The Effect of Mode of Test Administration on Computerised Assessment Results Using Proctored and Unproctored Test Administration Procedures

by

Francina Helena Nel

submitted in accordance with the requirements for

the degree of

Master of Arts

in the subject

Industrial and Organisational Psychology

at the

University Of South Africa

Supervisors:

Prof Marié de Beer

Ms Nomfusi Bekwa

December 2012

DECLARATION

December 2012

Student number: 32481691

I, undersigned, hereby declare that the dissertation entitled **"The Effect of Mode of Test Administration on Computerised Assessment Results using Proctored and Unproctored Test Administration Procedures"** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

_____ _____

Francina Helena Nel **Date**

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following individuals who supported and inspired me in completing this dissertation:

- My supervisor, Professor Marié de Beer for your invaluable guidance, your time and effort and for challenging me to learn throughout the process of completing research. Most of all thank you for being an inspiration and for sharing your passion for research with me.

- My co-supervisor Nomfusi Bekwa for your encouragement, assistance and support.

- My husband, André Nel, for your love, patience and support over the years and your ever amusing sense of humour much needed when times were tough.

- My parents, Chris and Dorothea Marais, and my parents in law, Val and Willie Nel, for being wonderful role models, for your love over the years and for your willingness to assist and support where necessary.

- Last but most important, my two little children Ruwan and Begonia Nel for being so supportive, for allowing me time to follow a dream, but most of all for your ever loving smiles no matter the circumstance.

SUMMARY

The purpose of this research was to investigate the effect that mode of test administration could have on computerised assessment results involving proctored and unproctored test conditions. Two South African test instruments, the Learning Potential Computerised Adaptive Test (LPCAT) and the Career Preference Computerised Adaptive Test (CPCAT) were used in the study. A quantitative, quasi-experimental design was used, and a convenience sample for LPCAT (N=82) and CPCAT (N=81) consisted of employees in the hospitality industry. Using a within-participants design, the dependent t-test was used for statistical analysis.

For the total group the LPCAT results yielded no statistically significant differences between the mean scores for the two different modes of administration. For the total group the CPCAT results yielded statistically significant differences in the mean scores per mode of administration for five out of 34 dimensions, however, for the majority of the CPCAT sub-dimensions, the mode of administration did not impact on results.

It was concluded that mode of administration did not impact on the cognitive test scores and only to a very limited degree on the non-cognitive test scores. Based on the results the null hypotheses for the effect of mode of administration were not rejected.

KEY TERMS

Computer adaptive testing, CPCAT, internet-based testing, LPCAT, mode of administration, proctored, unproctored, dependent t-test.

TABLE OF CONTENTS

List of Tables

CHAPTER 1

SCIENTIFIC ORIENTATION OF THE RESEARCH

Due to international and national developments related to computer-based and internet-based testing in the field of psychometrics, further research related to unproctored versus proctored administration was deemed necessary. The South African context consists of socio-economic and cultural issues, which needs consideration during psychological assessment, and as technology develops, research in computer and internet testing could be beneficial for future psychological assessment practices in South Africa. In the present study, the focus was on identifying whether the mode of administration, specifically the unproctored mode compared to the proctored mode can have a statistically significant effect on computerised test results. In Chapter 1, an introduction to and motivation for the study will be provided, the problem statement formulated, the paradigm perspective discussed, and the research methodology and ethical considerations reviewed. The general outline of the thesis is provided at the end of the first chapter.

1.1 INTRODUCTION

The 21st century work environment has changed radically due to the progress made in the field of the information technology and the internet (Bartram, 2000; Joubert & Kriek, 2009). As a result, information can be disseminated with incredible speed, and exciting business opportunities have developed worldwide and commensurately, the level of customer demand for products and services has increased (Foster, 2010; Lievens & Burke, 2011). Bartram (2006) explained that internet-delivered testing has increased as the market for internet and computer-based testing has developed quickly. The internet has allowed for new technological innovations in psychometric assessments (Eid & Diener, 2006; Joubert & Kriek, 2009).

Joubert and Kriek (2009, p. 79) reported "Over the past five years there has been a marked increase in employment tests available on the internet for recruitment, selection

and development". Due to the availability of the internet platform an increasing number of organisations have begun to recruit and select individuals via the internet (Bartram, 2000; Hense, Golden & Burnett, 2009). Lievens and Harris (2003) indicated that internet recruitment can be defined as when an individual relies on the internet when applying for a job. From a human resource perspective, to manage recruitment and selection processes efficiently in practice, implementations of technological testing infrastructures have become important (Bartram, 2000, 2006). A principal reason why internet-based methods became popular was the accessibility by means of which large numbers of participants could be assessed quickly (Bartram, 2006; Eid &Diener, 2006; Joubert & Kriek, 2009). Internet tests can be taken at places and times which are convenient to the test taker – an alternative form from supervised modes (Tippins et al., 2006). Also Foster (2010) has explained that testing delivered though internet browsers has the advantages of not being limited to fixed locations.

Due to the internet, business boundaries have merged and the internationalisation of testing has become possible which has resulted in a need for international control and some agreement on best business practice related to the internet (Bartram, 2006). Computer-based testing as well as internet-based psychological assessment, could not be ignored as a means to keep abreast with technology or the opportunities it offered. Bartram (2006) referred to the implication of new technology which enhanced development and professionalism within a globalised context but with focus on levels of control and standards for internet testing.

Despite the benefits that technology provided to the field of psychometric testing, many problems including cheating and security issues have been reported (Arthur, Glaze, Villado & Taylor, 2009; Bartram, 2000; Tippins et al., 2006). Aspects such as the level of control over the testing conditions and the possibility that tests could be made available to unqualified people raised concerns (Joubert & Kriek, 2009). The nature of the internet allowed for faster as well as easier selection and testing processes, however, consequently created ethical and legal challenges for the qualified practitioner (Foxcroft & Davies, 2006).

Tippins et al. (2006) defined proctored testing as an assessment event that is monitored, where participant identity is verified and where the degree of standardisation is high. On the other hand, unproctored testing is defined as an assessment event where candidates do not have a human proctor present during testing and where the standardisation of the testing environment is unknown (Tippins et al., 2006). The appropriateness and practical implementation of psychometric testing in the unproctored setting becomes the responsibility of the registered psychologist (Tippins et al., 2006). Those industries serious about keeping up with modern technology and related developments have had no choice but to define and re-define methodologies to retain their relevance in the fast moving and constantly changing technological world. In order to protect the nature of ethical practice in psychometric testing, whilst providing and using quality test products, it remains important for professionals to decide on using tests that are proven to be valid and reliable specifically on the internet platform.

1.2 MOTIVATION FOR THE RESEARCH

In the process of reviewing and further validating the Learning Potential Computerised Adaptive Test (LPCAT) as well as the Career Preference Computerised Adaptive Test (CPCAT), the test developer endeavoured to update the computerised version of LPCAT and to finalise CPCAT with the aim of making the tests available via the internet (De Beer, 2012). Such developments are partly in an effort to keep up to date with new technology and to provide updated products to test users, but also to manage the instruments from a systems perspective. The LPCAT is being updated to be compatible with new technology and operating systems (De Beer, 2012). In addition to the development of an internet test version of the LPCAT, the aim is also to improve the automated processes of test administration (De Beer, 2012). The internet versions of LPCAT and CPCAT would improve the overall management of the tests, make available more comprehensive research data and would be of use to test users, during test scoring and in terms of compatibility of software. With the aim of being able to make informed research-based decisions about the use of the aforementioned tests in the

future, this research was specifically focused on investigating the possible effect that mode of test administration might have on test results of the LPCAT and CPCAT as representatives of a cognitive and a non-cognitive measure respectively.

The South African legislation requires that test publishers only publish and offer internet-based tests once evidence of sufficient psychometric support has been provided (Foxcroft & Davies, 2006). Thus the aim of this research has been to provide information related to the two test instruments in compliance with South African legislation and the current test developments.

1.3 THE RESEARCH PROBLEM STATEMENT

Joubert and Kriek (2009) mentioned that psychological assessment has evolved along with the information technology. The benefits that computerised tests held for psychometric testing resulted in many paper-and-pencil tests being transformed into computer-based tests as the default medium (Bartram, 2000). Eid and Diener (2006) explained that as previously developed paper-and-pencil tests became computerised, whilst such tests offered little in terms of improved psychometric properties, certain administrative advantages transpired. According to Eid and Diener (2006), it is important that test scales should be validated for both paper-and-pencil as well as computer versions to ensure equivalent results and validity. Bartram (2000) mentioned that the research field has been dominated by the parallel use of computer-based and paper-based versions of the same tests. Studies in the past, specifically on personality tests, have been focused on measurement equivalence and equivalence of psychometric properties obtained following the transfer from paper-and-pencil tests to computer-and internet-based tests, yielding comparable results (Bartram & Brown, 2004; Salgado & Moscoso, 2003). Salgado and Moscoso (2003) investigated the equivalence of mean scores and standard deviations, as well as reliability coefficients and factors structures between the internet-based version and the paper-and-pencil versions of the Personality Inventory of Five Factors IP/5F). The results indicated that both versions of the IP/5F could be used because high equivalence of the measures

was reported. Joubert and Kriek (2009) conducted research on the construct equivalence of the Occupational Personality Questionnaire 32 (OPQ32i). Joubert and Kriek (2009) aimed to investigate the degree of equivalence of the OPQ32i when administered unproctored comparing the mean scores, reliabilities and analysis of the covariance structures. Results showed similar Cronbach's alpha coefficients and covariance structures for the OPQ32i.

Whilst some studies supported the equivalence of paper-and-pencil and computerised versions of the same tests, other studies did not confirm the equivalence of different versions. Arce-Ferrer and Guzmán (2009) studied the equivalence of raw scores from classical test theory reliability and factorial validity frameworks for the Ravens Standard Progressive Matrices (RSPM). The RSPM was researched for test equivalence between the paper-and-pencil and the computer-based form; the findings supported statistical equivalence of raw scores for both total score reliability, single factor structure and a preference for computer-based testing, therefore indicating that the different versions could be considered equivalent. Furthermore computer-based versus paper-and-pencil assessments of the third edition Self Descriptive Questionnaire (SDQ-III) were researched and yielded comparable results (Vispoel, 2000). The ICES PLUS has been researched for test equivalence from paper-and-pencil tests to computer-based tests and the findings indicated that, whilst overall ICES scales showed evidence of equivalence, the numeric, verbal and spatial ability scales lacked equivalence (Coyne, Warszta, Beadle & Sheehan, 2009). Eid and Diener (2006) reported on studies where computer-based speeded tests were not equivalent to their paper-and-pencil tests versions, but carefully developed power tests were equivalent.

In the domain of computerised assessment, many studies have been conducted on comparing paper-and-pencil results to computerised test results, but a literature search yielded fewer results for computerised adaptive test (CAT) results. Beaty, Dawson, Fallaw and Kantrowitz (2009) indicated that CAT could be a promising strategy to mitigate cheating and a viable method of rotating content which could enhance test security.

With the increase in the use of the internet, it became possible to make use of computer-based tests via this medium, but regulation of test use, professional standards provided by bodies such as the International Testing Commission (ITC) and specific guidelines for modes of administration were needed (Tippins, 2009).

Naglieri et al. (2004) conducted research on how the internet influences the practice of psychology as it relates to assessment. Traditional static paper-and-pencil test administration conducted in a supervised and well controlled environment (proctored), gradually transformed to computer-based and internet-based versions which could include unproctored mode (Eid &Diener, 2006; Joubert & Kriek, 2009). Computer-based testing as well as unproctored internet testing (UIT) became the alternative or possible future replacement for the traditional methods of test administration (Bartram, 2000).

According to Naglieri et al. (2004), concerns existed around test validity, test security, cheating, technical challenges and cultural challenges, all issues related to psychological assessment in general but more specifically in recent years to UIT. Bartram (2000) also referred to issues of good practice which included confidentiality, authentication and control over the test taking conditions.

The debate around UIT increased and the need for in-depth research into the topic of internet, computer-based and unproctored testing, including cognitive ability in the field of psychology, developed (Lievens & Burke, 2011). It seems that the benefits of internet testing in essence allows for rapid growth in unproctored testing methods. Many future opportunities exist, however whilst many organisations would use online testing processes and Industrial Organisational Psychologists use the internet for workplace interaction (Naglieri et al., 2004), debates indicate that the use of UIT is not yet fully accepted by all practitioners (Tippins et al., 2006). According to Beaty et al. (2011), a variety of variables impact on test validity when the proctor is removed from the test condition, and there have only been a few published studies showing what happened to predictive validity when a test was taken via the internet and offsite, (Kaminski & Hemingway, 2009, Lievens & Burke, 2011; Weiner & Morrison, 2009). The challenges that practitioners faced are to ensure that, in the application of unsupervised testing,

test validity and reliability are not compromised at any point during the unproctored internet-delivered process.

One of the main concerns to date has been whether internet-based tests, which are more likely to be used in the unproctored mode of administration, remain valid as well as reliable (Naglieri et al., 2004). In order to regulate computerised and internet-based testing, the International Test Commission (ITC) provided International Guidelines on Computer-Based and Internet Delivered Testing (ITC, 2005) distinguishing between open, controlled, supervised and managed modes of testing. Also in the most recent version, ITC guidelines for Quality Control in Scoring, Test Analysis and Reporting of Test Scores (ITC, 2011) it was suggested that practitioners had to have a broad understanding of quality control practices, and this would be of critical importance for tests to be used ethically, accurately and responsibly. The concerns that exist around the unproctored mode of testing and validity have warranted extensive research (Beaty et al., 2011).

As psychometric tests in general aim at providing objective information for purposes of occupational and other assessments for further decision making, research on the possible effect that the presence or absence of the proctor may have becomes necessary (Bartram & Brown, 2004). Tippins et al. (2006) defined high stakes testing as those testing situations where the consequences of testing affect other people or institutions beyond the individual who is tested. In the field of Industrial and Organisational Psychology testing for selection and placement, training opportunities or promotion qualifies as high stakes testing (Tippins et al., 2006). Cognitive tests which were viewed as high stake tests were specifically considered as being at risk of losing reliability and validity if used in UIT settings (Tippins et al., 2006). The practitioner is expected to consider the nature of the test (cognitive or non-cognitive) and the use of the test (selection or development) before allowing for internet-based testing (Tippins et al., 2006). Eid and Diener (2006) referred to multi-method measurement approaches which could indicate possible combinations of proctored and unproctored assessment techniques.

Bartram (2000) explained that very few examples exist of tests that have been published as computerised versions and not also produced as paper-and-pencil tests. Pertaining to this study, the literature search of tests that were designed in computerised form, and not initially designed in paper-and-pencil form or used in different modes of test administration, led to no results. Thus to refer to previous research this study addresses high stakes testing with the use of a cognitive test, computerised adaptive tests, online simulations of test administration and mean scores of the same participants are compared.

It is predicted that UIT is likely to increase in coming years (Tippins et al., 2006). More specifically for the South African context internet testing for disadvantaged groups could raise ethical and legal concerns (Joubert & Kriek, 2009; Tippins et al., 2006). In a recent national census conducted by Statistics South Africa in 2011 it was indicated that 64.8% of households in South Africa had no internet access. Out of the 35.2% who had internet access 8.6% had access from home, 4.7% from work, 16,3% from cell phones and 5.6% from elsewhere (Statistics South Africa, Census 2011). With reference to psychological assessment cell phone use is questionable and in general not used for psychometric testing which implies that only 18.9 % of South Africans possibly have access when having to complete online tests. This indicates that a high percentage of South Africans will be excluded from applicant pools where internet selection or testing is required. The shift towards UIT implies a certain level of change in psychological assessment (Naglieri et al., 2004) when human proctors are not present. Whilst the UIT continuum varies between positive benefits and questions around reliability (Lievens & Burke, 2011) psychological assessment is continuously changing.

Joubert and Kriek (2009, p.79) indicated that “there were no studies that examined measurement equivalence for paper-and-pencil versus unproctored internet test administration for previously disadvantaged groups”. Tippins et al. (2006) discussed the possible limited internet access that disadvantaged groups had and how those groups could be excluded from applicant pools during selection and recruitment processes if online procedures were to be used. This research will provide information regarding test results under different testing conditions on the LPCAT and CPCAT, tests that were not

designed for paper-and-pencil format and which exist only in computer-based form. Up until this point, no South African research comparing the effect on test results using the proctored and unproctored mode of administration has been conducted on either the LPCAT or CPCAT.

The LPCAT is considered a culture fair and language fair measure of learning potential by means of non-verbal figural reasoning (De Beer, 2000, 2005) it is therefore typically used for purposes such as:

a) recruitment and selection;
b) training and development; and
c) bursaries or learnerships.

CPCAT provides vocational information in terms of career preference (De Beer, 2011). The world of work has changed as the few existing trades and specific paths have now been replaced by multiple entry points, with career interest as one of the cornerstones of career counselling (De Beer, 2011). The dynamic three dimensional model of CPCAT which focuses on career fields, activities and environments allows for rich feedback on career preference (De Beer, 2011).

Future developments for LPCAT will include an internet-based analysis program and internet-based test administration (De Beer, 2012). The need for research on computerised tests, in this case computerised adaptive tests (CAT) administered via the internet was simulated to answer questions about the possible effect of the mode of administration on assessment results. Guidelines provided by the ITC (2005) for internet and computer testing included that studies of test equivalence and norming should be conducted over the internet representing those non-standardised or unproctored conditions that an intended target population will experience. In addition, evaluating research on the reliability and validity of the CPCAT could provide valuable comparative information to the test developer in terms of the psychometric properties of the instrument with the sample used in this study. To understand the need for the study, the appropriateness of psychometric testing in the organisational context and the implications of psychometric test use have to be considered. When practitioners started

having the option of using tests available on the internet as opposed to the traditional paper-and-pencil methods, the comparative validity of such a new method (the internet) needed to be investigated before decisions could be made on which method to use (Tippins et al., 2006). Similarly, the comparative results obtained with respectively the proctored and unproctored mode of administration on computerised and internet administered tests needed to be investigated and hard evidence provided (Tippins et al., 2006). The purpose of this study was to contribute to the field of Industrial and Organisational Psychology; more specifically psychological assessment and personnel psychology by means of an empirical investigation of test validity and the possible effects of the mode of test administration on test results obtained. The reason for including not only the field of psychometrics but also personnel psychology is because the fast-pace recruitment environments are likely to use psychometric tests online (Bartram, 2000; Kaminski & Hemingway, 2009; Lievens & Burke 2011; Lievens & Harris, 2003).

This study aims to provide further information based on empirical research results, which could assist practitioners when having to make choices as to the modes of administration.

1.4 PARADIGM PERSPECTIVE

This study was conducted within the theoretical paradigm of humanism in the discipline of Industrial and Organisational Psychology, more specifically the psychometrics field, and to some extent personnel psychology which relates to selection and placement procedures and often includes psychometric testing. Brockett (1997) explains that in psychology psychoanalysis and behaviourism are two schools of thought whereas humanism is known as the third force. Humanism which had great influence from Carl Rogers and Abraham Maslow's theories is concerned with learning, self actualisation, personality, motivation and potential, (Brockett, 1997) Humanism entails affect and cognition, intellect, feelings and emotions which forms part of metal health and educational practice in humanism.

In addition to humanism, psychological assessment included beliefs as well as hypotheses about behavioural problems and its relative importance, (Haynes & O'Brien, 2000). Behavioural assessment, personality assessment, intellectual or cognitive assessment and the causal variables in assessment could affect behaviour (e.g., early learning experiences, genetic factors, response contingencies) (Haynes & O'Brien, 2000).

The behavioural assessment paradigm as explained by Haynes and O'Brien (2000) stressed the use of well validated assessment instruments and assumed that social or environmental and cognitive variables are often central sources of behaviour variance. Behavioural assessment could be used to gather data for pre-intervention assessment, to evaluate effects of treatment and to analyse the conduct of individuals in terms of basic behavioural research.

Relevant to this study is the behavioural assessment because the test environment namely proctored and unproctored assessment which was manipulated, could have possibly affected participants through internal or external stimulus.

The empirical paradigm used was the positivism paradigm from the French philosopher Agustus Comte, who believed that knowledge can be obtained through experience and observation. In essence Comte believed that societies pass through three stages namely theoretical explanations, metaphysical or physiological stages and then positivism where scientific explanation is the rule for application related to human behaviour (Willis, 2007). The positivism paradigm was relevant for this study for the interpretation of empirical social science data (Terre Blanche, Durrheim & Painter, 2006) and as Krauss (2005) explained the objective of such a study is independent of the researcher, and knowledge is identified through measurement or direct observations. The purpose was to use quantitative methods to uncover laws regarding human behaviour within the science of Industrial and Organisational Psychology, and specifically to investigate the effect of different modes of test administration on computerised test results.

1.5 OBJECTIVES OF THE RESEARCH

In general the objective of the study was to identify whether mode of test administration, which entailed proctored and unproctored testing respectively, could affect test results of a cognitive and non-cognitive measure of the same participants using a repeated-measures design.

1.5.1 The research question

The study aimed to answer the following research question: Does the mode of test administration affect computerised assessment results when proctored and unproctored test administration procedures are used?

1.5.2 Variables

Salkind (2009, p. 22) explained the independent variable as "that which is manipulated or changed to examine its effect upon the dependent variable." The dependent variable is the measured outcome of the manipulation. The study was explanatory to determine the effect of the independent variable (mode of administration) on the dependent variable (test results on a cognitive and non-cognitive measure respectively).

1.5.3 Aims of the research

The aim of this study was therefore to investigate the effect of the mode of test administration on computerised assessment results using proctored and unproctored test administration procedures on a cognitive and a non-cognitive instrument respectively. The findings of the study should not be generalised to the larger population due to sample size limitations and the lack of representivity due to actual work related complications and convenience sampling used. If no differences were found between the results obtained by means of the proctored versus the unproctored assessment modes, it would indicate that the test providers of the LPCAT and CPCAT could consider either method of administration for the instruments concerned and could provide such information in the guidelines for the use of the measures. If results were to differ, decisions about administration procedures during online testing could be more prescriptive, and practical steps would need to be considered to safeguard the tests

from being used in a context such as unproctored testing, which could compromise the validity and interpretation of scores.

