rather co-variance). It may also summarise a lot of information into descriptive statistics such as the arithmetic mean (average), median (midpoint score in a set of scores) or mode (most frequently occurring score). It tries to maximise its validity by focusing on the ‘average’ behaviour of most people in a given population (Kerlinger, 1986; Rosenthal & Rosnow, 1991).

In contrast, qualitative researchers are more interested in how people experience and interpret the events in their lives. These inner or subjective experiences may differ from person to person, and qualitative researchers believe that the data yielded by this type of enquiry is rich and informative, and is often easier to relate to than numbers and scores (Mack et al., 2005).
Qualitative researchers can also modify their approach if the research situation changes; the structure of the research process depends largely on the main themes that emerge during the interviews. Researchers using a qualitative design often study the relationship of people’s behaviour to their social context, and attempt to discover how these aspects influence one another. Qualitative researchers also include an analysis of their own experiences, thoughts and values in the research process. The process thus becomes a relationship where data is both given and received (Neuman, 2000; Whitley, 2002).

Quantitative researchers essentially attempt to separate the natural context from the behaviour under study, and seldom use self-reflective analysis. Many, but not all, quantitative studies are done in laboratories where participants’ behaviour can be manipulated (Kerlinger, 1986; Neuman, 2000). By using a quantitative design, the researcher detaches him or herself from the research process, and promotes objectivity in a one-way relationship where the researcher gets information without sharing his or her own experiences with others (Whitley, 2002).

3.3 Research paradigm of the current study

A quantitative approach was employed in the current study because the raw data to be analysed was already available in the company’s database. This data comprised two sets: 1) the CTB2 numeracy subtest scores and 2) the work performance scores of call centre employees at the company. Given that the data had already been numerically measured and recorded, the choice of a quantitative methodology can be justified by citing Muijs’s (2004, p. 3) statement that “quantitative research explains phenomena by collecting numerical data that are analysed using mathematically based methods.” This method ensures high reliability levels of gathered data, and minimises the involvement of researchers while maximising objectivity during the research process (Matveev, 2002).
The epistemological assumptions underlying this study are informed by the positivist paradigm. The positivist paradigm explores social reality, and is based on the ideas of the French philosopher Comte (b. 1732, d. 1818), who emphasized observation and reason as the means for understanding human behaviour. The positivist theory is based on the premise that true knowledge is founded on the experience of the senses, and can be obtained through observation and experiment (Somekh & Lewin, 2004). Positivism assumes that only observable behaviour and competency can be studied and interpreted (Meyer, Moore & Viljoen, 1989). In the current study, the observables were job performance and assessment scores.

Positivist research remains formal and separate from the subjects who participate in the studies. Positivists believe that research produces truthful information about an objective world, and they commonly employ structured methods such as experiments or surveys to produce quantitative data. However, positivist research might also use structured interviews or observations to record qualitative data in a systematic fashion (Somekh & Lewin, 2004). This point is important in the current study, because such methods had contributed to the calculation of participants' work performance scores (see section 1.3.2).

In psychological testing, individual competencies are not viewed in isolation. Competencies are rather viewed as combinations of complementary skills and knowledge bases. Within the employee group, combined competencies give rise to critical business processes being performed in a satisfactory way. An employee follows the organisational culture and is able to deliver competent results at the end of each task (Mascarenhas et al., 1998). When interpreting assessment results, it is always important to measure people's ability to function optimally. The skills, knowledge and psychological attributes (such as behaviour, attitudes and values) of each candidate must complement the scope of the company in order for the candidate to provide competent service (Mascarenhas et al., 1998). Because of this subjective dimension, the work performance ratings of call centre agents at the company under study were determined partly by their managers' perceptions, rather than purely through
objective instruments such as computer-based questionnaires about product knowledge (see section 1.3.2).

Using aspects of quantitative and qualitative research designs together may yield rich, useful data which covers both the subjective experiences of people whilst supporting these findings with quantitative data in the form of numbers. However, in the current study the researcher decided to use only quantitative methodology because the available data was already in a quantitative format and was accessible for analysis. Therefore, the subjective or qualitative aspects of participants’ scores on the work performance variable were not directly examined in this study.

3.4 Study design

The research design of the current study was exploratory and descriptive (Kerlinger, 1986; Rosenthal & Rosnow, 1991; Rose & Sullivan, 1996). The aim of the study was to describe, both in statistical terms and through logical explanation, the phenomenon under study. The statistical relationship between the numeracy score variable and the performance score variable of call centre agents would be described and analysed. No attempts were made to manipulate variables or the scores of call centre agents, but the data was taken in its raw format from a pre-existing company database.

Although a quantitative research design and statistical analyses were used, it should be noted that some degree of subjective or qualitative assessment had contributed to the work performance scores.
In terms of time-frames, the research design was somewhat longitudinal as it made use of two sets of scores, obtained from the same individuals at different points in time. However, cross-sectional measurement methods were used, in that each participant’s scores were measured at only one point in time using one specific instrument or process (Kerlinger, 1986; Rosenthal & Rosnow, 1991).

3.5 Population and sample

This study was conducted within the call centre division of a large retail company employing about 1 800 call centre agents. The agents were involved in the selling of products and the delivery of services for consumers’ personal or family needs. The company is well known for its rewarding interactions with the end users of its products and services (the company’s database, 2006). The study focused on call centre staff because call centre management had raised a concern about the validity of the assessment tests when used with this group, specifically the numeracy subtest of the CTB2.

A sample of 300 individuals, drawn from the database of about 1 800 individuals, was sufficiently large to represent the entire group. The entire database was not used because of its unwieldy size and the amount of time that would have been required to check and clean the data before analysis. The drawing of a smaller sample is further justified by the fact that this was an exploratory study only, which could not meet the most rigorous of statistical requirements in other regards (e.g. controlling for the extraneous variables mentioned in section 3.9.1 below).

The research design made use of a matched sample, in that each research participant was paired with himself or herself on the two sets of scores being studied. This helps to control for potentially confounding factors such as gender, age and socio-economic variables (Kerlinger, 1986; Rosenthal & Rosnow, 1991).
3.5.1 Sample size

A research sample must have an adequate number of participants in order to be considered representative of the population under study (Kerlinger, 1986; Rosenthal & Rosnow, 1991; Whitley, 2002; Murphy & Myors, 2004). The sample should also be manageable by the primary researcher in terms of data cleaning and preparation, and time-frames for completing the research. According to Pelser (2002) the sample should be large enough to attain a statistical power of 0.80 and have an alpha no larger than 0.05. These guidelines minimise the researcher's odds of making either a Type 1 error (finding a relationship between variables when in reality there is none) or a Type 2 error (finding no relationship in variables when in reality there is one) (Pelser, 2002; Rosenthal & Rosnow, 1991).

A representative sample of 300 employees was drawn from the total nationwide call centre staff of 1 800, based on two techniques for determining sample size. The first was the sample-size calculator provided by SPSS\(^4\), the software program that was used to analyse the data. This calculation suggested that a sample of 300 would be suitable. The confidence level was set at 95%, which meant that the error margin and \(\alpha\) value was 0.05 (McCall, 1990; Kerlinger, 1986; Rosenthal & Rosnow, 1991).