The literature review aimed:

1) to define the relevant variables of mode of administration (proctored and unproctored) and test results in this study;
2) to discuss research conducted in the past that is relevant to the topic of mode of test administration and psychological assessment results with reference to UIT;
3) to clarify core concepts relevant to this study such as proctored and unproctored modes of administration, computer-based testing, computerised adaptive testing, and internet testing and also to clarify the advantages and disadvantages of computer-and internet-based testing and to report on and summarise the findings on these aspects to date; and finally
4) to provide information that can be used when having to decide on the mode of test administration.

The aims of the empirical study were:

- to determine the effect of mode of administration on LPCAT and CPCAT results respectively by statistically comparing the proctored and unproctored mean scores of the same group(s) ; and
- to determine if sequence effect played a role in the study.

1.5.4 The research setting

The organisation in which the research was conducted is a Hotel and Golf Resort in South Africa. Arrangements included approval for ethical purposes from the company, HR Manager and CEO as well as informed consent obtained from the participants related to security, confidentiality and the dissemination of results. The organisation and employees who participated in this study stood to benefit from the assessment made available to them of the LPCAT and CPCAT, which provided information on learning potential and career preference for training and development proposes, as well as for personal development or possible career guidance purposes. The participating

organisation specifically requested that employees within the Patterson grading A1-B3 participate in the study, as the identification for training and development opportunities for these grades in relation to the South African NQF levels would support the future planning related to the workplace skills plan and other growth and development opportunities. Some students that were completing learnerships at the time also participated.

Prior to the commencement of test administration procedures, the relevant departmental managers were informed about the study. They were asked to make the employees in their departments available for testing. In terms of ethical considerations, employees were told that testing was voluntary and that they could withdraw from the study at any point in time. Managers were to encourage participation in the study for those employees within the Patterson grading bracket. However several factors limited the sample size such as staff being on leave, hotel events taking place, work pressure and a protected strike when participants were being selected. In general employees within the Patterson A1-B3 grading system participated, and this meant that some of the participants had formal education lower than Grade 10. Whilst the LPCAT entry level for the version where standardised test instructions on the computer screen is at a Grade 10 level or mid-secondary level, the test adapts to the level of performance of the individual on the non-verbal figural pattern items. Those with lower than grade 10 level grades could possibly benefit from learning opportunities within the organisation such as completing matric; and were still included in the study. The LPCAT language score was used to identify whether participants below Grade 10 might have struggled to understand test instructions and terminology used during instruction and feedback on the example items. A person who did not have formal education at Grade 10 level would have received a first question at a Grade 10 level and could possibly have dropped in performance if the initial first question was not answered correctly. However, due to the adaptive nature of the LPCAT, questions of difficulty level commensurate with the performance level of the individual which would be interactively selected and administered. This could possibly have had an effect on initial anxiety levels; however with the difficulty level adjusted, the participant could have overcome such anxiety

during the pre-test. For CPCAT, De Beer (2011) investigated the tests' psychometric properties by assessing children at Grade 9 and Grade 11 levels, and whilst the limitations of assessing high school children were acknowledged, indications were that children at these above-mentioned school levels could understand the questions and were able to complete the CPCAT and the CPCAT showed good psychometric properties (internal consistency reliability and construct validity) for this group (De Beer, 2011).

Participants in the present study were asked to sign a consent form to provide written informed consent. Because of the comparative aim of the study, the results of particular employees were only included in the study if they had participated in both test administration conditions of proctored and unproctored testing. It was agreed that if results were to differ on the cognitive test, the best result would only be reported to the HR manager as well as to the employees and that result would be used for decision-making related to training or development. The CPCAT non-cognitive results of both test sessions were communicated to participants as it was anticipated that a self-rating questionnaire's results might vary. In feedback sessions the difference in CPCAT results was explained for the interpretation thereof.

The researcher is a registered psychometrist, experienced and accredited to administer and interpret LPCAT as well as CPCAT test results, which improves the credibility of the testing. It should be noted that a pilot study was conducted prior to the actual study commencing. Two people participated in the pilot study; both were Afrikaans females one with Grade 7 level of qualification and one with a Grade 10 level of qualification. The person with Grade 7 qualification indicated that the booklet information which provided steps for the CPCAT log on process (controlled mode) was not clear. Neither participant in the pilot study reported difficultly with the LPCAT instructions on the screen; however the person with Grade 7 reported some anxiety initially.

The participants received additional instructions in a booklet format for CPCAT on how to use the allocated keys on the keyboard and touchpad as well as steps that candidates had to follow in order to complete the test during unproctored testing. The

instructions booklet and steps were reviewed after the pilot study and it was decided to include screen shot pictures of the steps during log on. Some additional amendments to the booklet of instructions were made based on feedback from the participants of the pilot study.

The total group consisted of (N=82) for LPCAT and (N= 81) for CPCAT. Procedures included a proctored setting and a simulation of an unproctored setting. The supervised mode, more specifically the managed mode which is often used in test centres and implies a high level of supervision of test administration (Joubert & Kriek, 2009), was used during the proctored test session. For the LPCAT, the controlled mode was used to represent the unproctored mode of test administration. This meant that during the unproctored session the test venue and times of access to the venue were made available to known test takers and the test takers were assisted with the log on process, however during testing there was no supervision (Joubert &Kriek, 2009). It is mentioned in the ITC guidelines point 15 that for internet testing clear information to the test taker should be provided related to test log on and logging off from the system (ITC, 2005). For the CPCAT the controlled mode was used during the same unproctored assessment session, whereby during the same test session after LPCAT participants were required to enter into the CPCAT test without supervision by following the instructions in the booklet that had been provided.

Potential limitations of the research setting are presented below.

- Assessments took place within the HR offices in the training room. Whilst there was no administrator present during unproctored testing, the presence of HR personnel outside the training room was noticeable which could have impacted on the extent to which participants viewed the assessments as being unsupervised. Nevertheless there was no supervisor present in the testing room during the unproctored session unless a problem situation occurred. During unproctored mode some individuals called the researcher when the LPCAT had been completed to help save the test results; however, the researcher was not present during testing. There were several occasions when the internet connection was lost when participants were completing

the CPCAT. The researcher then had to re-set the router to regain internet connection. However in most cases the CPCAT resumed on the same screen where it was prior to loss of the internet signal. In general the support provided during unproctored testing was only related to troubleshooting and was systems orientated.

- The assessments were arranged to be conducted during the off-season when occupancy in the hotel was lower so as to cause minimum disruption to the business hours employees would spend away from work areas. Many new managers were appointed during this time and as a result of ad hoc events, conferences and annual leave taken not all participants completed both test sessions. This also meant that the number of the total group participants who completed the LPCAT in both proctored and unproctored mode differed from the number of participants who completed CPCAT in both modes of administration.
- Furthermore several challenges occurred during the time of assessment. Unfortunately, due to a protected strike and with dates being altered from union members' side, when the study commenced, the researcher was not able to randomly pre-assign individuals based on predetermined characteristics to the unproctored or proctored groups. Due to the fact that two thirds of the organisations employees were union members and mostly at A1 to B3 Patterson grading levels, the researcher had to commence with the study based on available and willing employees, as the duration of the strike could not be predicted. This in effect meant that the study did not adhere to requirements of a quasi-experimental design as originally planned and therefore resulted in a major limitation for the current study.
- The process by which managers decided to encourage employees to participate in the study was not controlled or monitored thereby ultimately further compromising the intent of a quasi-experimental design.

1.5.5 Hypotheses

The specific empirical aims were to test the following non-directional hypotheses:

$H1_0$: There are no statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for cognitive (LPCAT) results.

$H1_1$: There are statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for cognitive (LPCAT) results.

$H2_0$: There are no statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results.

$H2_1$: There are statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results.

The researcher was also interested to find out whether the sequence of testing had an effect on the test results for cognitive (LPCAT) and non-cognitive (CPCAT) results. Additional hypotheses stated were:

$H3_0$: There are no statistically significant differences between the mean scores of the first and second test sessions for cognitive (LPCAT) results.

$H3_1$: There are statistically significant differences between the mean scores of the first and second test sessions for cognitive (LPCAT) results.

$H4_0$: There are no statistically significant differences between the mean scores of the first and second test sessions for non-cognitive (CPCAT) results.

$H4_1$: There are statistically significant differences between the mean scores of the first and second test sessions for non-cognitive (CPCAT) results.

The researcher also obtained data on the biographical information, education levels and computer literacy of the sample group. These were, however, not used in formal hypotheses but merely to describe the sample.

1.6 RESEARCH METHOD

With the aim of providing empirical research, the results of two South African computerised adaptive tests were obtained in situations where the tests setting was altered by two modes of test administration (proctored and unproctored respectively). A quantitative quasi-experimental research design was planned. Terre Blanche et al. (2006) explained that experimental and quasi-experimental research involves the attempt to compare two or more groups of research participants on one or more variables, after the application of some type of intervention. The repeated measures (within-participants design) was used to collect primary data from the same sample group (N=82 and N=81). Due to various factors – described later – the realised samples with data for participants available for both the proctored and unproctored test administration, were slightly smaller for LPCAT (N=82) and CPCAT (N=81) results. Christensen (2001) explained that the within-participants design is used when participants participated in all treatment groups. Also, Whitely (2000) explained the within-participants design or (repeated measures design) where each participant took part in the experimental and control conditions.

1.6.1 Sample size

A convenience sample (N=82 and N=81) was used and the requirement of having data available for the same group for both the proctored and unproctored administrations leading to sample sizes of N=82 for LPCAT and N=81 for CPCAT. The sample group consisted of employees working in the tourism industry at a specific hotel and golf resort. Convenience sampling implied that the sample could not be considered representative – not having been randomly selected or randomly allocated to groups in terms of the sequence of test administration – and the level of power would be impacted on by the sample size (Tredoux & Durrheim, 2002).

Cohen (1992) suggested that four variables are involved with statistical inference namely significance criterion, power, sample size and effect size. In order to determine the effect size (d) for dependent t-tests Cohen's statistical method will be used to interpret the significance of the results. Comparing the mean scores and standard

deviations of the two groups for each test for mode of administration and sequence effect the mean difference will be calculated. Two groups are distinguished based on the mode of test administration that they first encountered – since all individuals included in the final sample groups for data analyses had completed testing in both proctored and unproctored mode of administration. This split was incorporated to allow for the possibility of sequence effect to be evaluated. The "unproctored" group first encountered the unproctored test session and in the second test session completed the proctored session. The "proctored" group first encountered the proctored test session and in the second test session the unproctored session. The descriptors UP for the "unproctored" group and PU for the "proctored" group are used to distinguish the "proctored" and "unproctored" groups – indicating which mode of administration was administered first – from the proctored and unproctored test sessions. The sub groups for LPCAT consisted of "unproctored" (UP) group (n=38) and the "proctored" (PU) group of (n=44). For the CPCAT, the sub groups were the "unproctored" (UP) sub group (n=36) and the "proctored" (PU) group (n=45). The sub-sample sizes were determined by the number of individuals in the total group for whom two sets of scores on the particular measure were available – as well as the sequence in which they had completed the two modes of test administration.

For LPCAT the total group biographical variables included participants that were Afrikaans speaking (61.0%), (31.7%) were Xhosa speaking and (4.9%) English speaking. In addition (57.3%) males participated and (42.7%) females. Employees within the A1 to B3 Patterson grading bracket as well as students completing learnerships (6.1%) across various departments were included in the study. Participants at a higher level than B3 (11.0%) participated as development opportunities were identified during time of testing. This meant that staff at lower levels up to junior management or supervisory level participated; however, heads of departments, senior management and executive management were not included. It was assumed that the formal education of senior and executive management was at higher levels and therefore less important to the organisation to identify their potential NQF levels related to the workplace skills plan. The decision to include employees at A1 to B3 grading

level meant that (6.1%) of participants had Grade 1 – 7 and (39.0%) had between Grade 8 and 11 as formal education.

For CPCAT the total group biographical variables included participants that were Afrikaans speaking (59.3%), (34.6%) Xhosa speaking and (3.7%) English speaking. In terms of gender (58%) males participated and (42.0%) females. Majority of participants were between A1 – B3 Patterson graining bracket with (6.2%) learnership participants and (9.9%) participants at a grading bracket higher than B3.

The organisation was interested in identifying potential for future development not only related to the specific job descriptions but also for those at a scholastic level where they had not completed Grade 12.

1.6.2 Measuring instruments

1.6.2.1 The Learning Potential Computerised Adaptive Test (LPCAT)

Relevant to the cognitive domain Murphy and Maree (2009) explained static intelligence tests as an ill-defined notion of intelligence as the culmination of environmental, socio-cultural or community and family concerns were not considered. Standard cognitive tests in general measure prior learning and not fluid ability to learn new skills, and neither do they allow for aspects such as the product of education, life experience or dynamic aspects of intelligence (De Beer, 2005; Gilmore, 2008). Binet, Vygotsky and Feuerstein are considered to be the founders of dynamic assessment and the Vygotskian theory specifically was applicable to diverse populations (Murphy & Maree, 2009). Kozulin, Gindis, Ageyev and Miller (2003) explained Vygotsky's zone of proximal development (ZPD) as one concept with three different contexts, namely the developmental context of emerging psychological functions in the child, the applied context in assessment and classroom learning, and the metaphoric space where the child's everyday concepts meets with scientific concepts learnt when provided by teachers. Based on Vygotsky's theory of the ZPD, De Beer (2005, p. 720) described the ZPD as "the difference between the level of achievement without help (actual developmental level) and the level of achievement with help (potential developmental

level)"whereas ability was viewed as an acquired skill on demand, potential was based on what could be (De Beer, 2005).

The LPCAT was designed specifically with the South African context in mind (De Beer, 2005). South Africa, with its multi-cultural background and 11 official languages, was likely to benefit from the measurement of learning potential in the cognitive domain (De Beer, 2012). De Beer (2012) indicated that the concept of assessing for learning potential is applicable to the South African legislation (Employment Equity Act No. 55 of 1998). The LPCAT is a cognitive power test that measures learning potential and general fluid ability domain by means of non-verbal figural, general reasoning ability (De Beer, 2012). The instrument uses the test-train-retest approach in which the pre-test score represents current levels of performance and the post-test score represents the projected future or potential level of performance (De Beer, 2005). Because the LPCAT measures learning potential over a wide range of ability levels; the improvement score alone should not be used as a measure of potential, but rather in combination with the present (pre-test) level of performance (De Beer, 2005). The LPCAT is a computer-based test and uses computerised adaptive testing (CAT) based on the Item Response Theory (IRT) in its administration. The concept of IRT is not new: Cianciolo and Sternberg (2004) referred to the history of the measurement of intelligence and explained that in the early 1900's Binet and Simon administered tests to children in order to identify which children would not have the mental capability to benefit from standard educational practices. Recognising the need for test item presentation to be adaptable, Binet and Simon (1916) assessed children individually by arranging a series of questions in increasing order of difficulty (Cianciolo & Sternberg, 2004). In later years Eggens and Straetmans (1996) explained that the purpose of CAT is to aim for efficient estimation of the individual's ability where IRT uses a calibrated item bank that controls the start, continuation and termination of a CAT. De Beer (2000) explained that in CAT, the selection of each consecutive item is based on the responses of the examinee to the previous item and the estimated level of ability at the point in time. De Beer (2005, 2010) indicated that psychometric features of dynamic assessment can improve if the IRT and CAT procedures are used.

What makes the LPCAT suitable for the South African context is that it consists of non-verbal figural items and having excluded language and scholastic content, the test is considered fair with regards to various sub-groups based on culture, language and the level of formal education – sub groups for which bias analysis was conducted during the development of the measure (De Beer, 2005). Test instructions are available in all 11 official languages; however the on-screen instructions are available only in English and Afrikaans (De Beer, 2000).

The coefficient alpha internal reliability consistency scores of the LPCAT range from 0.925 to 0.987 for different groups (De Beer, 2005). Gilmore (2008) presented results that confirmed the internal consistency reliability of LPCAT and construct validity of the LPCAT and in terms of predictive validity found positive correlations that LPCAT is a good predictor (r=0.66) for job performance.

When testing in groups, all participants should complete the same version of the test (De Beer, 2012). Due to the nature of this study the on-screen standardised instructions version of LPCAT was chosen which meant that candidates could read standardised instructions from the screen, thereby accommodating the unproctored mode. The LPCAT is not as yet available online; so for the purpose of this research, it was administered in a simulated online (unproctored) manner. In using the controlled mode the test administrator helped known test takers to log on but provided no direct help during the unproctored testing session (ITC,2005).

1.6.2.2 The Career Preference Computerised Adaptive Test (CPCAT)

The CPCAT, a newly developed test that measures career preferences and development was reaching its final phases during the time that the current empirical research was being conducted. The CPCAT was designed according to a three dimensional model which measures 16 career fields, 12 career activities and six career environments which allows for measurement of interest in these categories but also interest across pertinent dimensions.

De Beer (2011) explained that the CPCAT assesses the career preference of test takers which can be used for vocational guidance or career related decision-making such as screening and selection or training and development. In the first two stages of the test 68 questions are presented to the test taker with two rounds of questions covering all 34 sub-dimensions in the three broad dimensions. Test takers indicate their level of interest on a scale from 0 to 100 by rating their preferences for each statement. Subsequent to this first round of ratings, the sub-dimensions rated lowest in each of the broader dimensions are discarded. In stages three and four of the test the remaining top six fields, top six activities and top four environments - based on the ratings of the first two rounds – are then presented using more detailed questions upon which the test takers again rate each statement form 0-100. Note that while the CPCAT is computerised and to some degree adaptive, it is not based on IRT principles of CAT. A bar chart represents the individual's top 16 preferences in descending order mixed across the top six field preferences, the top six activity preferences and the top four environment preferences.

The coefficient alpha reliability results for CPCAT sub-dimensions reported by De Beer (2011) showed average coefficient alpha internal consistency reliability values of for these three dimensions 0.858 (fields), 0.818 (activities) and 0.808 (environments). Also 27 out of 34 sub-dimensions met the 0.80 with the other seven dimensions meeting the 0.70 level. De Beer (2011) indicated that further research for students and working individuals would be beneficial.

Due to the fact that the CPCAT was still in the final development phase, this study could also add additional information on the reliability and validity of the test with the same sample group and the possible effects of the online mode of administration on CPCAT results.

1.6.3 Statistical analysis

In the present study the effects of the mode of test administration on the test results were investigated for a cognitive (LPCAT) and non-cognitive (CPCAT) test respectively. Due to the small sample size the normality of distribution of the variables concerned

was checked by means of the Kolmogorov-Smirnov test on SPSS (Field, 2005). None of the LPCAT or CPCAT variables differed significantly from the normal distribution (Field, 2005) – even for the sub-groups - so the dependent t-test was used to compare the mean scores of the same individuals obtained with proctored and unproctored test administration respectively. Descriptive statistics provided information about variables, mean scores, variance and range (Terre Blanche et al., 2006). Participants completed biographical forms from which frequencies, minimum, maximum and mean scores, as well as standard deviation, were used to summarise the biographical data. In addition participants were asked to rate their computer competency and average number of hours of work on computers per week. It has been reported that levels of computer familiarity should not be ignored when assessing test performance (Davies, Foxcroft, Griessel & Tredoux, 2009; Joubert & Kriek, 2009).

Green and Salkind (2008) suggested that the dependent t-test could be used to evaluate the difference between repeated tests or the within-subjects design. Although the sample size for LPCAT (N=82) and for CPCAT (N=81) was small, Kolmogorov-Smirnov analyses (Fields, 2005) indicated that the data for the total group as well as for the sub-groups was normally distributed.

Field (2005) explained that the differences in two conditions can be compared in this manner. The “unproctored” (UP) sub-group and the “proctored” (PU) sub-group results were separately analysed and mean LPCAT pre-tests and post-test scores and mean CPCAT scores of both modes of administration were compared respectively for the two sub-groups separately as well as for the total group.

1.7 ETHICAL CONSIDERATIONS

The steps that were taken to ensure that the research was conducted in an ethical manner are listed below.

1) Consent for the study was obtained from the company involved prior to informing any managers or staff about the study.

2) Ethical clearance was obtained from the university ethics committee of the department and the college within which the department falls.
3) Informed consent was obtained from all the participants to be tested twice on the same measures.
4) Participants were informed that participation was voluntary and that they could decline to participate or withdraw at any time. The right to decline to participate was explained, as well as the right to withdraw from the study without any consequence to the individual involved. One person declined to participate in the study and eight people withdrew before completing the second testing due to various reasons.
5) Participants were informed about the South African legislation and Employment Equity Act no of 1998 regarding psychometric testing and the need for research in the use of assessments in the South African context.
6) Written agreements with the organisation were undertaken that the data of the research was to be treated with confidentiality on all levels.
7) The purpose for which test results were to be used was limited to training or developmental purposes.
8) Where participants scored less on one of the two test results of the LPCAT, only the best test results were reported on and these were the scores to be used for development.
9) In terms of testing individuals with the purpose for training and development related opportunities the open mode which entails no human contact with the candidates during testing was excluded but rather controlled mode used so as to uphold ethical practice related to the test environment.

1.8 CHAPTER LAYOUT

The structure of this study is outlined below.

Chapter 1 Scientific orientation of the research

Chapter 1 focuses on the motivation for the research, defining the problem statement and providing background information as well as relevant methodology.

Chapter 2 Literature Review

Chapter 2 provides a literature review relevant to the topic of psychometric testing, the mode of test administration, computerised testing and unproctored or proctored studies conducted in the past.

Chapter 3 Article

The article includes information on the background and the literature related to the topic as well as a presentation of results and findings.

Chapter 4 Conclusions, limitations and recommendations

Conclusions are discussed based on the findings of the research and the important limitations and future recommendations are mentioned.

1.9 CHAPTER SUMMARY

Chapter 1 was aimed at providing an overview of this research study, clarifying the purpose and objectives of the research. Chapter 2 will be focused on previous research in the field and aimed at providing relevant information regarding psychometric testing in the 21st century technological world with specific focus on modes of test administration, and computerised testing as well as internet-based testing internationally and in South Africa.

CHAPTER 2

LITERATURE REVIEW

The literature review is aimed at discussing past research that is relevant to the topic of mode of test administration and psychological assessment results with reference to UIT. Concepts relevant to this study such as proctored and unproctored modes of administration, computer-based testing, computerised adaptive testing, internet testing as well as the advantages and disadvantages of computer and internet based testing are explored.