The second method of selecting the sample size was to consult a power table based on Cohen's widely accepted guidelines (Rosenthal & Rosnow, 1991, p. 444). The table confirmed that a sample of 300 would be more than adequate to achieve a statistical power of 0.80 for a medium effect (e.g. \(r = 0.30\)) at the level of \(\alpha = 0.05\) (two-tailed test), or even at the more stringent \(\alpha = 0.01\) level. For \(\alpha = 0.01\) and power of 0.80, the sample needed to be only \(N = 125\). More will be said about statistical power in section 3.6 below.

\(^4\) Statistical Package for the Social Sciences: IBM® SPSS®. Website: [www.spss.com](http://www.spss.com)
3.5.2 Sampling method

There are many ways in which a sample can be drawn from a study population. One main distinction is probability versus non-probability sampling. In probability sampling, every person selected for the study has an equal and known non-zero probability of being picked for the study (Rosenthal & Rosnow, 1991; Somekh & Lewin, 2004), and there is some form of random selection involved (Kerlinger, 1986). In contrast, non-probability or convenience samples use no randomisation and therefore such studies are not as scientifically rigorous (Kerlinger, 1986). However, in reality non-probability sampling is used more often, partly because such samples are readily available (e.g. a high school class, a group of psychology undergraduates, or a pool of employees) and partly because they are usually much cheaper to organise. Probability samples, with all of the various types of randomisation that are possible, are generally far more difficult and expensive to obtain, although the results of such studies are of a higher standard statistically (Kerlinger, 1986).

In the current study, there was some randomisation in that only 300 participants were selected from the available pool of 1,800 employees. However, the sample should be considered a purposive and convenience sample, in that it was readily available and pre-existing. Purposive sampling is characterized by the use of judgement and a deliberate effort to obtain representative samples by including presumably typical areas or groups in the sample (Kerlinger, 1986, p. 120).

Randomisation of the sample was limited by the fact that all members of the population under study formed the pool from which data was taken. The subjects to be included in the current research were selected from this large pool of call centre employees using a systematic rather than random sampling technique (Kerlinger, 1986; Whitley, 2002). Kerlinger (1986, p. 120) describes
this method as follows: ÒHere the first sample element is randomly chosen from numbers 1 to \( k \) and subsequent elements are chosen at every \( k^{th} \) interval. For example, if the element randomly selected from the elements 1 through 10 is 6, then the subsequent elements are 12, 26, 36, and so on.Ó

In the current study the actual sampling was carried out by SPSS, with the sampling interval being 6 (\( 1800 \div 300 = 6 \)). The resulting sample was a group of call centre agents with mixed racial, gender and age attributes.

3.6 Statistical power

The power of a statistical test is defined as the probability of rejecting the null hypothesis when it is indeed false. In other words, power is the probability of not making a Type 2 error (Rosenthal & Rosnow, 1991; Kerlinger, 1986; Indiana University). As power increases, the chances of making a Type 2 error will decrease. Power indicates how often one would reach a correct interpretation if we were to repeat the study many times over. Studies with high levels of statistical power are likely to detect the effects of treatments and interventions correctly, whereas studies with low statistical power are known often to lead researchers to dismiss potentially important effects as sampling error (Kaplan & Saccuzzo, 2001; Rosenthal & Rosnow, 1991).

The power of a statistical test depends on factors such as the alpha (\( \alpha \)) level chosen for the study, and the size (\( N \)) of the research sample. The sample size is used to determine statistical significance, and the power of a test is a function of the size of the treatment effect (Rosenthal & Rosnow, 1991; Norman & Streiner, 2008; Pelser, 2002). The \( \alpha \) for this study is 0.05, and a statistical power of 0.80 was used to determine the number of subjects necessary (Rosenthal & Rosnow, 1991, p. 444; SPSS).
3.7 Biographical information

The researcher was given permission by the assessments executive manager to conduct the current study. In 2006 the study was commissioned at several business units within the company, specifically because call centre managers had questioned the relevance of the Clerical Test Battery (CTB2) numeracy subtest for screening call centre candidates. The use of the CTB2 was compulsory for screening all such candidates, and the assessment department assumed that call centre agents who did well on the CTB2 test would also perform well at work. The study was commissioned to verify whether there is in fact this positive relationship between CTB2 numeracy results and subsequent performance scores of call centre staff.

The following biographical information for the sample was sourced from the company's database:

- Gender
- Race
- Position
- Call centre name

This biographical information enabled the researcher to profile the sample. The sample was drawn from the company's three call centres in Johannesburg (46%), Durban (31%) and Cape Town (23%). The majority of participants were black (55.7%), the remainder being classified as Coloured (31.3%), Indian (7.3%) or white (5.6%). The sample was made up mostly of women (79%) with a smaller number of men (21%).

3.8 Research hypotheses for the current study

Section 1.8 (p. 14) summarised the nature of a research hypothesis and gave the hypothesis for the current study. The research hypothesis was formulated in non-directional terms for three reasons:
1. The presence or absence of a relationship between the two variables was to be investigated, rather than the directionality of any such possible relationship.

2. A two-tailed statistical test is more stringent and conservative, in terms of establishing significance, than a one-tailed test (Kerlinger, 1986; Rosenthal & Rosnow, 1991; Norman & Streiner, 2008). It was decided to use the more stringent option to reduce the chance of a Type 1 error. A two-tailed test implies a non-directional hypothesis.

3. Convention supports the use of two-tailed tests unless prior research has clearly established the directionality of a relationship between variables (Norman & Streiner, 2008; McCall, 1990; Indiana University). Furthermore, the null and alternate hypotheses should together cover all possible permutations of the data which may emerge (McCall, 1990; Kerlinger, 1986; Rosenthal & Rosnow, 1991). Using a directional alternate hypothesis would not cover the possibility that an inverse (negative) relationship may be found between the variables under study. In other words, higher CTB2 numeracy scores may have been associated with lower work performance scores, while lower numeracy scores may have been associated with better work performance. However unlikely and unexpected this scenario may be, it has not been ruled out by previous studies done with the CTB2 in the population in question. Therefore the prudent choice is to use a non-directional $H_1$ in the current study.

The research hypothesis was formulated to investigate whether a correlation exists between CTB2 numeracy subtest scores and work performance ratings. The following hypotheses were set:

$H_0$ There is no significant correlation between individuals’ scores on the numeracy subtest of the CTB2 and their later work performance ratings ($r = 0$)
**H₁** There is a significant correlation between individuals' scores on the numeracy subtest of the CTB2 and their later work performance ratings ($r \neq 0$)

Where $r$ is the correlation co-efficient.

### 3.9 Independent and dependent variables

Kaplan and Saccuzzo (2001, p. 104) define an independent variable as “the antecedent phenomenon in the cause and effect relationships” while the dependent variable is referred to as the “consequent phenomenon.” Babbie (2007) states that the independent variable can be seen as a hypothesized cause or influence on the dependent variable.

For the purposes of this study, the independent variable was participants’ scores on the numeracy subtest of the CTB2. The dependent variable was the work performance scores of the same individuals once they had become call centre agents. In this particular case, it was not possible to determine whether the independent variable would directly “cause” the dependent variable. The only type of analysis that could be done was correlational rather than causal, that is, to see if there was any co-variance in the two scores. Analysis therefore aimed at identifying whether there was a positive correlation between numeracy subtest scores and work performance scores.