2.1 INTRODUCTION

South African legislation as discussed in the previous chapter, par 1.2, requires that test instruments be researched when used in different modes of test administration, especially in different modes than what a test was designed for (Foxcroft & Davies, 2006). Tippins (2009, p. 7) made the following point: "If one relies on reliability and validity evidence of a test administered under proctored conditions, the psychologist cannot accurately describe the reliability and validity of the inferences made under unproctored conditions." In addition Tippins (2009) explained that one fundamental principle of good testing practice was to provide a candidate with a testing environment that could facilitate optimal performance. Unproctored testing, and specifically unproctored internet testing, could host testing conditions or environments that could limit or facilitate optimal performance (Tippins, 2009). Unique challenges such as internet connectivity or hardware and software problems of computer programs exist (Naglieri et al., 2004). In the literature review that follows previous arguments and empirical research results regarding the mode of test administration of computer-based and internet-based tests will be presented.

2.2 TECHNOLOGICAL DEVELOPMENTS AND PSYCHOLOGICAL TESTING

Bartram (2000) referred to the true beginning of the internet as having been in 1995 with a rapid growth in the use of the internet that brought about certain changes in society and with regard to computerised and internet-based psychological assessment. According to Eid and Diener (2006), computers and specifically the internet progressed greatly between 1995 and the 21st century. Furthermore in terms of psychometric testing Bartram (2000) referred to the impact of internet usage on computer-based tests as becoming the 'default' medium, with paper-and-pencil tests becoming the lesser preferred medium over time. Tippins et al. (2006) mentioned that as the internet became more accessible, the use of computerised testing increased. Foster (2010) referred to the prevalence of technology-based testing including computer-based testing, electronic testing, digital testing and online testing.

Salgado and Moscoso (2003) pointed out that many personnel selection procedures including tests and questionnaires were transformed into internet versions of the same measures. According to Naglieri et al. (2004), as Industrial and Organisational Psychologists used testing in many interventions, personnel selection is a broad field that related to psychological testing. Beaty et al. (2009) conducted surveys that indicated that more than two thirds of employers conducting testing for selection engaged in unproctored testing. Furthermore Beaty et al. (2009) reported a widespread acceptance of UIT and a need for improving the use thereof rather than debating UIT. The convenience that internet testing provides to organisations has resulted in increased online test usage during recruitment and similar processes; yet practical as well as ethical challenges for its correct use in practice still need attention (Davies et al., 2009; Tippins, 2009).

Coyne and Bartram (2006) reported that the past few years have seen major developments in stand-alone computer-based and internet-based testing. In addition Bartram (2000) predicted that computer-based assessment and online assessment could within time replace paper tests. Even though Bartram's prediction is not as yet entirely a reality for all psychometric tests worldwide, the continuous growth in the field

of computer-based and internet-based testing has resulted in discussions of how UIT can be used most effectively (Beaty et al., 2009; Lievens & Burke, 2011; Tippins, 2009). As a result encouraging many test providers to improve or modernise tests to keep up to date with the information technology and fast pace work environment.

The ITC guidelines for Quality Control in Scoring, Test Analysis and Reporting of Test Scores (ITC, 2011) identified the need for practitioners to broaden their knowledge of quality control practices and referred to quality control procedures regarding systematic processes for all stages, from test scoring to test analysis as well as reporting. Bartram (2006) explained that a move from the client-side to the server-side for control was to change the nature of the relationships, involving the test taker, test user, test distributor and test producer. Bartram (2006) referred to the pressure that HR professionals face in the hiring process and indicated that there was a need for a revival of psychologists as specialists needed within the assessment processes to uphold quality control procedures.

2.3 KEY CONCEPTS DEFINED

Relevant to this study the following key concepts will be defined: administration mode, computer-based testing, computerised adaptive testing and dynamic assessment, internet-based testing, learning potential and career preference.

2.3.1 Administration mode

The level of control during administration of psychological testing was particularly relevant to this study. Proctored modes of administration were defined as those times when testing was managed or supervised through direct human supervision during the test administration session (ITC, 2005). The unproctored mode of administration, on the other hand, was when tests were completed in open or controlled format and there was no human supervision present during the test administration session (ITC, 2005).

The following four modes of test administration have been identified:

1) *Open mode* has been defined as when anyone can access and complete a test without supervision. With open mode related to unproctored testing, the test taker has direct access to test materials with no involvement on the part of the test administrator (ITC, 2005). The issue at stake here is authenticity.
2) *Controlled mode*, related to unproctored testing, has been defined as when no direct supervision was provided but in order to access the test a logon code is required which is made available to a particular individual. The ITC guidelines describe this mode as control being exercised over who can access tests on the internet and how often (ITC, 2005).
3) *Supervised mode* has been defined as when a certain degree of supervision is present as the administrator deals with the logon process and verification of the test taker's identification. The mode involves face-to-face involvement with the test administrator but the test distributor has no means of directly controlling the location and equipment being used (ITC, 2005)
4) *Managed mode* has been defined as when high levels of supervision and control are achieved (Coyne & Bartram, 2006; ITC 2005). This mode infers that there is direct supervision and direct control over the equipment being used (ITC, 2005).

Tippins et al. (2006) defined unproctored testing as being when the selection instrument is made available to the candidate via the internet or computer and the testing event is not being monitored, thereby resulting in test takers not being identified and behaviour not observed or supervised. Furthermore Tippins (2009) referred to UIT as when a candidate completes an internet-based test without the traditional human proctor present but this could include cameras for observation or follow-up testing for verification of initial results.

Questions raised around UIT specifically focus on the degree of standardisation of the testing conditions that would be unknown during unproctored testing which in turn has relevance for reliability, validity and norms (Tippins, 2009). Do (2009) mentioned that it was important that the testing conditions should not cause bias such as measurement in equivalence or the possible presence of some effect across groups. Do (2009) also mentioned that, if little evidence of Differential Item Functioning (DIF) or Differential Test

Functioning (DTF) is found, there should be fewer concerns that the test scores are likely to be influenced by administration mode.

Where psychometric tests can be used for screening individuals during the selection process, the comparability of psychometric properties is essential to avoid wrongfully excluding test takers from applicant pools.

2.3.2 Computer-based testing

In contrast to static paper-and-pencil administration "computer based testing is when instructions appear on the computer screen and the computer prompts the client to answer a series of questions" (Gregory 2007, p. 580). Tippins et al. (2006) defined computer based testing as when instruments are presented to candidates via a computer. This could be done by pressing allocated keys on the keyboard or by using the mouse or touchpad to select the answer. Gregory (2007) discussed computer-assisted psychological assessment (CAPA) which was a term that referred to the entire range of computer applications in psychological assessment. CAPA includes computer-based testing, scoring and interpreting of results.

Foxcroft and Davies (2006) explained that computer-based testing enhanced the efficiency of testing and electronic test distribution. In addition computer-based testing allows for immediate scoring and reporting of results (Foster, 2010).

2.3.3 Computerised adaptive testing (CAT)

The origin of computerised adaptive tests dates back to as early as 1905 (De Beer, 2005). Binet and Simon measured intelligence, studied individuals' performance and attempted to improve future performance with the involvement of relevant interventions and mental exercises to assist individuals to show their optimal performance level (Cianciolo & Sternberg, 2004; De Beer, 2006; Murphy, 2002; Sternberg & Kaufman, 2011). According to Davies et al. (2009), adaptive testing individualises a test by adapting the level of difficulty of the items presented depending on the individual's response. Weiss (1982) described adaptive testing as various methods that permit

measurement of equal precision throughout the range of the trait being measured whilst maintaining high levels of efficiency.

Davies, Foxcroft, Griessel and Tredoux (2005) explained the method of computerised adaptive testing as follows: the test taker's correct answers will lead to more difficult questions while incorrect answers will lead to easier questions. This method matches the difficulty level of the items presented to knowledge or ability level of the test taker through the selection of items (Davies et al., 2005). The LPCAT is a South African test which utilises this technology of CAT (Van Eeden & De Beer, 2009). The CAT technique is particularly appropriate to ensure adequate measurement properties for dynamic psychological assessment (De Beer, 2010; Murphy, 2002).

2.3.4 Dynamic testing

Murphy and Maree (2009) explained that the philosophy of dynamic assessment relates to those changes individuals experience while developing expertise. In contrast to the static paper-and-pencil test methods, the approach called dynamic assessment, which involves a test-train-retest approach, was developed and is specifically relevant to the measurement of learning potential (De Beer, 2010; Murphy & Maree, 2009). Whilst the history of dynamic assessment tools dates back to the early twentieth century, mostly in the intelligence research community, dynamic assessment, both globally and in South Africa is receiving more attention from practitioners and educators (Murphy, 2002). Dynamic testing procedures are assessment procedures which include some form of learning experience as part of the testing. "In dynamic tests, what is tested is not merely previously acquired knowledge, but also the capacity to master, apply and reapply knowledge taught in the dynamic testing situation" (De Beer 2006, p. 9). Dynamic assessment is in general believed to improve culture fairness because most of these measures make use of non-verbal figural material and are aimed at measuring learning potential. The approach is based on Vygotsky's concept of the zone of proximal development (ZPD) (Murphy &Maree, 2009). Kozulin et al. (2003) explained that the centre of Vygotsky's theory implies that human cognition and learning are social and

cultural rather than individual phenomena. According to Kozulin et al. (2003), the current practice and application of the ZPD lies in dynamic assessment of learning potential.

2.3.5 Internet testing

Tippins et al. (2006) indicated that the moment computers were linked to the internet, web-based tests were going to be used. Also, whereas computer-based testing referred to mode of delivery, the internet indicated the source of the test content. Bartram (2006) described the internationalisation of tests as being when individuals and their countries are no longer closed mediums. This implies that as a result of the internet, tests from other countries could be used for assessment in South Africa (Joubert & Kriek, 2009). The internet has created the need for practitioners' worldwide to share a common understanding of best practices and standards regarding internet assessment (Foxcroft & Davies, 2006).

2.4 SOUTH AFRICAN LEGISLATION AND INTERNATIONAL GUIDELINES

With reference to the literature related to proctored and unproctored testing, a clear distinction between South African legislation and international rules and regulations was essential for this literature review. The reason for this is that South Africa is one of very few countries that were not only bound by professional boards for scope of practice, but also by strict legislation and statutory control (Foxcroft & Davies, 2006). In a country that has been crippled for many years by apartheid and inequality, the African National Congress (ANC) shifted the focus of the industry and education to redress the imbalances of the past (Foxcroft, Roodt & Abrahams, 2009; Murphy, 2002). Therefore, due to the potentially discriminatory nature of psychometric testing, the application of strict rules around the use and fairness of psychometric tests is governed by the Health Professions Council of South Africa (HPCSA) (Foxcroft et al., 2005).

The Employment Equity Act No. 55 of 1998 (Section 8) referred specifically to psychological tests and assessments as follows:

> *Psychological testing and other similar forms or assessments of an employee are prohibited unless the test or assessment being used has been scientifically shown to be valid and reliable; can be applied fairly to all employees; is not biased against any employee or group.*(Employment Equity Act 55 of 1998, p. 16)

The Health Professions Council of South Africa (HPCSA) was mandated to protect the public and to guide the profession of psychology (Foxcroft & Davies, 2006; Foxcroft et al., 2005). In addition the Health Professions Act 56 of 1974 restricts the use of psychological tests to registered psychology professionals. Those professionals are required to have undergone the necessary training and are only allowed to use psychological tests that have been classified by the Professional Board for Psychology (Foxcroft & Davies, 2006). It is therefore very important for South African registered professionals to understand that, whilst adopting ITC guidelines relating to computer-based testing and internet-delivered testing for the South-African context, the guidelines should be used in combination with South African legislation.

Foxcroft and Davies (2006) explained that countries differ widely in terms of legislation and how tests are controlled. It was highlighted that practitioners have to consider the specific country's legislation in which the test was published and in which it will be used.

2.5 TEST ADMINISTRATION

Why has the mode of test administration received so much attention in the past few years? Various studies have been conducted on the consistency of test results with diverse modes of test administration, yielding the following different end results: Lee, Moreno and Sympson (1986) investigated whether mean scores on a computerised version of an arithmetic reasoning test namely the Arithmetic Reasoning Subtest of the Armed Services Vocational Aptitude Battery (ASVAB-AR) and Experimental Arithmetic Reasoning Test (EXP-AR) were lower than those of the paper-and-pencil version. Participants were randomly assigned to either paper-and-pencil or computer mode of

administration. Results indicated that mode of test administration had a statistically significant effect on arithmetic reasoning test scores with the mean score obtained by computer was lower than that obtained by paper-and-pencil and that item difficulty was affected by mode. A possible explanation was given in terms of anxiety levels during computer testing. In more recent years Bartram and Brown (2004) found comparable results on the Occupational Personality Questionnaire (OPQ32i) and established that lack of supervision in high stakes situations had very little effect on test scores, however it should be kept in mind that non-cognitive tests could be more acceptable in UIT settings (Tippins, 2009). With regard to paper-and-pencil and internet-based testing, Salgado and Moscoso (2003, p. 200) indicated that:

> *previous research compared the similarity of the responses of the two versions using independent groups of examinees. This means that actual equivalence was not directly examined because the failure to detect differences between the groups does not mean that the two versions are really equivalents...*

In an attempt to address this shortcoming in previous studies, Salgado and Moscoso (2003) had 162 undergraduates complete the Personality Inventory of Five Factors (IP/5F) in both paper-and-pencil and internet form. In order to control the effects of the presentation some participants took the paper-and-pencil version first whilst others took the internet version first. All sessions were supervised. Results indicated that the mean scores were equivalent with a slightly larger standard deviation in the internet-based version. By using the same participants in their study Salgado and Moscoso (2003) provided strong evidence for the equivalence of mean scores.

Naglieri et al. (2004) highlighted the importance of preserving psychometric properties such as test reliability and validity during computer-based and internet-based testing. Do (2009) also suggested that the focus of comparisons between UIT and comparable proctored testing should be on psychometric properties. Joubert and Kriek (2009) reported that for the (OPQ32i) the construct equivalence on two different modes of administration was “strikingly similar”.

With reference to traditional test administration, Griessel, Jansen and Stroud (2009) discussed the practitioner's duties when tests are administered, which includes the relationship between test taker and assessment practitioner. Also, the administrator's duties include exercising control over the assessment environment, more specifically over group testing sessions (Griessel et al., 2009). Assessment is not a mechanical process but a psychological intervention, and the rapport that a practitioner has to establish with a test taker is considered important and encouraged (Griessel et al., 2009). The practitioner could offer motivation, empathy and solutions to emotional and environmental concerns during testing (Tippins et al., 2006). Griessel et al. (2009) referred to the importance of the relationship between the practitioner and the test taker to establish a rapport and motivate the test taker. In essence during the administration, the test duties of the administrator included dealing with anxiety, providing assessment instructions, adhering to time limits, managing irregularities and recording behaviour (Griessel et al., 2009). The question could perhaps be asked to what extent could 24/7 call centres and web-based support be possible in order to reduce the isolation experienced in UIT settings? Nonetheless experimenter effects could occur with the human factor present during test administration and should not be ruled out, whereas computerised testing allows for consistency of instructions (Tippins et al., 2006).

With the above in mind, the traditional method of administering tests has evolved into a potentially new era for psychometric testing yet it has also brought about some new opportunities as well as concerns and serious questions. Although internet-based testing has inherent limitations such as lack of both control and observation of conditions, many advantages exist over laboratory settings such as the ease with which large numbers of participants can be tested and such data researched (Eid &Diener, 2006).

Kaminski and Hemingway (2009) noted that internet testing provides quick turnaround time to fulfil the hard reality of business needs. Administrative and overall cost saving as well as selection platforms may drive UIT (Reynolds, Wasko, Sinar, Raymark & Jones, 2009). However, the broader framework of deployment conditions, hiring

contexts, selection systems, administration conditions and risk strategies should be taken into account as part of the decision to use UIT (Reynolds et al., 2009).

Test administration variations that could be present during UIT include anxiety levels, stress, illness, and restlessness as well as environment setting distractions such as noises, phones ringing or bad lighting which could all impact on the test administration process (Griessel et al., 2009). One major issue with the unproctored mode of test administration is the degree of, or lack of control over, the test situation (Tippins, 2009). Whilst in unproctored mode problems such as cheating could occur, the proctored mode in essence minimises such problems (Beaty et al., 2009).

Tippins et al. (2006) and Tippins (2009) reported that a panel of experts, with different perspectives on UIT, could not reach consensus about the ethics around UIT. Whilst some panellists embraced the benefits of UIT with open arms, others became wary and even rejected the UIT mode of test administration (Tippins et al., 2006). It would seem that the provision of sound research on test results for cognitive and non-cognitive measures in different modes of administration is a vital step towards defending test validity and reliability for different modes of testing and decisions as to when UIT may be used, especially because information technology develops so fast. Consequently the practical implementations of secure infrastructure must be considered with the relevant role players such as the developers, publishers and users to apply standardised practices and protect the use of tests (Bartram, 2006). Tippins et al. (2006) made an important observation in that the practitioner must contemplate the nature of the test (cognitive or non-cognitive) and the use of the test (selection or development) and research evidence available about the specific measures before deciding on the mode of administration. Davies et al. (2009) raised concerns regarding a trend that emerged from computer and internet-based testing where test distributors could deliver tests directly to the test taker thereby excluding the specialists. The issue includes “confusion regarding the roles and responsibilities of people involved in the assessment process and what knowledge, qualifications and expertise they require” (Foxcroft et al., 2009, p.17). As a result some countries distinguish between psychological assessment and competency-based assessment. It is indicated by

Foxcroft et al. (2005) that roles and responsibilities should be clearly defined in relation to legislation.

2.6 COMPUTER-BASED TESTING

Whilst the following discussion is not aimed at arguing the relevance of computer-based testing in the 21st century world of work, a few important advantages and disadvantages are discussed. The distinctions between computer-based and internet-based tests have to be made because, even though closely connected, the two concepts proved to be different.

2.6.1 Guidelines for computer-based testing

In the ITC guidelines (ITC, 2005) it has been suggested that test users must understand technological support documentation. Test developers are required to consider the technological, psychological and administrative challenges involved in computer testing.

According to Drummond (2004), the main objectives to be considered in computer-based tests are:

1) basic computer literacy;
2) knowledge of information sources;
3) an objective and evaluative attitude toward computer-based testing;
4) understanding of the individual's rights to privacy;
5) knowledge of and experience with computer-assisted testing; and
6) knowledge of the required computer hardware and software needed.

The above suggests a certain level of understanding from the administrator's side towards test takers' familiarity with computers and attitude towards computer testing. Also the administrator would need some background information and knowledge about the specific computer programs that are used. During computer assessment possible computer anxiety could be present, highlighting the role of interaction between the

administrator and the candidate, as well as the importance of empathy with regard to the stressors involved in the assessment process.

2.6.2 Advantages of computer-based testing

There are many advantages to computer-based testing (Davies et al., 2009). The administration procedures of computerised testing allow for computer-based tests being more enjoyable than paper and pencil tests (Davies et al., 2009; Foxcroft, Paterson, Le Roux & Herbst, 2004). Computerised tests take less time to complete in comparison to paper-and-pencil tests. Standardisation of instructions is achieved, fewer assessment practitioners are needed and errors from inaccurate scoring can be limited (Davies et al., 2009). Tippins et al. (2006) explained the advantages of computerised testing as enhancing consistency of delivery and improving efficiency of delivery. Tippins et al. (2006) furthermore explained that computer administered tests provide consistent instructions and accurate timing.

Gregory (2007) stated that in the past scoring of psychological tests by hand was monotonous, time consuming and error-prone but computer-based tests can be scored instantaneously with an effortless process compared to paper-and-pencil scoring. Tippins et al. (2006) also mentioned that computer tests provide quick and precise scoring. Once the test taker has completed the test, the administrator can generate the results instantaneously and automatically. Not only could a computer score results but it could assist with the interpretation and writing of reports with precise recording and storing opportunity (Davies et al., 2009). In addition, Davies et al. (2009) described computer-based test interpretation (CBTI) as being when the computer program, by using a clinical or research approach, alerts the practitioner of test scores relevant to diagnostic judgement, or norm group relevance, which could otherwise have been missed by the practitioner. Gregory (2007) referred to CBTI when services such as scoring reports, descriptive reports, actuarial reports and computer assisted clinical reports can be obtained from the program. Scoring reports allow for scores or profiles which include statistical significance and confidence intervals. Descriptive reports include a brief scale by scale interpretation of results, while actuarial reports make

predictions but also diagnoses of test takers (Gregory, 2007). CBTI could be defined as test interpretation and report writing with advantages such as quick turnaround time, low cost, near perfect reliability and complete objectivity (Gregory, 2007).

In general when computer-based testing is used scoring errors are reduced and standardisation increased (Naglieri et al., 2004). Furthermore, Davies et al. (2009) recognised that the biasing effect of the assessment administrator can also be eliminated and that fewer assessment practitioners are needed during the administration of computerised tests.

2.6.3 Challenges of computer-based testing

Computer tests pose many challenges relating to issues such as technical hardware or software problems (Tippins, 2009), program compatibility, the test takers' unease with computers, privacy issues and generic reports (Davies et al., 2009). Gregory (2007) referred to report writing as one of the major controversies linked to computer-based testing. One controversial issue is the automated and routine practice related to computer report writing. Davies et al. (2009) suggest that practitioners should not attach too much value to computer based reports but instead should complement and substitute reports with information from other sources such as observation. Some other concerns include arguments that computerised testing is a poor substitute for psychological assessment and that computer narrative reports are in general not validated (Gregory, 2007).

Leeson (2009) discussed the challenges related to computer-based testing such as the lack of familiarity with computers, computer anxiety and screen legibility. Whilst computer anxiety was a major concern, Eid and Diener (2006) predicted that the increasing prevalence of computers in society could in time overcome the issues related to computer familiarity. Foxcroft and Davies (2006) stated that consideration should be given to the impact that inequality of access to computers and technology can have on test performance, as many South Africans live in rural areas where electricity and technology are lacking. In addition Davies et al. (2009) mentioned the lack of computer sophistication (literacy) in South Africa. Foxcroft et al. (2004) discussed the

disadvantage that South Africa had with a lack of a wide range of computerised tests being available and they were also concerned about computer familiarity among test takers in South Africa. Moreover, Foxcroft and Davies (2006) highlighted the need for considerations regarding equality of access for all groups particularly in South Africa.