#### 3.9.1 Measuring numeracy scores: independent variable

As discussed in Chapter 2, the Clerical Test Battery (CTB2) consists of four tests bound into one booklet. It assesses a range of clerical aptitudes and skills: verbal reasoning, numerical ability, clerical checking and spelling. Designed for screening people for general clerical and administrative positions, the subtests can either be administered in isolation or the whole battery can be administered to produce a candidate profile. The numerical ability test evaluates the individual’s ability to use and manipulate numbers in
In a practical way. The questions in the test have been designed to reflect the kinds of numerical problems an administrator may experience on a day-to-day basis (Psytech, 2002 & 2010).

In the current study, the CTB2 numeracy subtest scores were used as the independent variable.

Numeracy scores were obtained at different points in time and under slightly different circumstances each time. Possible nuisance variables which may have resulted from this situation were not studied. That is, the possible effects of things like season of the year, time of day, day of the week, specific room used for the testing, number of candidates present during a testing session and characteristics of the person who administered the test were not investigated.

3.9.2 Measuring work performance scores: dependent variable

The company under study made use of two performance measures to assess its call centre employees. The productivity measure examined the number of phone calls an agent made or received in a day, while the adherence measure evaluated whether an agent was able to make calls and render services while adhering to the allocated time schedule (see section 1.3.2). Line managers in the call centre division also contributed to the work scores via their subjective ratings of an employee's performance.

3.10 Data collection

Data collection is the process of gathering data with the intention of obtaining information, so as to make decisions about important research questions. Data may be gleaned from primary sources such as databases for unique reporting requirements (Lee, 2000; Kerlinger, 1986), as in the current study. The researcher sourced data from the company’s database and individuals were selected systematically from the database, as described in 3.5.2 above.
This available raw data provided the basis for the current research via secondary analysis. No attempt was made to gather new raw data for the purposes of this study. The time frame for data collection was six months followed by data analysis.

3.11 Data cleaning

In the data cleaning stage, the researcher sorts through the data to eliminate errors and inconsistencies in datasets before they are analysed (Maletic & Marcus, 2000). When data is cleaned problems such as misspelling which may have taken place during data entry, and cells missing information, are detected or corrected. Missing values may require a participant to be deleted, or it may be possible to substitute the group mean into the empty cell. However, no subjective "guesses" may be entered which may compromise the integrity of the data (Rose & Sullivan, 1996).

Data cleaning is thus made up of three stages:

- Define and determine the types of error in the datasets (such as spelling mistakes or missing values)
- Search for and identify each instance of data error (this includes looking for inconsistent numbering in the data spreadsheet)
- Correct the errors before running the analysis (Maletic & Marcus, 2000).

3.12 Conducting validity and reliability studies

Reliability and validity were defined in Chapter 1 of this dissertation. Reliability studies assess the consistency of an instrument, either over time or with regard to the correlations (r) between the scores yielded by the instrument’s subscales. The reliability index (Cronbach’s α) of a test instrument is the extent to which the instrument yields the same results over repeated trials
Reliability thus refers to the accuracy of the instrument, which is evaluated according to whether the results yielded by it remain consistent on different testing occasions (Kaplan & Saccuzzo, 2001).

In contrast, validity studies assess the strength of a given conclusion and the likelihood that the instrument measures what it was developed to measure (Rose & Sullivan, 1996; Rosenthal & Rosnow, 1991). In a workplace selection context, test performance should predict future job performance; to achieve this it is essential to use the right assessment instrument (Whitley, 2002). Workplace validation studies are thus an attempt to identify predictor variables, and to determine the extent to which it is correct to deduce what an employee’s work-related behaviour or performance is likely to be, based on his or her assessment results. Therefore, it is essential to understand what a specific job will require a candidate to do (Kaplan & Saccuzzo, 1997).

Reliability must be established before validity can be investigated, because reliability is a prerequisite for claiming that an instrument is valid (Rosenthal & Rosnow, 1991; Kerlinger, 1986). The two dimensions are often assessed together in the same study.

Many test users and developers validate their measuring instruments in order to investigate the limits of their functionality and utility. Validity studies examine the usefulness and accuracy of an instrument in a particular environment, by using context-specific criterion measures. Validity cannot be assumed and must be statistically established with reference to a specific intended use of the instrument in that particular context (Anastasi & Urbina, 1997).

The validity of an instrument depends on three main criteria:

- identifying the intended purpose for which the test is designed
- questioning whether the test is being used for its intended purpose
- assessing how well the test performs that intended function (Anastasi, 1988).
The following are typical steps employed in validity and reliability studies done in the workplace:

**STEP 1** Conduct a thorough job analysis and clearly identify key dimensions (knowledge, skills and attitudes) and characteristics to be measured (Adler, 2007). The key dimensions and characteristics should be measured in terms of outcome performance, and one must ensure accurate measurement of outcomes or performance (also known as criterion measures).

**STEP 2** Decide on appropriate measurement tools and other available information to include, either those already in use or new tests that may be under consideration. All available biographical information, school results, job categories and so on should be included in data analysis. Evaluate tools and methods to check that the key dimensions are covered.

**STEP 3** Determine appropriate criterion measures to be used to assess outcome performance. Criterion measures might include, for example, all training results, all outcome results, and/or ratings of performance. This step is based on the research problem identified, in this case the validity of the CTB2 numeracy subtest when used with call centre candidates.

**STEP 4** Data gathering, capturing, and analysis. During this process data is collected and captured on a central system for cleaning and analysis; in the current study the collection had already been done by people other than the researcher. Next, the data is analysed using a statistical software package.

**STEP 5** Interpretation of results. The statistical results must be interpreted in plain language and written in a report format. This report should be given to all involved stakeholders.
STEP 6  Ongoing monitoring of the utility of tests and measures in the particular context of relevant groups. This may include:

- Gathering information to verify the instrument’s psychometric properties (especially predictive validity) and practical utility (qualitative and quantitative data).
- Critical evaluation of the instrument by using criterion or outcome performance measures (Anastasi & Urbina, 1997).

In the early 1990s, test developers showed that it was essential to assess a particular ability by using more than one instrument for the same group of respondents. This is the best way to ensure that the results give a true reflection of the respondent’s ability and knowledge (White, 2002). Though this process may seem like an unnecessary expense, it also assists in assessing each instrument’s ability to measure the attributes it is intended to measure (Psytech, 2002).

3.13 Previous reliability and validity studies of the CTB2 in South Africa

Sections 2.13 and 2.14 of the literature review reported the results of reliability and validity studies which have been done using the CTB2 in a South African context (Psytech, 2002; 2010). The information will not be repeated here and the reader is asked to refer back to that section if necessary.

3.14 Data analysis

Whitley (2002) describes data analysis as a process of gathering the data and then transforming it into a format that highlights useful information. This information may be used to suggest an intervention or support a decision. In the current research, and indeed in most or all quantitative research done these days, a computer program is used to analyse the data (Rosenthal &
Rosnow, 1991). The final interpretation of the results must be done by a human being, however. In the current study the Statistical Package for Social Sciences (SPSS®) software (Version 17) was used to analyse the results.

Descriptive statistics are coefficients or numerical values that summarise a given dataset (Graziano & Raulin, 2000). These mathematical summaries condense a lot of information into one value, such as the mean (average), median (midpoint score in a set) or mode (most frequently occurring score) (Kerlinger, 1986; Rosenthal & Rosnow, 1991).