From a legal perspective, Foxcroft and Davies (2006) emphasised the importance for test developers to document evidence of the equivalence between paper-and-pencil and computer-based tests.

2.6.4 Computerised adaptive testing

In CAT the computer typically selects items from a pre-determined item bank to match the difficulty level of the item presented to the candidate's estimated ability levels at the time (De Beer, 2010). The concept of constantly being measured at one's ability level has been executed more accurately and efficiently with the computerised adaptive method. This means that each person is challenged according to his or her own capabilities (De Beer, 2006). Kanjee and Foxcroft (2009) suggested that tailoring the tests is possible due to the Item Response Theory (IRT). Weiss (1982) reported that the application of the IRT consists of three components namely: a) the means of selecting the first item to be administered to the individual; b) a means of scoring the items during the process of administration in order to select the next item; and c) the means for determining the adaptive test based on a subset of items for each individual. With the use of the IRT, computerised adaptive testing is a process by which items are interactively selected to match to the test taker's estimated ability and this counters the floor-ceiling effect in LPCAT (De Beer, 2005, 2010).

De Beer (2006) asserted that a requirement for adaptive tests was that they have to be power tests and not timed tests, which would mean that those candidates tested had to have the opportunity to answer all items presented to them with no fixed time limit for the test. With reference to educational inequalities, the CAT process was seen as more fair and equitable (De Beer, 2006) because examinees were constantly faced with items of a difficulty level commensurate with their estimated performance level at the time. Styles and Andrich (1993) explained that due to the CAT process which is targeted at

each person's level, the presence of unexpected results was reduced. It should be noted that very few computerised adaptive tests have been available in South Africa (Murphy, 2002). However, with its time saving, improved measurement accuracy and appropriateness specifically for dynamic assessment, more psychological tests could make use of the computerised adaptive test administration technique in future. Drummond (2004) explained that reliability of computer adaptive tests was computed by the use of an internal consistency reliability index.

In a step by step process Antal, Erös and Imre (2010) explained CAT steps as follows.

1) The candidate starts from an initial ability level.
2) Selection of the most appropriate test item is based on the present level.
3) Re-estimation of the candidate's ability is based on the candidate's previous answer.
4) The previous two steps are repeated until the termination rules are met and a final level is established.

In general benefits of CAT could include aspects such as test takers being challenged equally, working at their own pace (Gregory, 2007) and testing becomes interesting and positive to the test taker as the test provides questions at an ability level that is consistently appropriate as well as challenging (Drummond, 2004). Whilst scoring is immediate and test security could be improved in computer based testing (Gregory, 2007), another major advantage has been that psychometric precision is increased (Drummond, 2004).

Another advantage is that CAT allows for shorter testing times (De Beer 2010; Gregory, 2007). De Beer (2010) suggested that CAT is particularly appropriate for measurement of learning potential as CAT allows direct comparison between pre-tests and post-tests specifically relevant to learning potential (De Beer, 2006). Furthermore, while variable numbers and different sets of items are administered to the candidate, scores remain comparable (De Beer, 2006).

The computer adaptive technique was regarded as one of the biggest contributions to computerised testing and whilst CAT was used during the 1990s, LPCAT was one of the early South African tests making use of the adaptive manner (Van Eeden & De Beer, 2009). It would seem that the computer adaptive technique is still growing in popularity and can only increase in future (Gregory, 2007). For this reason this study is quite unique, as the LPCAT is a fully adaptive (CAT) test in the cognitive domain, and the CPCAT measures interest through computerised and internet-based testing that is to some degree adaptive.

2.7 INTERNET-BASED TESTING

One of the leading debates in the world of psychometrics in the past few years has been the topic of internet-based testing (Beaty et al., 2011; Hense, Golden & Burnett; 2009; Lievens & Burke, 2011; Reynolds et al., 2009; Tippins, 2009). The benefits that internet testing provides are exciting. "The key advantage that the Internet offered was that test producers and publishers were able to assume the availability and accessibility of ubiquitous infrastructure through which to deliver new products and services" (Bartram, 2006, p. 130).

There are many advantages for organisations in the use of internet assessments, of which the two biggest driving forces seem to be time and cost effectiveness or otherwise stated better, faster and cheaper services (Davies et al., 2009; Naglieri et al., 2004). Internet-based assessment is in line with modern technology, new ways of thinking and social network systems, because the way people conduct business and communicate with others have changed (Foster, 2010; Naglieri et al., 2004). Internet assessment also provides alternatives in terms of assessment methodologies and it is changing models of delivery to the test taker (Bartram, 2006). In addition the results are easily obtained and immediately available, with the added luxury for test takers of being able to be assessed in one's own environment or home, after hours and in a relatively relaxed and known environment (Gregory, 2007). It should be kept in mind that this view is not necessarily applicable for the South African context where less than 10% of

individuals have access to internet at home (StatsSA, 2011). From a business perspective (Hense et al., 2009) found that internet assessment improved hiring efficiency dramatically. They concluded that the rewards of quality and efficiency of the internet outweighed the risks of using UIT as hiring efficiency improves (Hense et al., 2009). However, Naglieri et al. (2004) explained that advantages of internet testing became irrelevant when tests were used in ways that were not supported by the validity and reliability of the instruments.

Whereas computer-based testing in the past was mainly completed in a proctored and supervised environment, the internet facilitated certain modifications to computer assessment, not limited to, but allowing for, the possibility of the unproctored mode (Naglieri et al., 2004). Bartram (2000) and Coyne and Bartram (2006) stated that computer based testing has been delivered for many years with the accredited user purchasing the materials and exercising direct control over the use of the computer-based test. The internet, however, changed this mode of test delivery (Coyne & Bartram, 2006). Beaty et al. (2009) mentioned that survey results have revealed that more than two thirds of organisations conducting psychometric testing for selection have been engaging in UIT.

It is fair to say that the nature of the internet allows for unproctored testing conditions. Tippins (2009) explained that UIT was used to refer to internet-based testing when a candidate completed a test without a human proctor present. Serious concerns about the various impacts that internet-based testing had on the industry of psychology developed. Various perspectives on UIT and unproctored testing have emerged (Beaty et al., 2009, 2011; Pearlman, 2009; Tippins et al., 2006). While some practitioners supported UIT (Tippins et al., 2006) some pursued UIT as an option and some rejected UIT outright (Tippins et al., 2006). With a panel of experts (Tippins et al., 2006) little consensus existed on the topic of UIT and members of the panel were divided on the ethics related to the use of UIT; panellists could not agree as to whether or not UIT was an acceptable practice. The rapid growth of the internet, easy access to any information required and specifically the use of internet-based testing resulted in new challenges related to internet testing. There do not seem to be many immediate

solutions to the questions around internet-based testing and whilst the internet evolve almost daily, research contributions to questions on the topic have trailed behind.

2.7.1 Main concerns about internet-based testing

The debate about internet testing is ongoing. Reynolds et al. (2009) and Beaty et al. (2009) argued that rather than questioning UIT, the focus should be on the implementation and utilisation of internet-based testing in the workplace and finding ways to enhance the progress to better understand UIT. However, Reynolds et al. (2009) investigated the holistic process and other factors related to UIT. Concerns raised are that whilst HR systems continue using UIT, the appropriate guidance from workplace psychologists will trail behind as long as psychologists are for or against UIT (Reynolds et al., 2009). Also according to Gibby, Ispas, Mccloy and Biga (2009), the debate around UIT should move the focus from whether it can be used to how it can be validated, developed and managed. Reynolds et al. (2009, p. 52) furthermore stated that "UIT represents a range of test deployment conditions that vary so widely that they preclude accurate and unqualified statements about internet based testing." This abovementioned quote by Reynolds et al. (2009) provides a valid reason as to why UIT created certain unease with some practitioners, as standardisation procedures became a problem. In addition Kaminski and Hemingway (2009) referred to the hard reality of UIT that Industrial and Organisational Psychologists had to find a balancing act between academic practice and satisfying business needs. The pressure is on practitioners from a business point of view to recruit and select people quickly and efficiently with the most time and cost saving methods. Beaty et al. (2009) claimed that clients' needs for UIT was clear, that UIT was widely accepted and that debates should not be whether UIT is relevant but rather efforts should be focused on how UIT can be improved. On the other hand Tippins et al. (2006) reported that depending on certain contextual scenarios, some experts regard UIT as unacceptable, misguided and inappropriate. The varying aims of tests are to measure objectively domains of functioning such as ability, personality and interest, mostly undertaken with a supervisor present in order to ensure fairness, to standardise procedures and to confirm authenticity of answers by a

particular individual (Foxcroft & Roodt, 2005). Online testing challenges the authenticity of results and the process as a whole.

What are the major issues around UIT? Bartram (2006) stated that the main concerns related to issues of good practice which included adequate control over the management of the assessment process, feedback or reporting of high quality and control of tests delivered on the internet. Whenever a test situation presented itself where there was to be a lack of control, a window could open for problem situations and test credibility to be questioned, as well as non-standardised conditions affecting test results (Tippins et al., 2006). From the perspective of a test provider and distributor every test to be used online, should be researched for test reliability and validity and regularly updated (Foxcroft et al., 2004). Naglieri et al. (2004) and Joubert and Kriek (2009) concluded that with regards to psychometric standards, test reliability and validity still applied, even though tests were possibly used in a different way than had initially been intended.

2.7.1.1 Test security

Burke (2009) identified a key risk namely that content can be fraudulently accessed, memorised, learned or recognised. From a systems perspective Foster (2010) indicated that often other technological devices can affect testing such as cell phones, digital recorders or videos. Naglieri et al. (2004) explained that the Item Response Theory and computerised adaptive tests are specifically well suited to Internet testing as it can tailor a test to the individuals' responses. Naglieri et al. (2004) referred to levels of test security varying from highly secure and restrictive levels for high stakes testing to lenient and unsecured levels for low stakes testing. From a technological systems perspective on test security Naglieri et al. (2004) referred to the importance of reducing unauthorised intrusions into client test data. On the internet this meant that three independent servers were crucial for test security: an internet server, a test application server and a database server. Also Coyne and Bartram (2006) referred to stakeholders such as the test developers, publishers and users. Backups become essential in case data need to be recovered. Also when completing a test on the internet functions such

as copy, paste, export or print screen should be disabled or locked down to preserve the integrity of the test (Foster, 2009; Naglieri et al., 2004). However, what is to stop a person from using a cell phone to take pictures of item content of a test during an unproctored session? Pearlman (2009) also mentioned that test security and other problems in high stakes testing could possibly reduce the validity of the test. Bartram (2000) argued that whilst many were concerned about internet security a well designed internet system was possibly far more secure than local computer networks and intranets. The ITC guidelines (ITC 2005) and Joubert and Kriek (2009) referred to test security as a major issue, with specific focus on the security of testing materials, privacy, confidentiality and data protection.

2.7.1.2 Cheating

The ITC guidelines (ITC, 2005) referred to cheating as it relates to the open or controlled mode of unproctored testing and identifying this as a concern to the test publishers. With efforts to counter cheating, test users should be informed about honesty policies and subsequent validation assessments to verify the authenticity of results (ITC, 2011) which can be done by making use of a test verifiers after initial online testing. During UIT the possibility of cheating or fraudulent behaviour could not be ignored. It would seem that the most common solution to cheating is to allow for UIT as the first testing session with a verification test or re-examination to follow after the first testing session (Naglieri et al., 2004). Tippins (2009) mentioned that cheating associated with UIT might not be much greater than cheating associated with proctored conditions. Pearlman (2009) suggested that to overcome obstacles such as cheating in UIT there would really be no way to do so without technology. If video cameras and surveillance are used on a “live” setting then the situation is in effect no longer unproctored.

Furthermore Tippins et al. (2006) indicated that adaptive testing presumably enhances test security by eliminating the possibility of copying paper tests and scoring keys that can be distributed within organisations during operational procedures. Copying, memorising and photographing test content relates to the larger problem of only some

of the possible ways how tests can be compromised (Tippins et al., 2006). The fact is that if cheating has occurred in any form, the validity of the test has been compromised.

2.7.1.3 Candidate identification

Establishing the identity of the person completing a test has potential problems but one method to address this is verification testing (ITC, 2005; Lievens & Burke, 2011). Even in a technological world where very little is impossible, many ideas around identification such as eye scanners, fingerprint matches or keystroke monitoring were reasonably foolproof but also probably so expensive that it would be difficult to implement in practice. Foster (2009) suggested many alternative ways of identifying candidates, but it seems that there could be questions around whether the candidates are who they say they are. Tippins et al. (2006) discussed psychometric identification in the form of verification testing. However, some may argue that if verification tests are needed in the first place, it implies a certain level of uncertainty about the initial test result in the UIT setting. Pearlman (2009) explained that if cheating occurred methods such as verification after cheating are irrelevant. Burke (2009) described verification testing as being when short tests are administered in a proctored setting at a later stage to check the consistency of scores. Pearlman (2009), however, indicated that even if cheating occurred such as results being inflated or completed by someone other than the test taker, and if such cheating was identified during verification testing, the inappropriate disqualification (false negatives) of the qualified person would by then have taken place.

2.7.1.4 Culture Fairness

The internet and the implications of internationalisation mean that test users have to keep in mind that norm groups could be international and boundaryless rather than local and that implications of this need to be considered (Bartram, 2006). Foster (2010) suggested that more often high stakes testing were taken by individuals from different cultures and countries. Bartram (2006) raised important questions such as in which country does the test user need to have his/her qualifications and which country's test standards or codes of practice applied? Foxcroft et al. (2004) indicated that the test user remains responsible for ensuring that instruments being used are valid for the purpose

of testing and reliability, and norms should be considered. A further question arises about when test takers have been treated unfairly, to whom and in which country can such issues be addressed (Foxcroft & Davies, 2006).

2.7.2 Advantages and disadvantages of internet delivered tests

The internet has become part of people's daily lives. (Bartram, 2002; Davies et al., 2009). Advantages of the internet are that large numbers of participants can be tested quickly, that heterogeneous samples of people can be recruited. Moreover internet testing is cost effective as time, space and administration can be overcome and the test is brought to the participant rather than bringing the participant to the test site (Eid & Diener, 2006). The internet has made internationalisation of testing possible where an applicant in one country can be assessed for a position in another country, (Bartram, 2004). Furthermore new tests can be accessed by test publishers and updating tests becomes easier (Naglieri, et al., 2004). Also responses can be recorded and stored almost automatically (Naglieri, et al., 2004). Whilst the internet is rapidly growing concerns around internet-based testing are many.

One of the main concerns of internet assessment was that individuals are assessed in an uncontrolled (unproctored) environment. Joubert and Kriek (2009) suggested that some internet tests were readily available but were not necessarily scientifically validated. The completion of psychological tests during UIT provides a candidate friendly application process but offers exposure to cheating (Pearlman, 2009). In addition there were a number of uncertainties with regards to standardised instructions and methodology. This became a hurdle in terms of test fairness and could impact on the results (Davies et al., 2009). Some of the serious concerns reflected by different authors (Davies et al., 2009; Naglieri et. al., 2004) was that the test administrator could be bypassed, making it easy for unqualified people to have access to very personal information which opened up to the misuse of measures. Naglieri et al. (2004) implied that the tests available on the internet were possibly not appropriate for some groups. The concerns regarding internet-based assessment were exacerbated within third world and developing countries, as there were additional challenges such as high computer

illiteracy levels, as well as inaccessibility to computer-based facilities (Davies et al., 2009). Joubert and Kriek (2009) referred to the increase in internet usage in Africa as well as South Africa; however, it should be kept in mind that millions of Africans and South Africans do not yet have access or the means to use the internet (Davies et al., 2009). Technology is becoming more widespread in South Africa and adverse impact is probably diminishing, however the level of technological sophistication which can impact on test performance is still present (Davies et al., 2009).

In addition Bartram (2006) explained that the test distributor on the internet was responsible for the control practised regarding the internet-based test. Bartram (2006) furthermore highlighted that the end user should be considered as it sometimes is not the registered practitioner but the line managers who are making decisions based on the results of the assessment. Joubert and Kriek (2009) furthermore highlighted the importance of determining the equivalence of scale scores of measuring instruments in different modes of administration to ensure that the psychometric properties of a test have been accurately adjusted for the online version. Registered professionals have to be vigilant when making use of computer-based and online assessments and should take extra precautions to ensure that test use is in line with regulation as the absence of guidelines can lead to abuse of tests (Foxcroft et al., 2004). Moreover, if a test taker does not have a qualified person to assist with problem situations it raises issues such as standards of administration and control over the entire testing process (Coyne & Bartram, 2006).

2.8 CHAPTER SUMMARY

Foxcroft and Davies (2006) mentioned that adherence to international guidelines as well as local guidelines for test use is of high importance, but also that test developers should document evidence of the equivalence of computer-based and internet-delivered testing. The future of psychological testing with reference to computerised and internet testing methods offers many exciting opportunities, however, like everything else in psychology, the context and specific needs related to the purpose of testing should be

considered. At this point in time there cannot be clear cut answers to unproctored internet-delivered tests or the use of internet-based testing, unless such uses have been well researched. The fact is that ways and means to use internet and computerised testing practically and effectively in the workplace should go hand in hand with ethical standards of practice.

CHAPTER 3

RESEARCH ARTICLE

The Effect of Mode of Test Administration on Computerised Assessment Results Using Proctored and Unproctored Test Administration Procedures

Francina Helena Nel

Department of Industrial and Organisational Psychology

University of South Africa

ABSTRACT

Orientation: Computerised and internet-delivered psychometric testing have in the past few years increased rapidly due to technological developments worldwide. However, debates around the benefits and challenges related to internet-based testing are ongoing.

Research purpose: Test administration via the internet could involve unproctored circumstances. The moment the proctor is removed from the testing condition, the psychometric properties of tests in the unproctored compared to proctored mode needs investigation to ensure equivalence of measures for the sake of the validity of interpretation of the results. The purpose of this study was to investigate if the same sample group's test scores differ between modes of administration.

Motivation for the study: The focus of the study was on identifying whether the mode of test administration could have an effect on computerised assessment results for a cognitive test (LPCAT) and an interest test (CPCAT). This study could assist with certain decisions related to the online use of these measures.

Research design, approach or method: A quantitative study was conducted using a quasi-experimental repeated measures design. Convenience sampling was used and for the LPCAT (N=82) and CPCAT (N=81) participants from within the hospitality industry - for which scores on both the proctored and unproctored test administration sessions were available - formed the sample group for the data analysis.

Main findings: For LPCAT the total group mean scores showed no statistically significant differences for mode of administration. For CPCAT total group five out of 34 sub-dimensions yielded statistically significant differences in means scores for mode of administration. For sequence effect for the total groups, statistically significant differences between the mean scores of the two test sessions were found for all scores of the LPCAT and for some scores of the CPCAT.

Practical implementations: Findings are limited to this study, where mode of administration did not affect computerised assessment results significantly. For sequence effect on LPCAT it is concluded that for the total group and sub groups better performance was achieved during second testing sessions regardless of mode of administration. For the five CPCAT sub-dimensions where mean scores did differ significantly, the scores for the proctored session were consistently higher than the scores for the unproctored session. When completing the test unproctored a certain level of language proficiency could be needed if participants do not have assistance from administrators.

Contribution: These findings contribute to understanding the effect of proctored and unproctored test administration on cognitive and non-cognitive measures respectively.

KEY WORDS

Career Preference Computerised Adaptive Test (CPCAT), computerised adaptive testing, internet-based testing, Learning Potential Computerised Adaptive Test (LPCAT), proctored, unproctored.

INTRODUCTION

Both the field of Information Technology (IT) and the internet have seen rapid technological and scientific developments in the past few years (Coyne & Bartram, 2006). Conducting psychological assessment via the internet is becoming a convenient way for organisations to assess individuals globally (Beaty, Nye, Borneman, Kantrowitz, Drasgow & Grauer, 2011). Tippins, Beaty, Drasgow, Gibson, Pearlman, Segall and Shepherd (2006) explained that unproctored internet testing (UIT) allows test takers to take psychological tests at times and places convenient to them, thereby eliminating supervision and administrators, even eliminating the "class room" scenario. Whilst the internet provides many advantages for future psychological assessments, the acceptable use thereof is often debated - especially for high stakes testing (Tippins, 2009). There are many concerns related to internet-based testing such as test security, issues of good practice, confidentiality, control, validity and reliability (Beaty, Dawson, Fallaw & Kantrowitz, 2009; Naglieri et al., 2004; Tippins, 2009; Tippins et al., 2006).

Tippins (2009, p. 7) made the following point: "If one relies on reliability and validity evidence of a test administered under proctored conditions, the psychologist cannot accurately describe the reliability and validity of the inferences made under unproctored conditions." The realistic appropriateness and implementation of psychometric testing in the unproctored setting therefore becomes the responsibility of the registered psychologist (Tippins, 2009; Tippins et al., 2006).

While the internet in general allows for easy access of international products and crossing of international boundaries, for the field of psychology in South Africa adherence to national legislation remains essential (Foxcroft & Davies, 2006; Foxcroft, Paterson, Le Roux, & Herbst, 2004; Foxcroft, Roodt & Abrahams, 2009).

South Africa is one of the few countries with specific legislation regarding psychological assessment. The Employment Equity Act no. 55 of 1998 (section 8) states that:

> *Psychological testing and other similar forms or assessment of an employee are prohibited unless the test or assessment being used: a) has been scientifically*

shown to be valid and reliable; b) can be applied fairly to all employees; and c) is not biased against any employee or group.

Foxcroft and Davies (2006) referred to the ITC guidelines point 20.3 (ITC, 2005) which stipulates that test publishers should only publish and offer online tests that have sufficient psychometric evidence to support such use.

The Career Preference Computerised Adaptive Test (CPCAT) has been designed as an internet-based self-rating interest measure (De Beer, 2011). The Learning Potential Computerised Adaptive Test (LPCAT) measures learning potential and was originally designed as an adaptive computerised measure but was not initially designed for online use (De Beer, 2005). In order to keep abreast with new technology as well as to provide better system support, the LPCAT will in future be available in online form (De Beer, 2012). Whilst LPCAT will not be administered in unproctored mode as it is a cognitive and high stakes test, the questions around internet-based testing and the validity of tests through the medium of the internet should nevertheless be further explored.