Descriptive statistics are used to illustrate the basic features of the research data. They often provide summaries in the form of graphs or tables. This type of statistical information describes the data numerically or quantitative terms, which makes it easy to compare with other datasets and provides a basis for decision-making (Graziano & Raulin, 2000; Cohen, Cohen, West & Aiken, 2003).

Descriptive univariate (one-variable) statistics provide an overall picture of the research sample and measures. Such statistics are convenient in that they present data in a user-friendly way (Kaplan & Saccuzzo, 2001). The means and standard deviations of the predictor variable (in the current study, numeracy subtest scores) and criterion variable (work performance scores) are calculated to reveal the nature of the distribution of scores (Pelser, 2002).

The datasets in the current study were the two sets of scores obtained by the sample representative of the call centre staff. Descriptive statistics were used as they made it easier to present data in a simple tabular or graphic form (Graziano & Raulin, 2000; Cohen et al., 2003).

3.15 Correlations

Correlation coefficients are defined as an index of the degree of relationship between two variables (Whitley, 2002. p. 231). Correlational tests aims to reveal the degree to which changes in one variable are associated with
changes in another (Pelser, 2002; Psytech, 2002; Kerlinger, 1986). In the current study, CTB2 numeracy subtest scores were correlated with job performance appraisal scores.

Correlational data is presented in descriptive form by the use of a single index (Pearson’s $r$) which indicates the degree of correlation (Kerlinger, 1986; Rosenthal & Rosnow, 1991). A perfect (100%) correlation between two variables would be 1.00 (in either a positive direction or a negative / inverse direction), while variables that correlate at about 0.50 or higher indicate a large “effect size” or relationship. A correlation of 0.30 (30%) is moderate or medium, and a correlation lower than about 0.10 (10%) or approaching zero (0% or 0.00) indicates that the variables are minimally related to each other, or do not co-vary at all (Rosenthal & Rosnow, 1991). The values mentioned here apply to an assumed alpha level of 0.05 (Rosenthal & Rosnow, 1991).

3.16 Conclusion

This chapter clarified the purpose of the current study, and explained and explored the concepts of quantitative versus qualitative research designs. Further aspects of the research design were discussed, such as the presence of elements of both cross-sectional and longitudinal designs. The manner in which the sample was selected was described, together with the sample’s demographic characteristics (e.g. gender and race). The methods used for collecting and cleaning the data were discussed, followed by an outline of the methods of data analysis to be used. The steps involved in doing reliability and validity studies in a workplace context were also discussed, with particular reference to the current study.

The actual findings which were yielded by the data analysis are reported in the next chapter (Chapter 4), with the discussion and interpretation thereof being presented in the following chapter (Chapter 5).
CHAPTER 4
Results

4.1 Introduction

In this chapter, the results of the study are presented together with a discussion about the process of data analysis. The main focus was on seeing whether there were any statistically significant correlations between CTB2 numeracy scores and work performance scores, but analyses were also done according to gender and race group.

Descriptive statistics are used to present the research findings, due to the compatibility of this format with quantitative analysis. It also offers a way of presenting data in a manageable and convenient way to give a holistic view of the findings (Kaplan & Saccuzzo, 2001). Demographic information about the participants is presented first, followed by the results of the data analysis.

Although some of the statistical terms and symbols used in this chapter have already been described, some key definitions and explanations are repeated in Table 8 below, which is included here for ease of reference.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>α (Alpha)</td>
<td>The probability of making a Type 1 error (rejecting the null hypothesis when it is true). Alpha is usually set at 0.05 which means the chance of making an error is 5%</td>
</tr>
<tr>
<td>Cronbach's alpha</td>
<td>An index of internal consistency reliability</td>
</tr>
<tr>
<td>β (Beta)</td>
<td>The probability of making a Type 2 error (failing to reject the null hypothesis when it is false). The power of a test is given by the formula 1 - β Power is the odds of not making a Type 2 error</td>
</tr>
</tbody>
</table>
p-level
- After setting the alpha level, a statistic (like r) is calculated together with its p-value. The p-value shows the likelihood that the statistic was obtained by chance. If p \( \leq \alpha \) (p \( \leq 0.05 \)) the null hypothesis is rejected and the result is statistically significant. If p \( > \alpha \) (p \( > 0.05 \)) the null hypothesis is not rejected and the result is not statistically significant.

r (Pearson's product-moment correlation)
- A measure of the linear correlation between two variables

CL (confidence level)
- Confidence level is 1 - \( \alpha \). Thus for alpha of 0.05 we can state with 95% confidence (1 - \( \alpha \)) that a parameter (the score or measure obtained from the sample) falls in the same region as the true population statistic would.

<table>
<thead>
<tr>
<th>Table 8: definitions of statistical terms and symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Adapted from Rosnow &amp; Rosnow, 2009, p. 65; Rosenthal &amp; Rosnow, 1991)</td>
</tr>
</tbody>
</table>

### 4.2 Biographic information about the sample

#### 4.2.1 Gender breakdown

The total sample comprised 300 participants (N= 300). Their biographic data is shown in the following series of tables, with the first (Table 9a) indicating the gender composition of the sample.

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>64</td>
<td>21</td>
</tr>
<tr>
<td>Female</td>
<td>236</td>
<td>79</td>
</tr>
</tbody>
</table>

| Table 9a: Gender breakdown |

In South Africa, an estimated 51% of the general population is female (Statistics South Africa, 2010) but this pattern was not observed in the sample. It is not clear from the results of this research why there was such a preponderance of women working in the call centre (79%), but some possible reasons will be discussed in the next chapter.
4.2.2 Racial breakdown

<table>
<thead>
<tr>
<th>Race group</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>167</td>
<td>55.7</td>
</tr>
<tr>
<td>White</td>
<td>17</td>
<td>5.7</td>
</tr>
<tr>
<td>Coloured</td>
<td>94</td>
<td>31.3</td>
</tr>
<tr>
<td>Indian</td>
<td>22</td>
<td>7.3</td>
</tr>
</tbody>
</table>

Table 9b: Racial breakdown

Table 9b above shows the racial composition of the sample. As will be discussed in the following chapter, in South Africa the majority of the general population is black. The sample in the current study was in line with this, as most of the sample was black (55.7%), followed by Coloured (31.3%). Whites and Indians made up a small minority (5.7% and 7.3% respectively). The sample’s racial composition will be discussed further in the next chapter.

4.2.3 Participants’ positions

The participants held different positions in the company, although all of them were at the Band A or B level. Table 9c on the next page shows the types of jobs and the number of participants who worked in them.
<table>
<thead>
<tr>
<th>Position</th>
<th>N</th>
<th>% of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional services manager</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>New accounts manager</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Internet call centre manager</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Customer services manager</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Customer services</td>
<td>53</td>
<td>18</td>
</tr>
<tr>
<td>New accounts</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>Financial services</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Permanent part-timers</td>
<td>112</td>
<td>37</td>
</tr>
<tr>
<td>Secretary</td>
<td>1</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 9c: Call Centre Positions

Many of the call centre staff (37%) were "permanent part timers" which was defined as "employees who normally work less than 35 hours a week (the company's database). This type of employee did not receive full benefits from the company. The next highest number of participants were new accounts staff (33%) and customer service staff (18%). Together, these three groups made up the bulk of the sample at 88% (N=265). The remaining 13% of the sample were people working in other roles, some of whom may have initially been employed as call centre agents.
4.2.4 Participants’ regions (geographical area)

<table>
<thead>
<tr>
<th>City / Province</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cape Town</td>
<td>68</td>
<td>23</td>
</tr>
<tr>
<td>Durban</td>
<td>93</td>
<td>31</td>
</tr>
<tr>
<td>Johannesburg</td>
<td>139</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 9d: Geographical area

The majority of participants were based in the Johannesburg call centre (46%), followed by Durban (31%) and Cape Town (23%) respectively. One of the reasons attributed to a large number of employees in Johannesburg is the fact that Gauteng province has the highest population density in South Africa, at 22.4% of the national population. The Gauteng population is especially dense in the Johannesburg area due to the presence of job opportunities in that city (Statistics South Africa, 2010). The company’s head office is based in Johannesburg.

4.3 Descriptive statistics and t-tests

Simple descriptive statistics were calculated for the CTB2 numeracy scores to establish the mean (average) score and other measures of central tendency, as well as the standard deviation of the sample. The mode and median were 90 in every case, both for the total sample and for each of the gender subgroups. The mean for the total sample was 93.4 (rounded), while the average male score was 92.7 and the average female score was 93.6.

<table>
<thead>
<tr>
<th>Sample group</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Std Dev</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>93.37</td>
<td>90</td>
<td>90</td>
<td>6.62</td>
<td>80-120</td>
</tr>
<tr>
<td>Males</td>
<td>92.66</td>
<td>90</td>
<td>90</td>
<td>6.23</td>
<td>80-110</td>
</tr>
<tr>
<td>Females</td>
<td>93.56</td>
<td>90</td>
<td>90</td>
<td>6.72</td>
<td>80-120</td>
</tr>
</tbody>
</table>

Table 10: Measures of central tendency
The South African CTB2 numeracy subtest means given by Psytech-SA (2010) range from about 4 to 12. The company’s raw data had evidently been recorded in a different format, with the numeracy scores ranging from 80 to 120. Because of this discrepancy no direct comparison can be made between the sample’s results and the established South African norms.

The results for the two gender groups were tested by an F-test to see if their variances were equal or unequal (Rosenthal & Rosnow, 1991; Kerlinger, 1986). The F test showed no significant difference in variance for these two groups ($F_{235,63} = 0.49$, $p = 0.99$ using a two-tailed test). This indicated that the variance could be treated as equal.

A standard (Student’s) t-test was then done. The t-test investigates whether or not the mean scores of two groups differ to an extent that is statistically significant (Rosenthal & Rosnow, 1991; Kerlinger, 1986; DeCoster, 2006). The t-test was used here to find out if the difference in mean scores obtained by the male and female subgroups on the CTB2 numeracy test was statistically significant. The t-test for groups with equal variance was used.

The t-test indicated that there was no statistically significant difference between male and female scores on the CTB2 numeracy subtest ($t_{298} = 0.33$, $p = 0.74$ using a two-tailed test). However, the t-test should ideally be used with two samples of equal or approximately equal size (Rosenthal & Rosnow, 1991; Kerlinger, 1986). This was not the case with the current gender subgroups, therefore this result should be regarded with a little caution.

Regarding the differences between scores obtained by the various racial subgroups, analysis of variance (ANOVA) would have been the most suitable analysis to use. This is a form of F-test used to compare multiple groups rather than just two groups (Kerlinger, 1986; Rosenthal & Rosnow, 1991). However, in the present study ANOVA was not carried out due to the extremely unequal sample sizes of the racial subgroups (Rosenthal & Rosnow, 1991; DeCoster, 2006). For example, the white subgroup comprised only 17 people while there were 167 people in the black subgroup.
4.4 Correlations

Correlation measures were used to establish whether or not a relationship existed between the CTB2 numeracy scores and work performance scores. Correlation tests are used to measure the strength of the linear relationship between two variables. As discussed in section 3.15 in the previous chapter, correlation coefficients (Pearson’s $r$) range between -1 and 1. A positive valence means that as one variable increases the other variable similarly increases, while a negative valence means that as one variable decreases so the other variable decreases (Cohen et al., 2003; Kerlinger, 1986).

The larger the numerical value of Pearson’s $r$, the stronger the effect size or relationship strength. As discussed in section 3.15, a perfect correlation would be 1.00 (100%), while a zero correlation (0.00) would mean there is no statistical relationship at all between the two variables. Variables that correlate at 0.50 or higher usually indicate a large effect size or relationship. A correlation of 0.30 (30%) shows a relationship of medium strength, while a correlation lower than 0.10 (10%) means that the variables are not related or do not co-vary (Rosenthal & Rosnow, 1991).

A number of correlation analyses were performed on the data, with separate analyses being done for each gender and race group. The outcomes of these analyses are presented in this section. Pearson’s $r$ was calculated in each case, according to the objective of the study and to test the research hypotheses. It is recalled that the null hypothesis stated that there is no statistical relationship between the numeracy subtest scores of the CTB2 and the later performance ratings; the alternative hypothesis stated that the numeracy subtest scores of the CTB2 do correlate significantly with performance ratings.

The alpha level for the correlation analysis was set at 0.05. Anastasi (1988) states that alpha values of 0.01 to 0.05 (or smaller than 0.01) are considered
statistically significant. An $\alpha$ of 0.01 means there is only 1% chance of
drawing the wrong conclusion based on the test results. An $\alpha$ of 0.01 thus
means the researcher is being extremely careful and is willing to risk being
wrong only 1 in a 100 times, with regard to rejecting the null when it is true
(i.e. making a Type 1 error). An $\alpha$ of 0.05 means the researcher risks being
wrong 5 in 100 times, which is slightly less stringent but is the standard used
in most research. Any risk larger than 0.05 is considered unacceptable for
statistical significance (Rosenthal & Rosnow, 1991; Kerlinger, 1986; Anastasi,
1988).

It is important to analyse data bearing in mind these probability values, as
they guide the researcher both in interpreting the data and in being able to
establish if a relationship really exists between the variables. The values are
also seen as guidelines that are used in statistical analysis to make informed
conclusions about a dataset (Rosenthal & Rosnow, 1991; Kerlinger, 1986;
Anastasi, 1988).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Correlation</th>
<th>p-value</th>
<th>N (sample size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>0.014</td>
<td>0.805</td>
<td>300</td>
</tr>
<tr>
<td>Males</td>
<td>-0.004</td>
<td>0.973</td>
<td>64</td>
</tr>
<tr>
<td>Females</td>
<td>0.012</td>
<td>0.856</td>
<td>236</td>
</tr>
</tbody>
</table>

Table 11a: Correlations (Pearson’s r, two-tailed analysis) between CTB2
numeracy scores and work performance scores

Table 11a above shows the results of the correlation analysis for the entire
sample (N=300) as well as for each separate gender group (males N=64;
females N=236).