Background to the study

Coyne and Bartram (2006) explained that computer-based testing has been available for many years with the accredited user purchasing the materials and exercising direct control over the use of the computer-based test. The internet, however, changed this mode of test delivery and the question is how will the testing community react to the human element only being part of the process at the interpretation stage of testing (Coyne & Bartram, 2006). In this study the level of control involved in psychological assessment was considered important. The proctored mode of administration can be defined as the test administration process that is supervised by a qualified administrator (ITC, 2005). The unproctored mode of administration, on the other hand, is when there is no human supervision during testing, as test takers complete the test entirely by themselves. Tippins et al. (2006) defined unproctored testing as the testing event not being monitored, thereby resulting in test takers not being identified and their behaviour not observed.

The unproctored mode of administration, often related to internet-based assessment, could host an environment perceived to be less stringent than that of the traditional proctored administration. Salgado and Moscoso (2003) found that individuals show positive reactions and perceptions to internet-based versions of personality tests compared to paper-and-pencil versions. However, the question remains if unproctored conditions mean that test takers would experience the absence of supervision as positive, more relaxing or beneficial during cognitive high stakes testing. Moreover the test taker will not have a qualified person to assist with problem situations.

Various concerns around internet-based testing emerged such as ethics around the context of testing and standards of administration, test security and authentication of identity of test takers as well as validity of tests in unproctored mode (Foxcroft & Davies, 2006; Foxcroft et al., 2004; Naglieri et al., 2004, Tippins et al., 2006, 2009). Pearlman (2009) mentioned that test security and other problems in high stakes testing could possibly reduce test validity. Similarly Joubert and Kriek (2009) referred to test security as a major issue, with a specific focus on the security of testing materials, privacy, confidentiality and data protection.

Cheating is always a matter for concern as test validity could be compromised (Tippins, 2009). Pearlman (2009) suggested that to overcome obstacles such as cheating in UIT, there would really be no way to do so without the help of technology such as video cameras during testing. In addition the importance of identifying whether the right candidates complete tests online indicates concerns around identity. Foster (2010) believed that internet security risks can be effectively managed by the use of technological efforts to protect and secure tests. It was also implied that observation during unproctored testing through webcams of the torso and hands and audio devices could be used. The question could be asked, however, that if technological assistance is used to observe the test taker during testing then does the setting truly imply unproctored testing?

Whilst the concerns about internet testing are debated, benefits also exist. Hense, Golden and Burnett (2009) indicated that an increasing number of organisations recruit

and select individuals via the internet. From a human resource perspective, to manage recruitment and selection processes efficiently in practice, the institution of technological testing infrastructures became important (Bartram, 2000, 2006). In addition the internet is fast becoming an easily accessible medium and large numbers of participants could be assessed quickly (Eid &Diener, 2006; Joubert & Kriek, 2009). The main benefits include time and cost savings (Davies, Foxcroft, Griessel & Tredoux, 2009; Gregory, 2007)

The ITC guidelines for Quality Control in Scoring, Test Analysis and Reporting of Test Scores (ITC, 2011) suggested that practitioners had to have a broad understanding of quality control practices, which were of critical importance for tests to be used ethically, accurately and responsibly.

Salgado and Moscoso (2003) explained that the internet would be used for personnel selection procedures and that test instruments would be developed for web use. For LPCAT and CPCAT it would be beneficial to identify whether the mode of administration could affect computerised results so as to extend good practice related to the control of test use. The instruments are based on computerised adaptive test methods and Beaty et al. (2009) indicated that CAT is a promising strategy for efficiently rotating test content thereby reducing the chance that candidates will encounter the same items. The possible effect of unproctored conditions associated with internet-based testing would add valuable information about the tests in the alternative form.

Trends from the literature review

In terms of psychometric properties and equivalence of scores between the two versions of computer-based testing (CBT) and internet testing, the ITC (2005) provided guidelines such as comparable reliabilities for CBT and internet testing to be shown, that such tests should correlate with external criteria, and should produce comparable means and standard deviations. Past research was mostly focused on comparable scores on personality questionnaires (Bartram & Brown, 2004; Joubert & Kriek, 2009; Salgado & Moscoso, 2003).

Salgado and Moscoso (2003) investigated whether the Big Five personality questionnaire (IP/5F) paper-and-pencil version could be translated into an internet based version without loss of psychometric properties, and to explore perceptions about the internet-based testing. The findings revealed firm conclusions that the two versions mean scores and standard deviations were similar. Participants also perceived the internet-based tests as a positive experience. These conclusions were limited to personality questionnaires.

Joubert and Kriek (2009) investigated construct equivalence of the Occupational Personality Questionnaire 32 (OPQ32i) when administered in an online and paper-and-pencil mode of administration. No statistically significant differences were found between internet and paper-and-pencil results with regard to the constructs measured.

Furthermore Beaty et al. (2011) investigated whether predictive validity was shown when tests were taken off-site without a proctor present. The results showed similar magnitude for validity coefficients of a non-cognitive assessment (Beaty et al., 2011). When testing for equivalence between paper-and-pencil and internet-based modes Salgado and Moscoso (2003, p. 200) said that “the strongest evidence of the equivalence on both versions of the same measure is achieved using the same group of individuals (within-group experimental design)”. It was suggested by Salgado and Moscoso (2003) that where previous research was often addressed by the two versions namely paper-and-pencil and computer-based tests, comparing independent groups the actual equivalence was not directly examined. This study however addressed the within-participants design for both measures, specifically including a cognitive and non-cognitive test.

The debate around unproctored internet testing is ongoing; however, there are only a few published studies particularly on the validity of assessments during UIT (Beaty et al., 2011). Some literature has focused on the areas for concern regarding UIT such as test security, authentication and cheating (Beaty et al. 2009; Burke, 2009; Pearlman, 2009) while other literature referred to computer anxiety and ethical concerns, when unproctored testing is to be considered (Burke, 2009; Tippins et al., 2006).

Regarding systems and practical implementations of internet-based testing; Bartram (2006) referred to secure infrastructure, the development of more accessible or better software, and hardware and internet services with a shift in locus of control from the client-side to server side. In addition, the distributor who manages the internet server retains control by providing access to clients, but not maintaining control over the test taking conditions. Test users may request for a person to complete a test but never see the documentation, scoring or norms and would rely on the computer generated report, thereby changing the relationship between the publisher, test user and test taker (Bartram, 2006).

Use of internet-based testing is growing (Bartram, 2006; Foster, 2010). Beaty et al. (2009) suggested that instead of a debate as to whether or not to proctor, the focus should be on improving UIT. Advantages during computer-based testing could enhance consistency of delivery or improve the efficiency of delivery (Tippins et al, 2006). Computer tests provide consistent instructions, precise timing as well as accurate scoring, and also allow for precise processes (Davies et al., 2009; Gregory, 2007). Internet testing could result in cost savings, testing at convenient hours and complete and accurate data records (Tippins et al., 2006).

Advice for the future use of tests that are computer-based and internet-based involves adherence to professional and ethical guidelines regardless of mode of administration. (Tippins et al., 2006). Effective communication to line managers on the use of psychometric testing which could include online tests and thorough policies related to psychometric testing are important to navigate successful assessment programs (Foxcroft et al., 2004). Furthermore, it is suggested by some experts that UIT could be used in low stakes testing but should be rejected in high stakes testing situations (Tippins et al., 2006).

Research Objectives

The objective of the study was to determine if mode of administration could affect computerised test results. It is important for practitioners to consider the possible effects

when tests are being used via the internet so as to consider the possible differences of psychometric properties and test results.

The aims of the empirical study were:

- to determine the effect of mode of administration on LPCAT and CPCAT mean test scores respectively, thereby addressing equivalence questions around proctored and unproctored mode; and
- to determine if sequence effect played a role in the study

Measurements

The Learning Potential Computerised Adaptive Test (LPCAT) is a cognitive power test that measures learning potential by means of non-verbal figural, general fluid reasoning ability (De Beer, 2012). LPCAT is a test specifically used in South Africa to identify learning potential. Whereas ability is viewed as an acquired skill on demand, potential is based on what could be (De Beer, 2005). It is based on the test-train-retest approach and Vygotsky's theory of the zone of proximal development (De Beer, 2005). The zone of proximal development distinguishes between current performance and future potential performance. According to De Beer (2012), dynamic assessment is based on the same concept where the pre-test score is obtained without assistance and the post-test score is obtained after training. In practice LPCAT is used for obtaining information related to current and future levels of learning potential in line with South African National Qualifications Framework (NQF) levels. With the end purpose in mind, the current and projected future NQF levels could then be identified (De Beer, 2012). For example if a training and development opportunity exists, a certain NQF level would be required to cope with the level of training, thereby making it possible to use LPCAT t-scores linked to NQF levels to identify the level at which an individual could cope with training and learning tasks.

The Career Preference Computerised Adaptive Test (CPCAT) on the other hand could be used for vocational career guidance (De Beer, 2011). The test measures three domains namely career fields, activities and environments.

The potential contribution of the study

Several studies in the past investigated internet-based testing, related to personality (Bartram & Brown; 2004, Coyne, Warszta, Beadle & Sheehan, 2005; Joubert & Kriek, 2009; Salgado & Moscoso, 2003). Salgado and Moscoso (2003) highlighted the importance of comparing the same participants' scores when investigating the mode of administration for online equivalence. Bartram and Brown (2004) investigated the bias effect on the OPQ32i from data obtained during online administration where independent groups were tested without supervision and by means of the traditional supervised paper-and-pencil mode of administration. The findings of the investigation indicated that the web-based unsupervised controlled mode of administration had psychometric properties comparable to paper-and-pencil supervised mode of administration. Furthermore, Salgado and Moscoso (2003) examined whether the Big Five personality questionnaire can be adapted to an internet-based version with the maintenance of psychometric properties. Findings indicated that both versions were equivalent in terms of distribution, reliability and factor structure. Salgado and Moscoso (2003) indicated that personality measure equivalence between the modes of administration could not be generalised to cognitive tests or other personnel selection procedures.

Styles and Andrich (1993) investigated the Ravens Standard Progressive Matrices (RSPM) a cognitive non-verbal test of intelligence, to compare the paper-and-pencil version with the computerised form. Their findings showed that the CAT version covered a wider range of the continuum with item difficulty levels when selecting items than the paper-and-pencil version. Recently Lievens and Burke (2011) indicated that with verbal and numeric tests unproctored scores were higher than proctored scores, with aberrant scores for graduates on numeric tests and verbal tests. Suggestions for future research were that test security and honest responding strategies be scrutinised.

Tippins et al. (2006) stated that the kind of research required to facilitate internet testing included source of error variance effects on validity and descriptive statistics on possible changes of test scores. The level off validity that could be expected during UIT is

therefore a practical challenge. Salgado and Moscoso (2003) identified the need for more research regarding assessment procedures related to internet-based testing of, for example, cognitive tests. According to Bartram (2000), there are very few examples of tests that have been published for computer administration that could not also be produced in paper-and-pencil versions. Identifying whether test results are different based on unproctored testing as the choice of mode of test administration should be further explored (Salgado & Moscoso, 2003) which was the focus of the present study.

Future developments for both LPCAT and CPCAT will include internet-based programs for processing and access to results programs allowing users to access results information from databases online (De Beer, 2012). Internet-based adaptive test administration is another planned future development to improve test administration processes (De Beer, 2012). What makes this study unique is that LPCAT and CPCAT are examples of tests that have never been available in paper-and-pencil form, but rather were developed as measures only for computerised and internet administration respectively. Also the tests are computer adaptive and the LPCAT is also based on the dynamic test-train-retest assessment approach. The need for research regarding examples of cognitive (learning potential) and non-cognitive (interest) tests is addressed in this study by investigating the possible effects of the mode of administration on LPCAT and CPCAT results respectively.

RESEARCH DESIGN

The research design in this study refers to the research approach, the research method and information about the sampling, data collection and analysis.

Research approach

In the study a quantitative approach and quasi-experimental research design was used. The choice of quasi-experimental design often implies that full randomisation associated with experimental designs is not possible. In quasi-experimental research the participants are randomly pre-assigned to groups. However, due to challenges beyond

the researcher's control - a protected strike during the commencement of the study took place - participants were not randomly pre-assigned to groups. Convenience sampling was used with ad hoc assignment to the two groups used respectively to allow for monitoring of the sequence of proctored and unproctored test administration. The research setting and independent variable (mode of administration) was altered by the researcher between proctored and unproctored sessions including two test instruments. The small sample size (N=82 and N=81 for LPCAT and CPCAT respectively) with normal distribution of the variables of concern lead to the primary data being analysed using the dependent t-test as analysis method. The dependent t-test test (Field, 2005) was used to compare mean scores obtained by means of proctored and unproctored test administration for the same group of participants in the within-participants design.

Research method

The hypotheses are as follows:

$H1_0$: There are no statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for cognitive (LPCAT) results.

$H1_1$: There are statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for cognitive (LPCAT) results.

$H2_0$: There are no statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results.

$H2_1$: There are statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results.

$H3_0$: There are no statistically significant differences between the mean scores of the first and second test sessions respectively for cognitive (LPCAT) results.

$H3_1$: There are statistically significant differences between the mean scores of the first and second test sessions respectively for cognitive (LPCAT) results.

$H4_0$: There are no statistically significant differences between the mean scores of the first and second test sessions respectively for non-cognitive (CPCAT) results.

$H4_1$: There are statistically significant differences between the mean scores of the first and second test sessions respectively for non-cognitive (CPCAT) results.

The hypotheses were non-directional.

The research participants, measuring instruments, method of statistical analysis and research procedures will be discussed in the section that follows.

Research procedure

The study was conducted over a period of three months. Participants within the tourism industry were tested during work hours for development purposes. Ethical clearance was obtained from the higher education institution as well as from the participating organisation. The organisation agreed that results would be used for development of the employees in terms of possible training and development opportunities. The purpose of the research and intended use of the results were explained to all the participants and were agreed upon with the organisation. Participants were told that participation would be voluntary and that participation could be declined at any given point during the study. Participants were to receive individual feedback also for personal development purposes. Furthermore each participant provided written consent to participate in the study. Participants were asked to complete biographical forms as well as to report their level of computer competency.

Participants were assessed in groups with a maximum of six individuals being tested at one time.

The unproctored session implied controlled mode at the start of the session when LPCAT was the first test administered. Identity verification was first completed, and assistance provided with logging on. Participants were then left alone and were

expected to complete and exit out of LPCAT test and log into the online CPCAT test session which simulated the controlled unproctored mode of administration. The employee then in controlled mode needed to complete biographical information online and used a booklet explaining the steps to complete the CPCAT online as well as completing the CPCAT questions online.

The proctored session entailed the managed mode where high levels of supervision and control are maintained (Coyne & Bartram, 2006; ITC 2005). Employees received assistance with LPCAT log on, identity verification was done, employees were supervised and had the opportunity to ask questions during testing; to an extent building a relationship with the supervisor during testing. Employees were then assisted with entering information online for CPCAT and received any other necessary assistance.

In essence for the "unproctored" (UP) group (unproctored first session, proctored second session) meant that the first test session (test 1) required unproctored test administration and test 2 for this group required proctored test administration. For the "proctored" (PU) group (proctored first session, unproctored second session) the first test session (test 1) required proctored test administration and test 2 for the same group required unproctored test administration. Proctored mode of administration implies that a supervisor is present during the test session. In this study the managed mode applied. Unproctored mode of administration means that the participants had no human supervision during testing, however, the controlled mode was used whereby participants received assistance to log into the tests. For controlled mode used for the unproctored testing, no direct supervision is provided but in order to access the test a logon code is made available to a particular individual, thereby controlling who can access the tests on the internet (ITC, 2005).

Employees attended testing sessions during work hours at the company premises within the Human Resources department. Depending on the availability of participants, the number of weeks between the two testing sessions varied as employees attended sessions convenient to the business events and subject to the manager's agreement. Employees attended the second testing session between two weeks after and three

months after the first testing session. Due to the LPCAT being an adaptive test, any change in answers from the first test session to the second session could potentially lead to new items presented which to an extent limited potential recognition of test items on LPCAT. However it is possible that the training intervention during the test could have been remembered.

Several challenges that occurred during the test sessions included the loss of the internet signal, technical challenges and a few tests not having been saved accurately when participants had to exit from the tests when finished. In addition certain environmental challenges have to be mentioned such as union members being on strike during the time of the study and some of the participants going on annual leave during this time.

The results of the simulated internet unproctored mode were compared to the proctored mode results. Since the LPCAT is not available online, simulations of online mode on LPCAT were conducted with unproctored settings suggesting online circumstance, while CPCAT was completed online for both modes of administration.

Research participants

The research participants were employees in a Hotel and Golf Resort in the Western Cape, South Africa. The final samples for the LPCAT (N=82) and CPCAT (N=81) results respectively comprised those individuals who had participated in the research and for whom two sets of results for the respective measures were available. Table 1 provides biographical variables for the LPCAT total group, for which the ages of the participants ranged from under 21 years (4.9%), 21 to 30 years (39.0%), 31 to 40 years (31.7%), 41 to 50 years (22.0%) and older than 50 years (2.4%) with a mean age category score of 2.78 or between 21 to 30 years of age and standard deviation of 0.930. In terms of the two sub groups the proctored (PU) group had slightly more participants in the distribution of ages 21-30 (40.9%) versus the unproctored (UP) group (36.8%).The distribution for participants ages under 21 were 2.6% for the unproctored (UP) group and 6.8% for the proctored (PU) group, also ages 31 to 40 for the proctored (PU) group were 31.8% whereas the unproctored (UP) group distribution was 31.6%.

The unproctored (UP) group had more participants between ages 41-50 (26.3%) than did the proctored (PU) group (18.2%).

The gender distribution of the total group was 57.3% males and 42.7% females, not representative to the national gender ratio which was indicated in the 2011 national census in South Africa to be 48.7% males and 51.3% females (StatsSA, 2011). However within the organisation at the time the ratio of males were 56.5% and females 40.8% with a good organisational representation. The male distribution for the groups were somewhat balanced as the unproctored (UP) group consisted of 57.9% males and the proctored (PU) group of 56.8% males. The proctored (PU) group had 43.2% females and the unproctored (UP) group 42.1%.

In terms of the language of the group, the majority of participants (61.0%) were Afrikaans speaking, 4.9% English and 31.7% Xhosa speaking. The unproctored (UP) group consisted of more Afrikaans speaking participants (68.4%) compared to the proctored (PU) group which had Afrikaans speaking participants (54.5%). The proctored (PU) group had more Xhosa speaking individuals (34.1%) than the unproctored (UP) group (28.9%).

For school grade the majority (46.3%) of the total group was at the grade 12 level with (39.0%) of the total group for grade 8 to grade 11. The unproctored (UP) group had more participants in the grade 8 to grade 11group (42.1%) than the (PU) group of (36.4%). The (PU) group, however, had more participants at the grade 12 level of (50%) than the (42.1%) participants from the (UP) group.

The majority of the group was at a B1 Patterson grading level (52.4%) with the (UP) group consisting of (57.9%) and the PU group consisting of (47.7%) only.

For the CPCAT total group, the ages of the participants ranged from under 21 years (4.9%), 21 to 30 years (38.3%), 31 to 40 years (33,3%), 41 to 50 years (21.0%) and older than 50 years (2.5%) with a mean age category score of 2.78 or between 21 to 30 years of age and standard deviation of 0.922. In terms of the sub groups the unproctored (UP) and proctored (PU) group the proctored (PU) group had slightly more

participants in the distribution of ages 21-30 (38.9% and 37.8% respectively). The distribution for participants of ages under 21 were 2.8% for the unproctored (UP) group and 6.7% for the proctored (PU) group, also ages 31 to 40 for the proctored (PU) group were higher than the unproctored (UP) group (35.6%) whereas the unproctored group distribution was 30.6%. The unproctored (UP) group had more participants between ages 41-50 (25.0%) than did the proctored (PU) group (17.8%).

The gender distribution of the total group was 58% males and 42.0% females, a good representation of the organisation overall gender ratio. The male distribution for the groups were somewhat balanced as the unproctored (UP) group consisted of 58.3% males and the proctored (PU) group of 57.8% males. The proctored (PU) group had 42.2% females and the unproctored (UP) group 41.7%.

In terms of the language of the group, the majority of participants (59.3%) were Afrikaans speaking, 3.7% English and 34.6% Xhosa speaking. The unproctored group (UP) consisted of more Afrikaans speaking participants (66.7%) compared to the proctored (PU) group which had 53.3% Afrikaans speaking participants. The proctored (PU) group had more Xhosa speaking individuals (37.8%) than the unproctored (UP) group (30.6%).

For school grade the majority (45.7%) of the total group was at the grade 12 level with (39.5%) of the total group for grade 8 to grade 11. The unproctored (UP) group had more participants in the grade 8 to grade 11group (41.7%) than the (PU) group of (37.8%). The (PU) group, however, had more participants at the grade 12 level of (48.9%) than the (41.7%) participants from the (UP) group.

The majority of the group was at a B1 Patterson grading level (53.1%) with the (UP) group consisting of (58.3%) and the PU group consisting of (48.9%) only.

Table 1

Frequency distributions for biographical variables of participants

	Total Group LPCAT (N=82)		Unproctored 1st Group (n=38)		Proctored 1st Group (n=44)		Total Group CPCAT(N=81)		Unproctored 1st Group (n=36)		Proctored 1st Group (n=45)	
Gender	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Male	47	57.3	22	57.9	25	56.8	47	58.0	21	58.3	26	57.8
Female	35	42.7	16	42.1	19	43.2	34	42.0	15	41.7	19	42.2
Age	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
<21	4	4.9	1	2.6	3	6.8	4	4.9	1	2.8	3	6.7
21-30	32	39.0	14	36.8	18	40.9	31	38.3	14	38.9	17	37.8
31-40	26	31.7	12	31.6	14	31.8	27	33.3	11	30.6	16	35.6
41-50	18	22.0	10	26.3	8	18.2	17	21.0	9	25.0	8	17.8
50+	2	2.4	1	2.6	1	2.3	2	2.5	1	2.8	1	2.2
Home Language	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Afrikaans	50	61.0	26	68.5	24	54.5	48	59.3	24	66.7	24	53.3
English	4	4.9	1	2.6	3	6.8	3	3.7	1	2.8	2	4.4
Xhosa	26	31.7	11	28.9	15	34.1	28	34.6	11	30.6	17	37.8
Zulu	-	-	-	-	-	-	-	-	-	-	-	
Other	2	2.4	-	-	2	4.5	2	2.5	-	-	2	4.4
School Grade	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
1-7	5	6.1	2	5.3	3	6.8	5	6.2	2	5.6	3	6.7
8-11	32	39.0	16	42.1	16	36.4	32	39.5	15	41.7	17	37.8
12	38	46.3	16	42.1	22	50.0	37	45.7	15	41.7	22	48.9
Diploma	6	7.3	3	7.9	3	6.8	6	7.4	3	8.3	3	6.7
Degree	1	1.2	1	2.6	-	-	1	1.2	1	2.8	-	
Patterson Grading	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
A1	4	4.9	2	5.3	2	4.5	4	4.9	2	5.6	2	4.4
A2	7	8.5	3	7.9	4	9.1	8	9.9	3	8.3	5	11.1
A3	3	3.7	1	2.6	2	4.5	3	3.7	1	2.8	2	4.4
B1	43	52.4	22	57.9	21	47.7	43	53.1	21	58.3	22	48.9
B2	8	9.8	3	7.9	5	11.4	7	8.6	2	5.6	5	11.1
B3	3	3.7	2	5.3	1	2.3	3	3.7	2	5.6	1	2.2
Learnerships	5	6.1	-	-	5	11.4	5	6.2	-	-	5	11.1
>B3	9	11.0	5	13.2	4	9.1	8	9.9	5	13.9	3	6.7

Furthermore as reflected in Table 1 for the LPCAT group, school grades of the total group reflected that 6.1% of participants obtained school education between Grade 1 and Grade 7, 39.0% of participants' education ranged between Grade 8 and Grade 11 and the majority of participants (46.3%) had obtained a Grade 12, 7.3% had obtained a tertiary diploma and 1.2% had obtained a university degree. A mean score of 2.59 or Grade 8-11 range and standard deviation of 0.769 is reflected for the total group. The unproctored (UP) group had more Grade 8-11 school grade (42.1%) participants than the proctored (PU) group (36.4%). The proctored (PU) group had more Grade 12 participants (50.0%) than the unproctored (UP) group (42.1%). The unproctored (UP) group had proportionally more diploma and university level participants.

In terms of the Patterson grade, the majority of participants (52.4%) were at the B1 level, 17.1% ranging from A1-A3 and 24.5% above B1 level with 6.1% learnership participants. The mean score for the total group was 4.44, a B1 level with standard deviation of 1.792. The unproctored (UP) and proctored (UP) group, whilst somewhat balanced in terms of A1 and A2 levels differed in terms of B1 level. The unproctored (UP) group consisted of 57.9% of participants at this level, whereas the proctored group only comprised 47.7% of participants at this level. The proctored (PU) group consisted of students in learnerships (11.4%) busy completing a practical year with no learners in the unproctored (UP) group.

Table 2 below indicates that for both LPCAT and CPCAT total groups and sub groups for age, the mean age were between 21-30 years of age. The school grade for both LPCAT and CPCAT total groups and subgroups were between grade 8 -11. In addition the Patterson grade for both LPCAT and CPCAT total groups and sub groups were at the B1 Patterson grade.

Table 2

Descriptive statistics for some biographical variables

Descriptive Statistics	LPCAT Total Group (N=82)		Unproctored 1st Group (n=38)		Proctored 1st Group (n=44)		CPCAT Total Group (N=81)		Unproctored 1st Group (n=36)		Proctored 1st Group (n=45)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age	2.78	0.930	2.89	0.924	2.68	0.934	2.78	0.922	2.86	0.931	2.71	0.920
School Grade category	2.59	0.769	2.61	0.823	2.57	0.728	2.58	0.772	2.61	0.838	2.56	0.725
Patterson Grade category	4.44	1.792	4.37	1.777	4.50	1.824	4.36	1.777	4.36	1.823	4.36	1.760

With particular importance for developing countries such as South Africa, access to computers which is at present limited should be considered (Foxcroft & Davies, 2006). Whilst the LPCAT only requires the use of the space bar, and enter key and during administration early instructions provides the opportunity and time to exercise and become familiar with the two keys, CPCAT requires the use of a mouse or touchpad. As the tests were computerised, participants were asked to rate their computer competency and average number of hours spent per week on a computer so as to identify computer familiarity.

In Table 3 it is shown that in terms of reported computer competency for the LPCAT total group 42.5% participants reported their computer competency as poor and for CPCAT 41.8%. For the LPCAT the total group reported average computer competency (43.8%) whereas for the CPCAT total group it was 44.3%. For the LPCAT total group 13.8% reported good computer competency and for the total CPCAT group 13.9%. For the unproctored (UP) group, participants reported a poor (45.5%) computer competency, more than the proctored (PU) group (35.8%). The proctored (PU) group indicated higher perceived (18.4%) very good computer competency than the unproctored (UP) group (6.8%). The LPCAT unproctored (UP) group reported lowest computer competency and the LPCAT (PU) proctored group reported highest computer competency.

Furthermore the for the LPCAT total group 64.2% reported working 10 or less weekly hours on computers and for the total group for CPCAT 65.0% reported these same hours.

Table 3

Reported computer competency

Computer competency	LPCAT Total Group (N=82)		Unproctored 1st Group (n=38)		Proctored 1st Group (n=44)		CPCAT Total Group (N=81)		Unproctored 1st Group (n=36)		Proctored 1st Group (n=45)	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Poor	34	42.5	17	45.9	17	39.5	33	41.8	15	42.9	18	40.9
Average	35	43.8	17	45.9	18	41.9	35	44.3	17	48.6	18	40.9
Good	11	13.8	3	8.1	8	18.6	11	13.9	3	8.6	8	18.2
Missing	2		1		1		2		1		1	
Weekly hours on computer	LPCAT Total Group (N=82)		Unproctored Group (n=38)		Proctored Group (n=44)		CPCAT Total Group (N=81)		Unproctored Group (n=36)		Proctored Group (n=45)	
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
0-10	52	64.2	23	60.5	29	67.4	52	65.0	21	58.3	31	70.5
11-20	14	17.3	9	23.7	5	11.6	14	17.5	9	25.0	5	11.4
21-30	10	12.3	4	10.5	6	14.0	10	12.5	4	11.1	6	13.6
31-40	5	6.2	2	5.3	3	7.0	4	5.0	2	5.6	2	4.5
Missing	1		-		1		1		-		1	

Measuring instruments

A biographical questionnaire was used to obtain data on gender, age, language, school grade, Patterson grading levels, and computer competency. Two psychometric instruments were used, namely, the Learning Potential Computerised Adaptive Test (LPCAT) as a cognitive (learning potential) measure and the Career Preference Computerised Adaptive Test (CPCAT) a non-cognitive (interest) questionnaire. The LPCAT has been reported to be an acceptable measure of learning potential (De Beer, 2004; Gilmore, 2008) and non-verbal figural reasoning and the CPCAT of career preference (De Beer, 2011).

The Learning Potential Computerised Adaptive Test (LPCAT)

The LPCAT is a non-verbal figural reasoning test which measures learning potential (De Beer, 2005). It is a dynamic test which uses a test-train-retest approach and enables test takers' potential to be identified (De Beer, 2005). The coefficient alpha internal consistency reliability scores of the LPCAT range from 0.925 to 0.987 for different groups (De Beer, 2005). For the current sample group (N=82), test-retest reliability for the LPCAT was satisfactory with correlation values of r=.730 (p=.000) for the LPCAT pre-test and r=.898 (p=.000) for the LPCAT post-test scores respectively. The scores that were used for statistical analysis in this study were the t-scores of the pre-tests and post-tests from each mode of administration respectively. The difference of LPCAT post-test results and the NQF level at which opportunities related to the purposes of testing are targeted can be interpreted as the extent of effort that the individual will need to apply in order to perform at optimal level (De Beer, 2012).

The Career Preference Computerised Adaptive Test (CPCAT)

The CPCAT is a test that is in its final stages of development and has not yet been released for general use. It focuses on measuring career interest but more specifically in three domains namely: career fields, career activities and career environments (De Beer, 2011). Due to the fact that many people do not have one career or one job in their lifetimes but rather different vocational experiences, they face many challenges in

adapting to multiple new entry points (De Beer,2011). The coefficient alpha internal consistency reliability values for the CPCAT range between 0.716 and 0.915 for the 34 sub- dimensions of the three main domains (De Beer, 2011). For the current sample group (N=81), the test-retest reliability values were somewhat low with test-retest correlation values for the CPCAT field sub-dimensions ranging between .510 and .820 (average .595). For the CPCAT activity sub-dimensions test-re-test correlations ranged between .304 and .578 (average .464) while for the environment sub-dimensions, values ranged between .357 and .676 (average .514). Since the CPCAT measures multiple sub-dimensions, with the level on each determined by the response to between either two or four questions (as a result of the adaptive test administration process), lower consistency over time could be explained to some degree.

The scores that were used for the purpose of this study were the averages respectively of all 34 factors in terms of the three main dimensions, namely fields, activities and environments on both modes of administration sessions.

Statistical analysis

The statistical analysis was performed using SPSS version 20.0. The 0.05 significance level was accepted for the interpretation of results.

The dependent t-test was appropriate as analysis method for the repeated measures design with two sets of results (proctored and unproctored) obtained for the LPCAT and the CPCAT respectively for reach participant. The effect of the independent variable, namely the mode of administration, on the dependent variable of test results for both LPCAT and CPCAT was analysed by comparing the mean scores obtained in the unproctored and proctored test administration session respectively for the LPCAT pre-test and post-test scores and for the 34 CPCAT sub-dimension scores.

RESULTS

In order to establish whether the distribution of the study variables deviated from comparable sets of scores with the same mean and standard deviation that are normally distributed, the Kolmogorov-Smirnov test was used (Field, 2005). The results

indicated that the LPCAT and CPCAT scores for the sample in this study are not significantly different from a normal distribution and the dependent t-test sufficed. In addition Cohen's (d) was used to calculate the effect size of the unproctored and proctored groups for both LPCAT and CPCAT mean differences for mode of administration and sequence effect. Based on Cohen (1992) a small effect size is .20, a medium effect size is .50 and a large effect size is .80 .

Presentation of results

Firstly the LPCAT results for the total group and sub groups will be presented and the CPCAT total group and sub group results will thereafter be presented. For each test results the following presentation will be provided:

- The descriptive statistics and dependent t-test results and associated effect size indicators of the total group for mode of administration comparisons are presented first. Thereafter the descriptives and dependent t-test results of the total group for the sequence effect comparisons are presented.
- Secondly the descriptive statistics and dependent t-test results for the sub groups are presented and identified as the "unproctored" (UP) group or the "proctored" (PU) group. There is no differentiation between the sub-groups for mode or sequence effect as test 1 for the unproctored (UP) group would be the unproctored test administration session and test 2 for the same group would be the proctored test administration session. The results for the sub-groups (UP and PU) are presented separately for the LPCAT results only for the sake of clarity – despite some repetition. Similarly, test 1 for the proctored (PU) group would be the proctored test administration session and test 2 for the same group would be the unproctored test administration session. For the CPCAT results – due to the large number of variables involved - this repetitive presentation is not done.

Learning Potential Computerised Adaptive Test (LPCAT) results

Table 4

a) Total group LPCAT descriptives and dependent t-test comparisons for the mode of administration (N=82)

	Pre-test							Post-test						
	Unproctored		Proctored					Unproctored		Proctored				
Total Group	M	SD	M	SD	t	Sig#	d	M	SD	M	SD	t	Sig#	d
	49.12	7.976	49.18	7.503	-.089	.929	.01	49.27	8.249	49.35	7.791	-.180	.858	.01

#2-tailed

b) Total group LPCAT descriptives and dependent t-test comparisons for the sequence effect (N=82)

	Pre-test							Post-test						
Total Group	Test session 1		Test session 2					Test session 1		Test session 2				
	M	SD	M	SD	t	Sig#	d	M	SD	M	SD	t	Sig#	d
	47.85	7.643	50.45	7.621	-4.192	.000	.34	48.18	8.279	50.44	7.592	-5.601	.000	.28

#2-tailed

Interpretation of LPCAT results

Table 4 a) indicates the total group results for the mode of administration. For the unproctored session a pre-tests mean score of 49.12 was obtained and for the proctored session for the pre-tests a mean score of 49.18 was obtained. A p-value of .929 was obtained indicating that there was not a statistically significant difference between the mean pre-test scores obtained in the unproctored and proctored sessions respectively. For the unproctored sessions for the post-test a mean score of 49.27 was obtained and for the proctored sessions for the post-test a mean score of 49.35 was obtained. A p-value of .858 was obtained indicating that there was not a statistically significant difference in the mean post-test scores obtained in the unproctored and proctored sessions respectively. The null hypothesis ($H1_0$) is thus not rejected, implying that there are no statistically significant differences in the mean LPCAT pre-test and post-test scores obtained in the proctored and unproctored mode of administration respectively.

Table 4 b) indicates the total group results for the sequence effect. During the first test sessions for the pre-test a mean score of 47.85 was obtained and for the pre-test during the second test sessions a mean score of 50.45 was obtained. A statistically significant p-value (p=.000) was obtained and an effect size of .34 indicating moderate practical effect. During the first test sessions for the post-test a mean score of 48.18 was obtained and for the post-test during the second test sessions a mean score of 50.44 was obtained. A statistically significant p-value (p=.000) was obtained and an effect size of .28 indicating moderate practical effect. $H3_0$ is therefore rejected. The sequence effect for the LPCAT scores for the total group showed statistically significant differences between the mean scores of the first and second test sessions respectively for both the pre-test and the post-test scores.

Table 5

a) "Unproctored" (UP) group comparisons for the mode of administration and sequence effect for LPCAT results

N=38	LPCAT pre-test							LPCAT post-test						
UP group:	Unproctored		Proctored					Unproctored		Proctored				
Mode	M	SD	M	SD	t	Sig*	d	M	SD	M	SD	t	Sig*	d
comparison	47.79	7.308	50.66	6.679	-2.920	.006	0.41	47.84	8.035	50.37	6.756	-4.247	.000	0.34

UP group:	Test session 1		Test session 2					Test session 1		Test session 2				
Sequence	M	SD	M	SD	t	Sig*	d	M	SD	M	SD	t	Sig*	d
comparison	47.79	7.308	50.66	6.679	-2.920	.006	0.41	47.84	8.035	50.37	6.756	-4.247	.000	0.34

*2-tailed

b) "Proctored" group (PU) comparisons for the mode of administration and sequence effect for LPCAT results

N=44	LPCAT pre-test							LPCAT post-test						
PU group:	Unproctored		Proctored					Unproctored		Proctored				
Mode	M	SD	M	SD	t	Sig*	d	M	SD	M	SD	t	Sig*	d
comparison	50.27	8.423	47.91	8.005	-2.983	.005	0.29	50.50	8.323	48.48	8.566	-3.669	.001	0.24

PU group:	Test session 1		Test session 2					Test session 1		Test session 2				
Sequence	M	SD	M	SD	t	Sig*	d	M	SD	M	SD	t	Sig*	d
comparison	47.91	8.005	50.27	8.423	-2.983	.005	0.29	48.48	8.566	50.50	8.323	-3.669	.001	0.24

*2-tailed

Table 5 indicates results per sub group. For the unproctored (UP) group during the first test session (unproctored) a mean score of 47.79 for the pre-test and 47.84 for the post-test was obtained. The same group during the second test session (proctored) obtained a mean score of 50.66 for the pre-test and 50.37 for the post-test. The statistical comparison of the mean pre-test scores for the unproctored and the proctored sessions respectively by means of the dependent t-test, showed a statistically significant difference between these mean scores (t=-2.920; p=.006). The statistical comparison of the mean post-test scores for the unproctored and the proctored sessions respectively showed a statistically significant difference between these mean scores (t=-4.247, p=.000). Thus for the unproctored (UP) group, LPCAT pre-test results as well as LPCAT post-test results improved significantly in the second (proctored) test session. The above results are identical for the comparison of the sequence of testing, since for the unproctored (UP) group test session one was the unproctored one and test session two was the proctored one.

For the proctored (PU) group during the first test session (proctored) mean scores of 47.91 for the pre-test and 48.48 for the post-test was obtained. The same group during the second test session (unproctored) obtained mean scores of 50.27 for the pre-test and 50.50 for the post test. The statistical comparison of mean pre-test scores for the unproctored and proctored sessions respectively by means of the dependent t-test showed a statistically significant difference between these mean scores (t=2.983, p=.005). The effect size was .29 – indicating moderate practical effect. The statistical comparison of the mean post-test scores for the unproctored and the proctored sessions respectively showed a statistically significant difference between these mean scores (t=3.669, p=.001) with a moderate practical effect size of .24. Thus for the proctored (PU) group, LPCAT pre-test results as well as LPCAT post-test results differed statistically significantly. Closer inspection shows higher mean scores for the unproctored sessions than for the proctored sessions for the “proctored” (PU) group – which is the opposite of the results of the unproctored (UP) group). However, when sequence of testing is taken into account, the results confirm that of the unproctored”

(UP) group in that the mean scores of the second test session are consistently statistically significantly higher than mean scores of the first test session.

Results for both sub-groups showed statistically significant differences in the mean LPCAT pre-test and LPCAT post-test scores for both the unproctored and proctored test sessions as well as for the first and the second test sessions. Therefore $H1_0$ can be rejected for the LPCAT pre-test and post-test results for the sub-groups with groups based on mode (or sequence) of administration. Comparison of the mean scores for the unproctored and proctored test sessions were mixed, however, in terms of which session yielded the higher results.

Comparison of the mean scores with regard to the sequence of administration, showed statistically significant differences between the mean scores for the first and second test sessions. Therefore $H3_0$ can be rejected for the LPCAT pre-test and post-test results for the sub-groups with groups based on mode (or sequence) of administration. Furthermore, with regard to the sequence of testing, the mean scores for the second test sessions were consistently (and statistically significantly) higher than the mean scores for the first test sessions.

Table 6

Total group CPCAT descriptives and dependent t-test comparisons for the mode of administration and sequence effect

a) CPCAT Fields

CPCAT Fields	Mode of administration								Sequence effect							
	Unproctored			Proctored			t	Sig. 2-tailed	First test session			Second test session			t	Sig. (2-tailed)
	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
Law	81	49.34	23.321	81	55.63	22.953	-2.639	**.010**	81	52.31	22.353	81	52.65	24.317	-.136	.892
Business	81	48.01	26.327	81	52.41	24.571	-1.651	.103	81	48.36	25.580	81	52.05	25.406	-1.377	.172
Science	81	42.61	24.592	81	43.44	24.463	-.335	.739	81	41.51	23.819	81	44.54	25.132	-1.225	.224
Art	81	47.84	26.406	81	46.34	29.108	.567	.573	81	44.27	27.520	81	49.91	27.790	-2.191	**.031**
IT	81	58.67	27.030	81	61.47	25.577	-.968	.336	81	58.78	27.331	81	61.36	25.266	-.893	.375
Numeric	81	46.19	24.015	81	50.57	24.652	-1.797	.076	81	49.04	24.223	81	47.72	24.627	.534	.595
Language	81	49.09	24.515	81	55.88	23.784	-2.618	**.011**	81	52.27	24.559	81	52.70	24.223	-.160	.873
Sport	81	52.36	29.735	81	51.10	29.815	.621	.536	81	49.75	30.841	81	53.70	28.546	-1.981	**.**051
Tourism	81	54.98	25.603	81	56.42	26.619	-.574	.568	81	56.51	25.668	81	54.89	26.551	.648	.519
Technical	81	50.19	23.149	81	49.06	24.810	.473	.637	81	48.53	23.486	81	50.71	24.455	-.918	.361
Historical	81	41.36	24.248	81	46.71	25.399	-2.056	**.043**	81	43.83	24.296	81	44.24	25.636	-.156	.876
Agriculture	81	41.74	26.950	81	43.32	26.729	-.613	.541	81	39.80	26.265	81	45.26	27.148	-2.186	**.032**
Conservation	81	46.59	26.726	81	49.17	30.488	-.971	.335	81	45.73	28.355	81	50.03	28.875	-1.639	.105
Teaching	81	57.56	24.356	81	61.79	21.828	-1.724	.089	81	59.24	22.131	81	60.11	24.260	-.346	.730
Medical	81	47.44	25.853	81	47.39	24.519	.017	.987	81	46.42	25.599	81	48.41	24.743	-.719	.747
Security	81	46.59	21.348	81	49.44	23.001	-1.383	.170	81	46.53	22.560	81	49.51	21.806	-1.444	.153

b) CPCAT Activities

CPCAT Activities	Mode of administration								Sequence effect							
	Unproctored			Proctored			t	Sig. (2-tailed)	First test session			Second test session			t	Sig. (2-tailed)
	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
Managing	81	63.26	22.871	81	67.02	20.360	-1.551	.125	81	63.89	22.800	81	66.39	20.537	-1.021	.310
Service	81	68.35	24.984	81	73.44	23.416	-1.927	.058	81	72.56	25.461	81	69.23	23.060	1.245	.217
Precision	81	64.85	24.387	81	69.80	21.217	-1.721	.089	81	66.08	25.757	81	68.56	19.768	-.852	.397
Administration	81	58.94	21.612	81	60.86	20.534	-.844	.401	81	59.89	21.398	81	59.91	20.802	-.007	.995
Holistical	81	53.77	21.084	81	57.28	21.379	-1.432	.156	81	55.51	20.974	81	55.54	21.633	-.012	.990
Autonomy	81	61.47	23.981	81	65.57	21.330	-1.383	.170	81	62.11	24.059	81	64.92	21.349	-.940	.350
Entrepreneurial	81	55.69	28.512	81	58.75	25.275	-1.042	.300	81	56.71	29.341	81	57.73	24.394	-.345	.731
Practical	81	69.29	23.844	81	72.52	22.084	-1.203	.232	81	71.53	23.251	81	70.28	22.807	.463	.645
Creativity	81	64.29	21.591	81	70.48	21.440	-2.666	**.009**	81	68.94	21.606	81	65.83	21.761	1.294	.199
Challenge	81	62.93	22.821	81	64.89	21.491	-.761	.449	81	63.95	23.818	81	63.87	20.429	.030	.976
Public speaking	81	55.35	24.594	81	56.10	26.450	-.283	.778	81	56.50	24.441	81	54.95	26.573	.591	.556
Task Variety	81	66.48	23.096	81	69.14	21.366	-1.013	.314	81	68.67	22.806	81	66.94	21.723	.657	.513

c) CPCAT Environment

CPCAT Environment	Mode of administration								Sequence effect							
	Unproctored			Proctored			t	Sig. (2-tailed)	First test session			Second test session			t	Sig. (2-tailed)
	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
People	81	67.59	26.432	81	71.19	22.986	-1.154	.252	81	70.08	26.733	81	68.70	22.758	.438	.663
Thing	81	46.59	23.536	81	44.92	23.551	.619	.537	81	43.61	24.342	81	47.90	22.542	-1.617	.110
Informal	81	53.60	26.993	81	60.40	26.234	-2.206	**.030**	81	53.81	29.413	81	60.19	23.550	-2.058	**.043**
Formal	81	68.83	23.173	81	72.64	23.182	-1.441	.154	81	71.19	23.611	81	70.28	22.887	.340	.735
Indoors	81	53.61	25.523	81	56.48	22.046	-1.268	.208	81	53.61	25.594	81	56.48	21.962	-1.268	.208
Outdoors	81	55.73	26.142	81	56.74	24.202	-.444	.658	81	54.32	26.186	81	58.15	24.011	-1.697	.094

Interpretation of CPCAT results

Table 6 indicates results of the total group for both the mode of administration and sequence effect.