None of the correlation indices reported in Table 10a showed any evidence of
a relationship between CTB2 numeracy subtest scores and performance
scores. The results would have been significant only if the p-values were
equal to or less than 0.05, whereas the obtained values were more than 10
times higher than this limit (at 0.805, 0.973 and 0.856 respectively).
In addition, the correlation coefficients \( (r) \) indicated non-existent relationships. A small effect size would require a correlation of at least 0.1 or more \( (r \geq 0.1) \), with medium and large effect sizes becoming evident only when the correlation index reaches 0.30 and 0.50 respectively, or higher. The obtained values were considerably less than this, ranging from 0.004 to 0.014, thus well under 0.1 in each case.

<table>
<thead>
<tr>
<th>Race group</th>
<th>Correlation</th>
<th>p-value</th>
<th>N (sample size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.056</td>
<td>0.475</td>
<td>167</td>
</tr>
<tr>
<td>Coloured</td>
<td>0.055</td>
<td>0.598</td>
<td>94</td>
</tr>
<tr>
<td>Indian</td>
<td>-0.095</td>
<td>0.673</td>
<td>22</td>
</tr>
<tr>
<td>White</td>
<td>0.171</td>
<td>0.526</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 11b: Correlations (Pearson’s r, two-tailed analysis) between CTB2 numeracy scores and work performance scores: racial breakdown

Table 11b above shows the results for the different racial subgroups of the sample. In terms of Pearson’s r, only the white subgroup showed any degree of correlation between CTB2 numeracy scores and work performance scores. This was a weak effect of 0.17, but the p-value (0.5) was too high to render any significance to this result. The remaining three race groups showed no correlation between the scores, and also no significant p-values.

Therefore it is safe to conclude that none of the racial subgroups showed any correlation between their scores on the CTB2 numeracy subtest and their later work performance scores.

One final correlation analysis was done in which all staff other than call centre agents were eliminated, such as team managers, secretaries and people being trained in financial services. The three main categories listed in section 4.2.3 were retained, namely permanent part timers, new accounts staff and customer service staff. Together, these three groups made up 88% of the original sample and comprised 265 participants. This subgroup can be considered highly representative of entry-level call centre agents at the
company under study. This secondary analysis was done using Microsoft® Office Excel (2003)\(^5\) and the result was as shown in Table 11c below.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Correlation</th>
<th>p-value</th>
<th>N (sample size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry-level staff</td>
<td>0.026</td>
<td>0.05</td>
<td>265</td>
</tr>
</tbody>
</table>

Table 11c: Correlations (Pearson’s r, two-tailed analysis) between CTB2 numeracy scores and work performance scores: entry-level call centre agents

The results showed that for this subgroup, there was no evidence of a statistically significant relationship between CTB2 numeracy scores and work performance scores. The default alpha of 0.05 was used.

Based on the various results reported above, it is safe to conclude that there is no statistical relationship between the two variables under study. This was the case whether the total sample was used, or separate gender, race or job category groupings.

4.5 Conclusion

This chapter outlined the biographical information of the sample, and all the statistical terms used were defined. Overall, the results showed that there is no statistically significant relationship between the CTB2 numeracy subtest scores and the work performance scores of call centre agents at the company under study. This lack of relationship between the two variables (test scores and performance scores) was indicated both by the extremely low correlation coefficients and also by the high probability values. The findings were essentially the same across all gender and racial subgroups of the sample, as well as for the entire sample when analysed as a single group.

The t-test was used to determine if there was a statistically significant difference between the results for the gender subgroups. The result revealed

---

\(^5\) MS Excel results were verified against the SPSS results by repeating 3 of the correlation tests carried out during the primary analysis. Pearson’s r produced by Excel and SPSS were identical in each case.
no significant difference, which suggests that these subgroups were drawn from the same underlying population group. ANOVA was not performed on the different racial groups because the sample sizes were so unequal.

The null hypothesis of the study asserted that there is no correlation between the CTB2 numeracy subtest scores and work performance ratings, while the alternative hypothesis asserted that the numeracy scores do correlate significantly with performance ratings. Based on the results of this study, the null hypothesis cannot be rejected, as no relationship was found to exist between the two variables. The following chapter will take this discussion further and will also present the limitations of the study, as well as recommendations for the future.
CHAPTER 5
Interpretation of the findings

5.1 Introduction

In this final chapter, the results reported in the previous chapter are discussed in more detail. An attempt is made to explain the findings theoretically and from a practical point of view. Conclusions about the findings are presented and the limitations of the research situation are discussed. The chapter concludes with some recommendations for future research, as well as specific recommendations regarding the use of the CTB2 numeracy subscale at the company’s call centre.

5.2 Discussion of results

In this research project, correlation analysis was used to investigate whether or not the CTB2 numeracy subtest scores of prospective call centre agents was related to the later job performance of those same individuals. Job performance was rated by the call centre managers some time after the candidates had been employed by the company.

The results reported in section 4.4 of the previous chapter showed that there was consistently no statistically significant relationship between the two scores (variables). This was the case for all race groups, both genders, and for the total sample as well as a sample that was more strictly representative of entry-level call centre agents. In other words, moderator variables such as gender and race appeared not to have any direct effect on the dependent and independent variables, as the results remained the same when analysing these subgroups separately.
Summary analyses of the participants’ biographical data also yielded some interesting information. The researcher did not have access to the exact racial or gender composition of the company’s staff at the time of the data collection in 2006, either for the entire company or for its call centre staff nationally. Nonetheless an attempt is made in the following sections to gain insight into how representative the sample was not only of the call centre staff but also in the South African context.

5.2.1 Gender breakdown

The gender breakdown (Table 9a, section 4.2) showed that the majority of call centre agents are female (79%) with only 21% being male. It is not clear what the reasons may be for this unequal gender split, since in the general population the ratio is fairly equal, with 51% being female and 49% male (Statistics South Africa, 2010). However, two tentative explanations are suggested.

Firstly, it may be that call centre positions are perceived as being more suitable for females than males, in the same way that several other occupations (e.g. nursing and teaching) are perceived as being particularly suitable for women. One of the reasons may be that the job does not require muscular strength.

Secondly, it is possible that the kinds of personality characteristics necessary to become a good call centre agent are found more often in women than men. These traits might include conscientiousness, agreeableness, emotional stability, good communication and extraversion (Bhattacharya & Bhattacharya, 2007; Skyrme, Wilkinson, Abraham & Morrison, 2005).

Section 4.3 reported the results of tests (F-test and t-test) which were used to find out if there were statistically significant differences in the data for the male and female subgroups. No significant differences were found, either in the
variance of these two sample groups or in their CTB2 numeracy scores. It therefore appears that the sample was drawn from a homogenous population group (Rosenthal & Rosnow, 1991; Kerlinger, 1986) which included both males and females, without their gender differences affecting the variables of interest to the study. Psytech-SA (2010) also reported no significant differences in the scores of men and women on the CTB2 numeracy subtest.

5.2.2 Racial breakdown

According to Statistics South Africa (2010), the racial composition of South Africa is estimated as follows: African 79.4%, Coloured 8.8%, Asian 2.6% and White 9.2%. The company under study has a policy of employing staff in accordance with these demographics so as to be completely representative of the populations which it serves (the company’s website, 2010). There are management structures and targets in place to ensure that this policy is translated into action.