For mode of administration statistical significant results were obtained for five out of 34 dimensions. Three CPCAT fields showed statistically significant differences between the mean scores obtained with the proctored and unproctored administration: Law with a mean score of 49.34 for unproctored mode of administration and a mean of 55.63 for proctored mode of administration, (t=-2.639, p=.010) and effect size d=.27 indicating moderate effect; Language with a mean score of 49.09 during unproctored mode of administration and a mean score of 55.88 during proctored mode of administration (t=-2.618, p=.011) and effect size d=.28 indicating moderate effect; Historical field with a mean score of 41.36 during unproctored mode of administration and a mean score of 46.71 during proctored mode of administration (t=-2.056, p=.043) and effect size d=.22 indicating between low and medium effect or practical significance. One CPCAT activity showed a statistically significant difference between the mean scores obtained with the proctored and unproctored administration, namely creativity with a mean score of 64.29 during the unproctored mode of administration and a mean score of 70.48 during the proctored mode of administration (t=-2.666, p=.009) and effect size .29 indicating moderate effect or practical significance. One of the environment sub-fields, informal, also showed as statistically significant difference between the mean scores of the proctored and unproctored administration with a mean score of 53.60 during the unproctored administration and a mean score of 60.40 during the proctored test administration (t=-2.206, p=.030) and effect size d=.26 indicating moderate effect or practical significance. For each of the abovementioned five dimensions, the mean score for the proctored session was higher than that of the unproctored session. Based on the above results despite the fact that five out of 34 dimensions on the CPCAT showed statistically significant differences between the mean scores obtained with the proctored and unproctored test administration respectively, the null hypothesis ($H2_0$) cannot be rejected as the majority of the dimensions did not show statistically significant

differences between the mean scores obtained with the proctored and unproctored test administration respectively.

Furthermore table 6 indicates that for the total group results the sequence effect of three out of 34 CPCAT sub-dimensions showed statistically significant differences between the mean scores obtained during the first and second test administration respectively. With regard to the fields, for art and agriculture statistically significant differences in the mean scores were shown. For art, the mean score for the first test session was 44.27 while the mean score obtained in the second test session was 49.91 (t=-2.191, p=.031) and an effect size (d=.20) indicating between low and moderate effect size or practical significance. The field of agriculture showed a mean score of 39.80 in the first test session and a mean score of 45.26 in the second test session (t=-2.186, p=.032) with an effect size (d=.20) indicating between low and moderate effect or practical significance. One of the environment sub-dimensions, informal, also showed a statistically significant difference between the mean scores obtained in the first and second test sessions – a mean of 53.81 for the first test session and a mean of 60.19 for the second session (t=-2.085, p=.043) and an effect size (d=.24) indicating between low and moderate effect or practical significance. Based on the above results although some statistically significant differences between the mean scores of the first and second test sessions were obtained, for the majority of CPCAT sub-dimensions, no statistically significant differences were found between the mean scores obtained with the first and second test sessions respectively. $H4_0$ is therefore not rejected for the majority of the CPCAT results.

Table 7

CPCAT descriptives and dependent t-test comparisons for sub groups

a) CPCAT fields

CPCAT Fields	UP Group Unproctored			UP Group Proctored			t	Sig. (2-tailed)	PU Group Unproctored			PU Group Proctored			t	Sig. (2-tailed)
	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
Law	36	50.00	22.772	36	57.47	24.175	-1.831	.076	45	48.81	23.994	45	54.17	22.092	-1.895	.065
Business	36	43.09	27.502	36	52.19	26.309	-1.979	.056	45	51.94	24.958	45	52.58	23.389	-.212	.833
Science	36	41.60	24.922	36	45.94	26.089	-1.35	.264	45	43.42	24.578	45	41.44	23.182	.606	.548
Art	36	44.44	25.670	36	49.10	29.163	-1.258	.217	45	50.56	26.956	45	44.14	29.203	1.787	.081
IT	36	54.90	27.097	36	60.94	23.441	-1.302	.201	45	61.69	26.894	45	61.89	27.421	-.054	.958
Numeric	36	43.51	23.749	36	46.94	25.382	-.870	.390	45	48.33	24.276	45	53.47	23.939	-1.666	.103
Language	36	47.71	24.854	36	55.83	23.886	-1.955	.059	45	50.19	24.464	45	23.97	3.574	-1.728	.091
Sport	36	52.05	33.282	36	55.07	30.757	-.899	.375	45	52.61	26.951	45	47.92	28.991	1.949	.058
Tourism	36	56.49	26.053	36	56.28	28.147	.051	.960	45	53.78	25.468	45	56.53	25.652	-.889	.379
Technical	36	48.68	22.389	36	49.86	25.421	-.322	.749	45	51.39	23.922	45	48.42	24.579	.949	.348
Historical	36	38.13	24.674	36	44.62	28.038	-1.418	.165	45	43.944	23.862	45	48.39	23.261	-1.496	.142
Agriculture	36	39.72	28.937	36	47.64	29.309	-1.683	.101	45	43.36	25.464	45	39.86	24.253	1.411	.156
Conservation	36	44.24	27.138	36	51.98	31.825	-1.828	.076	45	48.47	26.546	45	46.92	29.543	.474	.638
Teaching	36	56.15	24.163	36	61.88	23.898	-1.658	.106	45	58.69	24.723	45	61.72	20.297	-.873	.387
Medical	36	48.47	28.811	36	50.66	26.352	-.450	.655	45	46.61	23.523	45	44.78	22.912	.578	.566
Security	36	45.31	20.440	36	51.88	21.344	-2.046	**.048**	45	47.61	22.223	45	47.50	24.307	.042	.967

b) CPCAT Activities

	UP Group Unproctored			UP Group Proctored			t	Sig. (2-tailed)	PU Group Unproctored			PU Group Proctored			t	Sig. (2-tailed)
CPCAT Activities	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
Managing	36	62.08	25.528	36	69.13	20.217	-1.633	.111	45	64.19	20.751	45	65.33	20.542	-.428	.671
Service	36	75.69	25.923	36	77.67	20.702	-.508	.615	45	62.47	22.824	45	70.06	25.092	-2.113	**.040**
Precision	36	63.37	29.089	36	71.74	19.115	-1.726	.093	45	66.03	20.126	45	68.25	22.855	-.648	.520
Administration	36	58.09	20.425	36	60.28	18.432	-.578	.567	45	59.61	22.723	45	61.33	22.267	-.610	.545
Holistic	36	54.76	22.228	36	58.75	23.029	-1.066	.294	45	52.97	20.342	45	56.11	20.149	-.955	.345
Autonomy	36	57.88	25.951	36	65.66	20.580	-1.526	.136	45	64.33	22.159	45	65.50	22.143	-.339	.736
Entrepreneurial	36	50.97	33.46	36	55.56	25.580	-.975	.336	45	59.47	23.547	45	61.31	25.020	-.490	.626
Practical	36	69.58	25.516	36	71.81	23.158	-.508	.347	45	69.06	22.709	45	73.08	21.434	-1.198	.237
Creativity	36	63.61	23.312	36	67.08	23.628	-.953	.347	45	64.83	20.362	45	73.19	19.357	-2.797	**.008**
Challenge	36	62.19	24.512	36	64.31	19.106	-.516	.609	45	63.53	21.636	45	65.36	23.429	-.553	.583
Public speaking	36	53.89	24.622	36	52.99	28.889	.198	.844	45	56.53	24.786	45	58.58	24.369	-.681	.499
Task Variety	36	67.60	24.994	36	68.65	21.930	-.285	.778	45	65.58	21.706	45	69.53	21.145	-1.059	.259

c) CPCAT Environments

	UP Group Unproctored			UP Group Proctored			t	Sig. (2-tailed)	PU Unproctored			PU Group Proctored			t	Sig. (2-tailed)
CPCAT Environ	N	M	SD	N	M	SD			N	M	SD	N	M	SD		
People	36	66.77	31.476	36	69.27	24.049	-.479	.635	45	68.25	21.935	45	72.22	22.253	-1.178	.245
Thing	36	44.41	23.453	36	47.36	21.299	-.638	.528	45	48.33	23.720	45	42.97	25.276	1.752	.087
Informal	36	51.35	28.824	36	66.18	19.388	-2.907	**.006**	45	55.39	25.621	45	55.78	30.052	-.110	.913
Formal	36	67.82	23.654	36	71.08	23.026	-.855	.399	45	69.64	23.016	45	73.89	23.489	-1.152	.256
Indoors	36	46.77	27.664	36	53.23	21.117	-1.819	.077	45	59.08	22.510	45	59.08	22.658	.000	1.000
Outdoors	36	57.08	27.579	36	62.53	22.003	-1.524	.137	45	54.64	25.195	45	52.11	25.109	.871	.388

Table 7 indicates the results for the unproctored (UP) and proctored (PU) subgroups respectively for comparison of the mean scores obtained during the proctored and unproctored test sessions. To avoid duplication for the large number of CPCAT test scores, the results – which are identical for the comparison of sequence of testing (see earlier results for LPCAT in Table 5) – a separate table for the sequence results is not presented.

For the unproctored (UP) group, for only two of the 34 dimensions of the CPCAT were the differences between the mean scores obtained with the proctored and unproctored administration respectively statistically significantly different. These were the field dimension security, where a mean score of 45.31 for the unproctored administration and a mean of 51.88 for the proctored administration was obtained (t=-2.046, p=.048) and an effect size score (d=.031) showing a moderate effect or practical significance. The second sub-dimension for the unproctored (UP) group where a statistically significant difference was shown between the mean scores of the proctored and unproctored administration was the environment sub-dimension informal where a mean score of 51.35 for the unproctored administration and a mean score of 66.18 for the proctored administration was found (t=-2.907, p=.006) with an effect size (d=.62) shows a large effect or practical significance.

For the proctored (UP) group, only two activity sub-dimensions showed a statistically significant difference between the mean scores obtained with the unproctored and proctored test administration. These are service for which the mean score of 62.47 for the unproctored and a mean score of 70.06 for the proctored sessions was found (t=-2.113, p=.040) with an effect size (d=.32) showing moderate effect or practical significance; and creativity with a mean score of 64.83 for the unproctored and a mean score of 73.19 for the proctored test sessions respectively (t=-2.797, p=.008) and an effect size (d=.42) showing between moderate and large effect or practical significance.

For both the unproctored (UP) and the proctored (PU) sub-groups, where there were statistically significant differences between the mean scores obtained in the proctored

and unproctored test sessions respectively, the mean score obtained in the unproctored session was lower than that of the mean score obtained in the proctored session. With regard to the sequence of administration, the results were mixed. For the unproctored (UP) group, the mean scores obtained in the first (unproctored) administration were lower than the mean scores obtained in the second (proctored) administration for those two dimensions (security and informal) where statistically significant differences between the mean scores were found. For the proctored (PU) group, however, the mean scores obtained in the first (proctored) administration were higher than the mean scores obtained in the second (unproctored) administration for those two dimensions (service and creativity) where statistically significant differences between the mean scores were found. Despite some dimensions on which statistically significant differences between the mean scores were found, these were in the minority (only two or three of the 34 sub-dimensions), so $H2_0$ (mode of administration) and $H4_0$ (sequence of administration) could also not be rejected for the unproctored (UP) or for the proctored (PU) sub-groups.

DISCUSSION

The aim of this study was to investigate the effect of mode of administration on computerised assessment results for LPCAT and CPCAT. Within-participants groups were used to compare similarities of the responses thereby addressing limitations of previous studies when independent groups were used (Salgado & Moscoso, 2003). Joubert and Kriek (2009) suggested that high stakes selection settings can add value as opposed to laboratory settings, this study entailed a cognitive test related to high stakes testing and a test setting *in vivo*.

The study was carried out to address the need to investigate a cognitive (learning potential) test and non-cognitive (interest) test equivalence in the field of UIT and to compare results from the same participants for the two modes of administration. No study has been conducted on LPCAT and CPCAT for the effect of the mode of administration on computerised assessment results. Identifying possible effects on test results when the unproctored or online mode of administration is used could assist the

test provider and practitioners in making informed decisions regarding online use during high stakes testing and for general test development. Also identifying possible effects on test results could assist with the interpretation of test results that have been obtained via a different mode of administration. This could also help practitioners to deal with the practical challenges related to assessments for selection and placement as well as training and development purposes in the workplace.

Joubert and Kriek (2009) provided evidence for comparable reliabilities, means and standard deviations between the different modes of administration for a personality test. Coyne et al. (2005) indicated that comparable results for the ICES interest was promising showing equivalence between supervised paper and unsupervised online mode. The need for research on cognitive tests was discussed (Lievens & Burke, 2011) and it was recommended that the validity of unproctored and proctored scores should be scrutinised for cognitive measures as findings suggested that unproctored scores were higher than proctored scores for numerical and verbal tests (Lievens & Burke, 2011). The LPCAT did not yield statistical significant differences for mode of administration but rather for sequence effect and it was indicated that improvement in second test sessions could imply that memory or learning played a role. In this study statistically significant differences in the CPCAT could be explained by the verbal reasoning and understanding of vocabulary or terminology of the statements in the test that could have affected results. An additional explanation of the interest self rating questionnaire could mean that participants change their ratings based on certain feelings or perceptions on the day of testing. However, no explanations were clear for the differences in the sub dimensions of the CPCAT results and it should be kept in mind that only a very small group of sub-dimensions yielded statistically significant differences between the mean scores.

Statistically significant differences of mean LPCAT scores

For hypothesis 1 the null hypothesis stated that there is no statistically significant difference between the mean scores obtained with proctored and unproctored test administration for cognitive (LPCAT) results. There were no statistically significant

differences between the mean scores for the LPCAT results for the total group when proctored and unproctored results were compared. This was the case for both the pre-test and the post-test scores - thus the null hypothesis ($H1_0$) is not rejected.

Results for sequence effect for the total group irrespective of mode of administration showed statistically significant results for the LPCAT pre-test and LPCAT post-test which indicates that the mean scores for the second test session were consistently higher than the mean scores for the first test session – indicating that results for the total group improved during the second test session irrespective of mode of administration.

With regard to the sub groups, mixed results were shown for the mode of test administration while for sequence of administration mean scores for the first test session were consistently (and statistically significantly) lower than the mean scores for the second test session. It could be inferred that possible anxiety levels associated with cognitive tests in general could be present, and may be lower on the second administration when the individual is more familiar with the testing procedure.

For the proctored (PU) group mean scores also improved in the second test session (unproctored) and this could suggest that certain familiarity or confidence after having had support and supervision the first time around, during the second test session could have produced less anxiety.

It should be noted that where group mean scores improved, the improvement in terms of NQF levels related to t-scores was not enough to improve towards a next NQF level on LPCAT. Nevertheless mean scores improved significantly during the second test session for each group every time.

Whilst the computer adaptive nature of the test implies that participants would not necessarily have received the same items during the first and second test sessions the improvement in scores provides further support for Vygotsky's zone of proximal development and the potential to learn and improve – which is at the heart of the measurement of learning potential. Also it could imply that whilst test items could have

differed from the initial testing due to the adaptive nature of the test, the possibility of participants remembering the information provided in the training part of the test could not be ruled out.

The significant differences in scores showed that sequence of testing played a role in mean score differences, however mode of administration did not. It could be suggested that, for LPCAT results in the initial test session could differ significantly from results obtained during a second (verification) testing – although practically the differences generally do not affect the NQF level interpretation.

Statistically significant differences of mean CPCAT scores

For the total group for mode of administration statistically significant differences between the mean scores obtained with the unproctored compared to the proctored test sessions results were obtained for five CPCAT sub dimensions namely law, language and history in the CPCAT field dimensions, for creativity in the CPCAT activity dimension and informal in the CPCAT environment dimension. At first glance it would seem that the null hypothesis could partially be rejected, however, five out of 34 dimensions are not enough to fully reject the null hypothesis ($H2_0$) and was therefore not rejected. Upon further investigation it was clear that the mean scores for the five sub-dimensions where statistically significant differences between the unproctored and proctored session results were found were consistently higher for the proctored session compared to the unproctored session.

For the total group for sequence effect significant differences between the mean scores obtained with the first compared to the second test session were obtained for two CPCAT field sub-dimensions namely art and agriculture. Statistically significant differences between the mean scores of the first and second test session were also found for the informal environment sub-dimension with a higher mean score in the second than in the first test session.

For the sub groups the unproctored (UP) group statistically significant differences between the mean scores were found for one field (security) and one environment

(informal) sub-dimension when the proctored and unproctored test session results were compared. In both cases the unproctored session mean was lower than the proctored session mean. For the proctored (PU) group, statistically significant differences between the mean scores were found for two field sub-dimensions (service and creativity) only. In both cases the unproctored session mean scores were lower than the proctored session mean scores.

Unlike the cognitive measure (LPCAT) in this study where sequence effect consistently showed statistically significantly higher mean scores for the second test sessions compared to the mean scores of the first test sessions, for the CPCAT results with regard to sequence of testing was mixed. However, with regard to the mode of administration for the CPCAT, for those dimensions where statistically significant differences between the mean scores were found, the mean scores for the unproctored test session were consistently lower than the mean scores for the proctored test session. Whereas LPCAT measures non-verbal figural reasoning, CPCAT required statements in English to be read and understood which could indicate that levels of verbal competency is necessary to understand the statements fully. The fact that at least 45% of participants did not have grade 12 and that the majority of participants first language was not English could have impacted on understanding of statements on CPCAT. Participants rated the abovementioned dimensions higher during proctored (supervised) sessions indicating that where assistance was available confirmation of meanings or words were provided, possibly leading to higher ratings.

Reported computer competency

Computer competency was not the main focus of the study but this information provided additional information. The majority of participants reported their computer competency as being poor. During the two tests different keystrokes were used such as the space bar and enter bar for LPCAT, but for CPCAT all keys on the key board for the typing of biographical detail and the touchpad to rate preferences on a sliding scale. The significant differences noted for LPCAT in the second test sessions could indicate that unfamiliarity in the first test session could have caused some degree of anxiety. Also

for CPCAT where typing of biographical information using the key board was needed the unproctored sessions could have caused some degree of anxiety.

CONCLUSIONS: PRACTICAL IMPLICATIONS

The results of the study are somewhat mixed. Based on this study it is concluded that since the cognitive test results did not differ significantly for mode of administration future use of internet-based testing for LPCAT could be considered. Unproctored administration for LPCAT is not planned as it is a cognitive test where the identity of those being tested and the security of test items are of critical importance to maintain the integrity of the instrument. However, internet-based proctored administration could be considered. Also the results should be interpreted with caution as cheating and other forms of test security were not researched. There is clear evidence that rather than mode of administration sequence effect impacted on the cognitive measure's computerised test results. Statistically significant findings for LPCAT for sequence effect implies that where test results improved in a second test session it is inferred that learning can take place during a testing session and memory could have resulted in improvement of second test mean scores. Several factors could have limited recognition or overexposure of test items as indicated by Tippins (2009) of test content for the LPCAT such as the dynamic assessment methods, item response theory, adaptive techniques and rest periods of two or more weeks between sessions possibly restricted recognition of test content. However the basic principles taught to participants during the training session of the test could have been remembered by participants as pre-test (current ability) scores of the second test session improved significantly. This is also in line with the principle of dynamic (test-train-retest) assessment of learning potential, based on Vygotsky's concept of the zone of proximal development.

Further indications are that the cognitive measure (LPCAT) a high stakes test which was used in the context for development purposes, along with familiarisation of computer testing could provide some explanation for the participants' improved scores

because of the learning opportunity. The LPCAT cognitive test results on sequence effect and improvement in a second test session is explained that even the testing session could have lead to learning taking place. However, in practice it would seem that LPCAT verification testing could imply that participants based on their results in a first test session may be rejected falsely in cases of high stakes testing. Keeping in mind that the mean scores did not improve to higher NQF levels, nevertheless when a group of individual scores are compared certain first test scores could lead to applicant rejection, whereas second tests score may differ.

CPCAT results reveal that self-rating questionnaires could imply that the possible emotional state on the particular day may impact on ratings and should be further researched. Furthermore where the questions entail statements and English sentences a degree of understanding of terminology and verbal competency is needed. Whereas the interest self-rating measure (CPCAT) was possibly experienced as a low stakes measure the possible difficulty of reading and understanding terminology of statements could have impacted on the results when assistance was not available.

The findings indicate that test familiarity and anxiety may affect test results. The mode of administration for a non-cognitive interest test may partially have an impact on test results. An initial unsupervised mode may still be experienced as stressful by candidates. Keeping in mind that the majority of the participants were Afrikaans speaking with many Xhosa speaking participants it is inferred that levels of language proficiency needed be considered for test use as language and terminology may be a barrier for test takers.

Limitations of the study

One limitation of the study was the fact that specific events at the company prevented adherence to the true or quasi-experimental design prerequisites. In essence the study's limitation is that technically it did not adhere to the stipulations for a quasi-experimental design. Due to a protected strike in the time when testing commenced, employees were not pre-assigned to groups as initially planned. Two thirds of the employees were union members and mostly within the A1-B3 Patterson grading

bracket. The study was undertaken as the events of the strike unfolded and as employees attended work during and after the strike. Also due to the long hours and the shifts which individuals work in the hospitality industry, participants were not randomly assigned at any point to control groups, but attendance was based on convenience for managers and employees based on workload and availability. Because of the non-representivity of the sample of convenience of participants in this study, the statistical significant scores could not be generalised to the larger population.