The data was gathered for this research in 2006 and it is quite possible that in the years 2006-2010 the racial profile of the company may have changed slightly. At the time of collecting the data, it was obvious that efforts had been made to employ people in line with the general SA population racial distribution, but this objective had not been fully accomplished. Section 4.2.2. of the previous chapter showed that the sample included 55.7% blacks, 31.3% Coloureds, 7.3% Indians and 5.7% whites. These racial proportions are shown together with the SA general population estimates (from Statistics South Africa, 2010) in Table 12 below.

<table>
<thead>
<tr>
<th>Population group</th>
<th>% of SA population (2010)</th>
<th>% study sample (2006)</th>
<th>Difference in % points</th>
<th>Over / under-represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td>79.4</td>
<td>55.7</td>
<td>23.7</td>
<td>Under</td>
</tr>
<tr>
<td>Coloured</td>
<td>8.8</td>
<td>31.3</td>
<td>22.5</td>
<td>Over</td>
</tr>
<tr>
<td>Indian / Asian</td>
<td>2.6</td>
<td>7.3</td>
<td>4.7</td>
<td>Over</td>
</tr>
<tr>
<td>White</td>
<td>9.2</td>
<td>5.7</td>
<td>3.5</td>
<td>Under</td>
</tr>
</tbody>
</table>

Table 12: Racial composition of SA population and the sample
The biggest discrepancy was for blacks and Coloureds, with these two groups being either over or under-represented by more than 20 percentage points; blacks were under-represented while Coloureds were over-represented. Asians were slightly over-represented (by 4.7 percentage points) while whites were slightly under-represented (3.5 percentage points).

It is possible that black job applicants are still at a disadvantage when applying for jobs, including call centre jobs. This may be due not so much to a company’s lack of affirmative action as to educational and language barriers that continue to dog many of the previously disadvantaged communities. Language barriers may be a particular problem for black applicants to call centre positions, given that many businesses require English to be used as the main language of a workplace, and call centre employees spend most of the day on the phone. It is possible that Coloured communities tend to use more English than black communities, which might give them an advantage over blacks in terms of securing call centre jobs. Affirmative action policies require that blacks, Coloured or Asians are employed rather than whites, but the competition between members of those three “non-white” groups is still a hurdle on the individual level.

Of more direct relevance to the current research is the possibility that a lack of numerical ability, as measured by the CTB2, might also disadvantage black applicants to call centres. Once again it is not clear why blacks might be more disadvantaged in this area than Coloured applicants, other than because of unequal historical educational deficits.

All of the above suggestions are purely speculative, however, and further research would be required into these areas to validate or refute the ideas.
5.2.3 Participants’ positions

Section 4.2.3 indicated that while the bulk of the sample (88%, N=265) was made up of actual call centre agents, some participants held other jobs within the call centre, such as being administrative staff or team managers. Because these candidates had all been through the same selection process using the CTB2, and were all working in the call centre environment at Bands A or B, they were included in the sample.

As discussed at the end of section 4.4, a secondary analysis was carried out where all other job categories were removed. This left a sample of 265 individuals who were all working at entry level and only as call centre agents. The same results were found regardless of whether the original sample or this new subgroup was used for the correlation analysis, namely there was no statistically significant correlation between CTB2 numeracy scores and later work performance scores.

The null hypothesis, which stated that no relationship exists between call centre agents’ numeracy scores and their work performance scores, can therefore not be rejected. It appears that people’s ability to score highly on the CTB2 numerical ability subtest did not relate to their work performance once they were employed in the call centre. This finding confirms the call centre managers’ impressions that this particular subtest is not a good choice for screening job applicants for the call centre at the company under study.

5.3 Statistical analyses

5.3.1 Internal consistency reliability

Sections 2.13 and 2.14 of Chapter 2 presented the results of reliability studies done on the CTB2 in South Africa by Psytech-SA. The important numerical figure in these results is Cronbach’s alpha, which reflects the instrument’s internal consistency reliability. As noted in 2.13.1, a Cronbach’s alpha of at
least 0.70 is needed to show internal consistency reliability, while Cronbach's alpha of 0.85 to 0.90 is considered excellent (Rosenthal & Rosnow, 1991).

It was noted that Cronbach's $\alpha$ for the CTB2 numeracy subtest was adequate for the total South African sample (Psytech-SA, 2010). However, an analysis for separate racial groups showed that Cronbach's $\alpha$ was too low for both the black and Coloured subgroups. This finding suggests that the test was normed on populations that are more similar to South African whites and Asians, and less similar to SA blacks and Coloureds. The numeracy subtest may need to be refined for use in the South African context if the internal consistency reliability is to be improved for these indigenous groups.

In the current study there was no attempt to establish the internal consistency reliability of either the CTB2 battery or its numeracy subscale, when used with the company's call centre sample. The reason for this was the researcher was only granted access to the total numeracy subtest scores. Testing for internal reliability would have required raw data that contained the scores for each question item as well (Rosenthal & Rosnow, 1991). It was therefore assumed, based on the findings reported by Psytech-SA (2010), that the numeracy subscale may have some problems with internal reliability when used with the current sample.

5.3.2 Descriptive statistics and t-tests

The descriptive statistics for the mean CTB2 numeracy scores of the sample, as well as the results of the t-test for gender subgroups, were reported in the previous chapter. These results do not require extensive interpretation, and the reader is asked to refer back to that section if necessary. It was also mentioned that no ANOVA was performed due to the unequal sizes of the racial subgroups of the sample.
5.3.3 Correlational analysis

Section 4.4 of the previous chapter presented the findings of the correlation analysis. This formed the core of the research project, as it provided a way to assess whether or not a relationship existed between the study variables.

It was noted that Pearson’s r was consistently too small to indicate any covariance between the CTB2 numeracy scores and work performance scores. This was the case for the overall sample as well as the black, Indian and Coloured subgroups. The only exception was the white subgroup, which had an r of 0.17. This might have indicated a small amount of correlation, with a very weak effect (Rosenthal & Rosnow, 1991), but the large p-value of 0.53 rules out even this interpretation.

The results of the current study therefore confirm that it is questionable to use the CTB2 numeracy subscale when screening call centre candidates. Using the approach of criterion validity, it can be seen that the scores which people obtain on the CTB2 numeracy subscale do not appear to relate to their ability to become competent call centre agents. The null hypothesis was therefore not rejected. There was no statistical evidence to indicate a significant relationship (either positive or negative) between numerical ability and work performance as a call centre agent.

One of the implications of this finding is that the company’s human resources department will need to carefully consider whether the CTB2 numeracy subscale test is adding any value to their selection process. It may be the case that this particular test should not be used with call centre candidates. A thorough job analysis of the kind of work which a call centre agent does must be done (Adler, 2007), and it must then be determined whether in fact there is any need for strong numerical ability and skills in this type of work.

A second and related implication of using the CTB2 numeracy test to assess call centre candidates is that people who are in fact suitable for this role may be screened out, if they score poorly on the numeracy subtest. It was not
clear to the researcher exactly how much weight was placed on a candidate’s CTB2 numeracy results by the human resources decision-makers. This aspect would need further investigation in its own right, for example by interviewing human resource staff and call centre managers, as well as reviewing the company’s policy documents in this area. This was beyond the scope of the present study. Should the CTB2 numeracy subtest continue to be used by the company when screening call centre applicants, such research would need to be done in order to justify this decision.

The current study has suggested that the use of the CTB2 numeracy subtest is inappropriate for identifying the type of person who would make a good call centre agent. The decision to hire or reject call centre applicants cannot be made based on this specific test alone. The use of more relevant tests such as verbal reasoning and data checking should be given preference and the candidate must also be interviewed on a one-to-one basis.

5.4 Ethical considerations

Participants in the current study were not aware that their CTB2 scores and work performance scores were being used for research purposes. For this reason the data was treated as highly confidential, and only the researcher and her supervisor had access to the raw data for purposes of statistical analysis. No information has been reported that might have identified an individual or group. Furthermore, the name of the company under study has been omitted from this report for ethical reasons.

5.5 Limitations of the study

The research was limited in the following areas:
5.5.1 Limitation of the literature

In terms of the literature review, limitations were identified with regard to the availability of publications and articles about the numerical subtest of the battery under study (CTB2). Specifically, there was very limited literature on the validity and reliability of the numerical subtest battery in an African context. There is also little published evidence about the use of the CTB2 subtests in the retail sector. Due to a lack of sufficient literature in the early stage of the research, the researcher consulted Psytech-SA (the company holding the licence for the CTB2 in South Africa) and requested relevant information which would benefit the study. At a later stage, some of this information became more easily available on the Psytech-SA website.

5.5.2 Limitations of the data

The study was carried out using only quantitative data and methodology. It would have been beneficial to use qualitative surveys and face-to-face interviews with the key players as well, to get a better picture of the situation being studied (Kerlinger, 1996; Babbie, 2007). The stakeholders, including the call centre agents themselves, would then have felt themselves to be part of the study and they may have voiced highly pertinent comments and opinions. However, this option was beyond the reach of the researcher due to the amount of time which both she and other employees would have had to take off from work while attending such interviews. Furthermore the additional data analysis would have extended the time-frames of the research project beyond what was feasible.

It was also noted in previous sections of this report that the data itself was limited to only two sets of scores, namely the CTB2 numeracy scores and later work performance scores. Ideally, the researcher would have had at her disposal not only the full results (raw data) for all items of the numeracy test, so as to be able to assess the internal consistency reliability, but also the scores on other CTB2 subtests.
5.5.3 Limitations regarding the findings

The study was conducted at a large retail company that specialises in clothing and textiles in southern Africa. A sample of entry-level call centre employees was used. The findings of this research cannot be generalised to any other contexts or population groups, and are relevant only within a South African clothing and textiles call centre environment. Therefore, the interpretation of the results should be limited to the group studied at the time of this research.

Specific problems with the type of data and sampling used in this study also mean that the findings should not be generalised to other contexts. The limitations in this regard were as follows:

- The sampling method entailed the use of a purposive convenience sample plus a systematic selection strategy. Pre-selection of participants took place in that only some candidates working at the company’s call centre, on the A and B Bands (i.e. the lowest or entry-level bands) were included in this study. This was the particular group that was being tested with the Clerical Test Battery (CTB2). Other applicants in other areas of the company were assessed with different tests, such as the Cognitive Process Battery.

- One set of data contained an unknown element of subjective bias. That was the work performance scores, which were influenced to some extent by managers’ personal opinions. However, the scores were also determined by objective measurement of the employees’ productivity and product knowledge. Each employee was given a final rating in the form of a single numerical score; therefore the dataset was deemed to be suitable for quantitative research.

Despite these limitations, the current study contributes to understanding more about the use of the CTB2 numeracy subtest in a South African call centre environment. This is especially valuable given the lack of pre-existing
literature in this area. The findings are also of value to the company’s human resources department, in terms of guiding future decisions about which instruments to use for the screening and selection of call centre employees.

5.6 Dissemination of results

The outcomes of this study will be shared with the assessment managers and practitioners at the company under study.

5.7 Recommendations

The following recommendations and suggestions are offered to human resource managers at the company under study, and more generally to researchers in relevant fields:

1. The use of the CTB2 numeracy subtest with call centre candidates should be carefully reviewed. It is possible that an alternative instrument might provide a more relevant assessment in this area of candidates’ ability, or that an additional test battery should be used in conjunction with the CTB2 to provide more trustworthy results.

2. The assessment department should identify other instruments or tests that could replace or supplement the CTB2 numeracy subtest. One possibility that should be investigated is the Learning Potential Computerised Adaptive Test (LPCAT), which is designed more specifically for staff such as call centre agents.

3. Decisions should not be taken about the hiring or repositioning of staff based only, or mainly, on their CTB2 results.

4. The assessment department should engage in discussions with call centre managers, so as to identify the core qualities and skills that are needed for the role of a call centre agent. These discussions should form part of a thorough job analysis of call centre positions (Adler, 2007).
5. The job analysis should inform the assessment process. Instruments (or subtests) which can demonstrably measure the relevant skills and abilities should be selected, rather than simply because they may exist as part of an assessment battery.

6. Another feasible option could be to outsource assessments to more experienced psychologists who are specialists in test development, administration and reporting. Such specialists could also develop an "in-house" instrument for the company to use, based on similar tools used at other companies in the clothing and textile retail sector (Ngidi, 2009).

7. Future research in this field should include qualitative data in order to understanding more about how call centre agents perceive their own roles and responsibilities. Such research can also yield valuable insight into the perceptions of call centre managers.

8. A comparison study could be conducted by taking a sample from another similar organization in South Africa, and comparing CTB2 numeracy scores with work performance scores as was done in the current study. The results of this hypothetical study should also be compared with those of the current study. This process would highlight any general trends and may suggest more conclusive findings.

5.8 Conclusion

This chapter presented a discussion of the results and findings of the study, and an overview of the academic and methodological limitations thereof. The statistical analysis found no evidence of a significant relationship between CTB2 numeracy subtest scores and later work performance scores of call centre agents. This suggests that the use of the CTB2 numeracy test with call centre candidates is limited and questionable.

The recommendations and suggestions arising from this research include the possibility of ceasing to use the CTB2 numeracy subtest when screening call centre agents. Other more appropriate instruments should be considered,
either as a supplement or as a replacement. Qualitative research should ideally take place in an effort to find out exactly how call centre agents and their managers view the roles and responsibilities of a call centre agent. In addition, a thorough formal job analysis needs to be done, and the results of this should guide the choice of assessment instruments. Several other suggestions have been made to improve the assessment process of call centre agents at the company.
REFERENCES


Company Database: Information from the company under study pertaining to employees during 2006. Date of Access March 2007. In addition, the company’s website (accessed January 2007 to December 2010).


Cunningham, E. G., Brandon, C. M. & Frydenberg, E. (1999). Bright Ideas: A school-based program teaching optimistic thinking skills in pre-


Foxcroft & G. Roodt (Eds.), An introduction to psychological assessments in the South African context (pp 11-33). Cape Town: Oxford University Press.


Indiana University website, biographies accessed from at Human Intelligence: biographical profiles on 10 December 2010:
http://www.indiana.edu/~intell/yerkes.shtml
http://www.indiana.edu/~intell/binet.shtml
http://www.indiana.edu/~intell/terman.shtml
http://www.indiana.edu/~intell/rthorndike.shtml
Hypothesis Testing summary
http://www.indiana.edu/~educy520/sec5982/week_12/hypothesis_test_summary011109.pdf


Kehoe, L. (2002). Psychometric methods in marketing and organizational data: Generalizing data from research to practice. California: University of Southern California