In addition it should be noted that at least two weeks rest periods in between testing were planned, yet in practice attendance of employees was based on availability. The fact that participants did not all have the same number of weeks rest period between sessions reflects that the study was conducted *in vivo* and therefore represents an indication of real events and not a laboratory setting. Whilst the period between the two test sessions was not too short for participants to have remembered or recognised test items, the longer test periods could have suggested some changes in the participants' lives and possibly their learning experiences or interests.

The small convenience sample (N=82 for LPCAT and N=81 for CPCAT) impacts on the generalisability of the findings. The approach to using proctored versus unproctored or internet-based testing is however still relevant in terms of current debates and future challenges. The findings are valuable in relation to researching decisions when a cognitive or non-cognitive computer test is used in different modes of administration. Concerns in the literature which related to cheating in unproctored mode for example were not investigated in this study, however, such possible scenarios cannot be excluded.

Finally the LPCAT was conducted in a simulation manner to simulate UIT but where the CPCAT testing entailed actual connection failures and other challenges related to internet testing, the LPCAT was not online, only an online simulation. Also whilst the unproctored session was unsupervised the researcher had logged test takers into the LPCAT and for that reason the requirements for open or controlled mode might not reflect all the challenges experienced with unproctored testing accurately.

Recommendations for future research

The limitations in this study were many, nevertheless future research could be considered as discussed below. Whilst comments such as those of Beaty et al. (2011) indicated a move from the debate whether or not to proctor, findings in this study keeping in mind the limited sample size indicated that for CAT tests mode of administration did not seem to affect test results. The need identified by Salgado and Moscoso (2003) for research related to internet-based testing and specifically for cognitive tests was to some extent addressed in this study. Whilst opinions such as that of Pearlman (2009) imply that UIT can promote production of poorly developed or unvalidated tests the strong interests of organisations to use tests in UIT form cannot be denied (Tippins et al., 2006).

- Further studies could be conducted in the g-factor and learning potential domains to identify possible knowledge retention of such tests and ability to learn from the test experience could add value.
- The unproctored mode did involve assisting the (UP) group with log on of LPCAT based on potential computer literacy limitations of the test takers, which meant that the unproctored mode was not fully in open or controlled mode and could be kept in mind for future studies.
- Larger sample groups in a similar study could be beneficial.
- A longitudinal study of CPCAT differences could be valuable. Personality which relates to inherent traits is probably less likely to change over time than one's preferred interests. The few statistically significant differences found for CPCAT (a self rating interest) measure provides possible explanations of a certain verbal competency needed to understand the questions and possible fluctuations in participants based on emotional states on the day.
- Further studies in the true experimental domain including within-participants on test-re-test validity for the CAT and IRT tests in online mode are recommended not only for personality tests but also for cognitive tests.
- Use of practice tests for cognitive measures to increase familiarity with test processes and similar content to lower test anxiety could be considered.

REFERENCES

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, *8*(4) 261-274.

Bartram, D. (2004). Assessment in organisations. *Applied Psychology: an International Review*. *53*(2), 237-259.

Bartram, D. (2006). The internationalization of testing and new models of test delivery on the internet. *International Journal of Testing*, *6*(2) 121-131.

Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ31i scores. *International Journal of Selection and Assessment*, *12*(3) 278-284.

Beaty, J.C., Dawson, C.R., Fallaw, S.S &Kantrowitz, T.M. (2009). Recovering the scientist– practitioner model: How IO's should respond to unproctored internet testing. *Industrial and Organizational Psychology*, 2, 58-63.

Beaty, J.C., Nye, C.D., Borneman, M.J., Kantrowitz, T.M., Drasgow, F., &Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive as job performance? *International Journal of Selection and Assessment*, *19*(1), 1-10.

Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, 2, 35-38.

Christensen, L.B. (2001). *Experimental Methodology*. (8th ed.). Massachusetts, Pearson Education.

Coyne, I., &Bartram, D. (2006). Design and development of the ITC guidelines on computer based and internet delivered testing. *International Journal of Testing*, *6*(2), 133 -142.

Coyne, I., Warszta, T., Beadle, S., & Sheehan, N. (2005). The impact of mode of administration on the equivalence of a test battery: A quasi-experimental design. *International Journal of Selection and Assessment*, *3*(13), 220-224.

Davies, C., Foxcroft, C., Griessel, L., & Tredoux, N. (2009). Computer Based and Internet-delivered assessment. In C. Foxcroft, & G. Roodt (Eds.), *An introduction to psychological assessment* (3rd ed., pp. 185-200). Cape Town, Southern Africa: Oxford University Press:

De Beer, M. (2000). *Learning Potential Computerised Adaptive Test (LPCAT): User's Manual.*Pretoria:University of South Africa.

De Beer, M. (2004). Use of differential item functioning (DIF) analysis for bias analysis in test construction. *South African Journal of Psychology*, *30*(4), 52-58.

De Beer, M. (2005). Development of the Learning Potential Computerised Adaptive Test (LPCAT).*South African Journal of Psychology*, *35*(4), 717-747.

De Beer, M. (2011). Initial review of psychometric properties of a computerised career preference test for career guidance assessment. *Journal of Psychology in Africa*, *21*(2), 311-314.

De Beer, M. (2012). The Learning Potential Computerised Adaptive Test in South Africa. In S. Laher, &Cockroft, K. *Psychological assessment in South Africa:*

Research and applications 2000 – 2010. (pp.137-157)

Eid, M., &Diener, E. (2006).*Handbook of multimethod measurement in psychology*. Washington: United Book Press.

Employment Equity Act, 55 of 1998.*Government Gazette*, 400 (19370). Cape Town, South Africa, 19 October 1998. Government Printer

Field, A. (2005). *Discovering statistics using SPSS.* (2nd ed.). Oxford, The Alden Press.

Foster, D. F., (2010). Worldwide testing and test security issues: Ethical challenges and solutions. *Ethics & Behaviour*, *20*(3/4), 207-228.

Foxcroft, C.D., & Davies, C. (2006). Taking ownership of the ITC's guidelines for computer-based and internet-delivered testing: A South African Application, *International Journal of Testing*, *6*(2) 173-180.

Foxcroft, C., Paterson, H., le Roux, N., &Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis.* Pretoria, South Africa: Human Sciences Research Council.

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

Foxcroft, C., Roodt, G., & Abrahams, F. (2009).The practice of psychological assessment: controlling the use of measures, competing values, and ethical practice standards. In C. Foxcroft, & G. Roodt, (Eds.), *An introduction to psychological assessment (3rd ed., pp. 9-26*). Cape Town, South Africa: Oxford

University Press.

Gilmore, N. (2008). *The relationship between learning potential and job performance.* (Unpublished Masters thesis).University of South Africa.

Gregory,R,J. (2007). *Psychological testing: History, principles, applications* (5th ed.). Boston: Pearson Educational Inc.

Hense, R., Golden, J.H., &Burnett, J. (2009). Making a case for unproctored internet testing: Do the rewards outweigh the risks? *Industrial and Organizational Psychology*, 2, 20-23

IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

International Test Commission (2005).*International guidelines on computer-based and internet delivered testing.Version 2005*. [http://www.intestcom.org]

International Test Commission (2011).*International guidelines for quality control in scoring, test analysis, and reporting of test scores.*Version 2011. [http://www.intestcom.org]

Joubert, T., & Kriek, H.J. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *South African Journal of Industrial Psychology*, *35*(1), 78-88.

Lievens,F., & Burke, E. (2011). Dealing with the threats inherent in unproctored internet testing of cognitive ability: Results from a large scale operational test program. *Journal of Occupational and Organizational Psychology*, 84, 817-824.

Naglieri, J.A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist*, 59, 150-162.

Pearlman, K. (2009). Unproctored internet testing: Practical, legal, and ethical concerns. *Industrial and Organizational Psychology*, 2, 14-19.

Salgado, J.F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of Measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, *11*(2/3), 194-205.

Salkind, N.J. (2009). *Exploring research.*(7th ed.). Upper Saddle River, New Jersey: Pearson Education Incorporated.

Styles, I., & Andrich, D. (1993). Linking the standard and advanced forms of the Raven's Progressive Matrices in both the pencil-and paper and computer adaptive testing formats. *Educational and Psychological Measurement, 1993, 53* Doi:10.1177/0013164493053004004.

Tippins, N. T., Beaty, J., Drasgow F., Gibson W. M., Pearlman, K. Segall, D.O., & Shepherd, W.J. (2006).Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 189-225.

Tippins, N.T. (2009). Internet alternatives to traditional proctored testing: Where are we Now?*Industrial and Organizational Psychology*, 2, 2-10.

CHAPTER 4

CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

In this final chapter, the focus will be on the conclusions reached, based on this study. The limitations connected to the empirical research and literature reviews will be highlighted and recommendations will be discussed for practical implementations of the findings.

4.1 CONCLUSIONS RELATED TO THE LITERATURE REVIEW

In terms of the literature review the aim of the study was to discuss:

- previous research conducted in the past that was relevant to the topic of mode of test administration and psychological assessment results with reference to unproctored internet testing (UIT);
- core concepts relevant to this study such as proctored and unproctored modes of administration, computer-based testing, computerised adaptive testing, internet testing and the advantaged and disadvantages of computer- and internet-based testing; and
- information that can be used when having to decide on the mode of test administration.

In general these aims were achieved by conceptualising and discussing concepts and debates around the literature review.

4.1.1 Previous research on the topic of mode of test administration and UIT

In Chapter 2 it was reported that in the past several studies were focused on comparing paper-and-pencil tests equivalence to online or internet-based test results (Bartram & Brown, 2004; Coyne et al., 2005: Joubert & Kriek, 2009; Salgado & Moscoso, 2003). Research conducted in the field of internet-based testing, for example on personality questionnaires, provides much information related to the topic of unproctored testing (Bartram & Brown, 2004; Coyne et al., 2005; Joubert & Kriek, 2009; Salgado &

Moscoso, 2003). Literature searches provided more research articles in terms of comparable results for personality questionnaires with few findings on cognitive test results in UIT. Tippins et al. (2006) indicated the need for research on cognitive tests results when conducted either as computerised tests or by means of an alternative mode such as internet or online mode. Lievens and Burke (2011) indicated a shift from whether cognitive testing in UIT settings is feasible to how to deal with reliability and validity in such settings. By computing the level of change in individual scores for numeric and verbal tests, Lievens and Burke (2011) reported that unproctored scores were higher than proctored scores.

Conclusions drawn from the literature review for mode of administration are as follows.

Several authors seem to agree that the future use of online an internet-based testing will increase as technology expands (Bartram, 2000; Beaty et al.,2011; Gibby et al., 2009; Lievens & Harris, 2003; Tippins, 2009). Whilst indications are that the use of UIT can assist with cost savings related to HR and business practices, maintaining psychometric properties and validity with a focus on ethical and legal guidelines in the process remains important (Reynolds et al., 2009). It was highlighted that, although internet-based testing in the globalised context can merge business boundaries, adherence to international guidelines and legislation in South Africa is essential.

The debate around both the challenges and the benefits of UIT is ongoing as the risks involved are vast. There do not seem to be clear cut answers or foolproof practical solutions to issues such as test security, cheating, confidentiality, cultural sensitivity and fairness. Whilst options are being explored in terms of security plans and technological infrastructures, (Coyne & Bartram, 2006; Foster, 2010) good practice is essential.

4.1.2 Core concepts, disadvantages and advantages of computer-based and internet-based testing were clarified

Focus has shifted from the traditional paper-and-pencil methods towards computerised testing and internet-based testing (Bartram, 2000; Tippins et al., 2006; Vispoel, 2000). The benefits of computer testing include consistency of delivery (Tippins et al., 2006) while other advantages relate to administration procedures, standardised instructions,

quick and accurate scoring as well as report writing (Gregory, 2007). The challenges relating to computer testing could include systems problems with hardware and software, backup procedures and practitioners having to be trained in the computer program (Coyne & Bartram 2006). In the South African context lack of computer literacy and familiarity cannot be ignored (Foxcroft & Davies, 2006).

There is however a clear distinction between computer-based testing and internet testing. Internet-based testing implies accessibility of infrastructure or the medium through which test products are presented to the client (Bartram, 2006). The advantages of using internet assessments include time and cost effectiveness as well as better, faster and cheaper services (Naglieri et al., 2004). In addition the results are easily obtained and immediately available, with the added luxury for test takers of being able to complete tests in their own time and convenience. This could imply that test takers can complete tests whilst at work or home or after hours and in a relatively relaxed and known environment. However, the recent national census conducted in 2011 in South Africa indicated that internet access is still not available to the majority of South Africans at home or in work settings (StatsSA, 2011). From a business perspective Hense et al. (2009) found that in the financial sector internet assessment improved hiring efficiency dramatically.

Whereas computer-based testing in the past was mainly completed in a proctored and supervised environment, the internet facilitated certain modifications to computer assessment. Bartram (2000) and Coyne and Bartram (2006) explained that computer-based testing has been delivered for many years with the accredited user purchasing the materials and exercising direct control over the use of the computer-based test. The internet, however, changed this mode of test delivery (Coyne & Bartram, 2006). Beaty et al. (2009) indicated that an increasing number of organisations conduct psychometric testing, engaging in UIT. Naglieri et al. (2004) explained that the advantages of internet testing have become irrelevant when scores were used in ways that were not supported by the validity of the instruments. Whilst some argue that the benefits of UIT outweigh the risks of UIT and that the debate should not be around whether or not to proctor, there are still many areas of concern in UIT.

The concerns around UIT are too many to accept unproctored mode fully at this point in time (Foxcroft et al., 2004; Naglieri et al., 2004; Tippins, 2009) and further research to address specific questions is needed.

4.1.3 Information that can be used when having to decide on mode of test administration

Suggestions are that certain important guidelines should be adhered to when deciding on the mode of administration, namely international, national and legal guidelines (Foxcroft & Davies, 2006; ITC, 2005, 2011; Tippins, 2009).

Tippins' (2009, p. 3) summary of expert discussions provide guidelines relevant for UIT and the main points appear below.

- The nature of the test, namely cognitive or non-cognitive, plays an important part when having to make decisions related to UIT.
- High-stakes situations and UIT alone should never be accepted.
- Many question the benefit of UIT if second testing verification is needed.
- The ethics around UIT are questioned.

4.2 CONCLUSIONS RELATED TO THE EMIPRICAL STUDY

Salgado and Moscoso (2003) found comparable psychometric properties for the big five personality dimensions of IP/5F between the paper-and-pencil version and internet version of the test. Joubert and Kriek (2009) reported similar psychometric properties for the OPQ32i personality test in high stakes settings for paper-and-pencil and internet mode. Coyne et al. (2005) reported equivalence between paper-and-pencil and online testing modes for interest and personality on the ICES. However for cognitive numeric and verbal scales lack of equivalence were reported with indications that the psychometric properties of the numerical and verbal scales were affected by the computerised version of the tests. Arce-Ferrer and Guzmán (2009) found support for equivalence between paper-and-pencil and computer-based administrations for the Raven Standard Progressive Matrices test. In terms of this study the difference in

results on computer-based and internet-based delivered testing was not significant. Studies in the past mainly focused on comparing results between paper-and-pencil versions with internet-based versions of tests which include personality, interest and some cognitive tests. This study is unique in terms of comparing computer-based and computer adaptive tests (CAT) with the online version of the tests rather excluding paper-and-pencil test administration.

The study aimed to answer the research questions stated in the form of hypotheses:

$H1_0$: There is no statistically significant difference between the mean scores obtained with proctored and unproctored test administration respectively for cognitive (LPCAT) results. Based on the evidence from the current study, this null hypothesis was not rejected.

$H1_1$: There is a statistically significant difference between the mean scores obtained with proctored and unproctored test administration for cognitive (LPCAT) results. Based on the evidence from the current study, this alternative hypothesis was not accepted.

$H2_0$: There is no statistically significant difference between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results. Based on the evidence from the current study, this null hypothesis could not be rejected for all dimensions of the CPCAT – as only five of the 34 sub-dimensions showed statistically significant differences between the mean scores. This null hypothesis could therefore only be rejected for those five sub-dimensions.

$H2_1$: There is a statistically significant difference between the mean scores obtained with proctored and unproctored test administration respectively for non-cognitive (CPCAT) results. Statistically significant differences between the mean scores obtained with proctored and unproctored test administration respectively were only found for five of the 34 CPCAT dimensions. This hypothesis could therefore only be accepted for those particular dimensions.

The empirical findings lead to no statistically significant differences for LPCAT and the null hypothesis was not rejected. The results are similar to those of Arce-Ferrer and

Guzmán (2009) as the mean scores were in general equivalent for the mode of administration of a non-verbal figural reasoning test. For CPCAT statistical significant differences were found on five out of 34 sub dimensions form mode of administration however, since the majority of sub dimensions did not yield significant differences the null hypothesis was not rejected overall. The results are similar to that of Coyne et al. (2005) as the majority of sub dimensions for CPCAT, an interest measure, were not affected by mode of administration. The mean scores of both tests were generally comparable for mode of administration.

With regard to the sequence of administration, the results were mixed.

$H3_0$: There are no statistically significant differences between the mean scores of the first and second test sessions for cognitive (LPCAT) results. This null hypothesis was rejected, as there were statistically significant differences between the mean scores obtained in the first and second test sessions. In all cases, the means of the second test administration were all higher than the means of the first test administration. This was the case for both the pre-test and the post-test. This study is different from other studies as the indication that a computerised adaptive test and non-verbal figural reasoning test could have been affected by sequence effect and more specifically that results improved during second test sessions, indicates to the learning potential measured by the LPCAT that can improve and memory that can play a role in cognitive testing.

$H3_1$: There are statistically significant differences between the mean scores of the first and second test sessions for cognitive (LPCAT) results. In line with the above rejection of the null hypothesis, this alternative hypothesis was accepted based on the evidence obtained in the current study.

$H4_0$: There are no statistically significant differences between the mean scores of the first and second test sessions for non-cognitive (CPCAT) results. Since there were only three of the 34 sub-dimensions of the CPCAT on which statistically significant differences were found between the mean scores obtained with the first and second administration respectively, there was insufficient evidence to reject this hypothesis in

general – it could only be rejected for those three sub-dimensions on which statistically significant differences were found.

$H4_1$: There are statistically significant differences between the mean scores of the first and second test sessions for non-cognitive (CPCAT) results. In line with the above rejection of the null hypothesis is limited or only a partial rejection of the null hypothesis, this alternative hypothesis could also only be practically accepted – for only the three out of a possible 34 sub-dimensions – for which statistically significant differences between the mean scores for the first and second test administration were found.

In light of previous research and guidelines provided by Naglieri et al. (2004) and Tippins et al. (2006) concerns relevant to high stakes testing which often includes cognitive testing and the challenges related to UIT should carefully be considered by practitioners. In this study results showed no significant differences for mode of administration or the online and simulation of online settings on a cognitive and non-cognitive test.

4.2.1 The first aim was to determine if the mode of administration had an effect on LPCAT and CPCAT respectively

It was discussed in the research article (Chapter 3) that the quantitative research design was relevant and that primary data was analysed using the dependent t-test. Findings differed in terms of the cognitive and non-cognitive test results.

The following conclusions were drawn.

a) Conclusions about LPCAT

For LPCAT for the mode of administration no statistically significant differences between the mean scores obtained with unproctored and proctored mode respectively, were found. Therefore, the mean scores for mode of administration can be considered comparable.

b) Conclusions about CPCAT

Findings indicate that for the mode of administration statistical significant difference on a total of five dimensions were evident. For the other 29 dimensions mean scores could be considered comparable. Indications are that test mean scores for the 34 sub-dimensions were mostly not statistically significantly different and for this reason the null hypothesis for CPCAT was not rejected – it could only be rejected for the specific sub-dimensions for which statistically significant differences between the mean scores (for unproctored and proctored test administrations) were found.

In practice indications are that the interest questionnaire may be used unproctored and in UIT settings but levels of verbal competency may be necessary when test takers are tested.

4.2.2 The second aim was to determine whether sequence effect played a role in the study

The following conclusions were drawn.

For LPCAT sequence effect the total group and sub groups' scores for pre-tests and post-tests differed significantly and the first test sessions results were consistently lower than those of the second test sessions. Based on the sub groups (unproctored (UP) and proctored (PU) groups) at first glance it seemed that the unproctored (UP) group performed worse during the unproctored condition whereas the proctored (PU) group performed better in the unproctored session than in the proctored session. It became clear that both groups' mean scores improved with the second test session.

For LPCAT for the sequence of administration, statistically significant differences were found between the first and second test administrations for all scores. In all cases, mean scores on the second test administration were higher than the mean scores on the first test administration.

Findings suggest that when scores improved during the second test session regardless of mode of administration memory played a role and the learning effect and zone of proximal development between test sessions were possible.

For the CPCAT the total group yielded statistically significant differences for sequence effect on three sub dimensions of CPCAT. For the unproctored sub group two sub dimension presented a statistically significant difference between the mean scores of the first and second test sessions. For the proctored group two sub dimensions yielded statistically significant differences between the mean scores of the first and second test sessions. For the sub-groups, evidence indicated that mode of administration showed more consistency than sequence of administration for those sub-dimensions for which statistically significant differences between the mean scores were found.

4.2.3 Additional information

By obtaining self-ratings from participants on their computer competency and considering the fact that two tests were completed during every test session the following conclusions can be drawn.

- Indications were that computer competency, whilst not necessarily affecting test scores directly (as this was not formally investigated), could have possibly added to anxiety, specifically during the first test sessions of both tests.
- Whilst only the space bar and enter key was used to complete LPCAT, CPCAT involved typing biographical information by using the keyboard and selecting answers by using the touchpad. Considering that unproctored CPCAT sessions involved use of more keys on the keyboard than LPCAT, the fact that findings were significantly different could indicate that the use of computer keyboards could have added additional pressure during the test session.

Several observations noted during the study included the points set out below.

- Problems with internet connection for CPCAT caused problems with log on and saving of results.
- Some online CPCAT tests did not save based on candidates not following steps during the unproctored process which ultimately lowered the sample size as not all participants data could be used for statistical analysis.

- Whilst most groups were quiet, a few groups engaged in discussions during testing.
- Test anxiety could have been present when employees attended the first test session.
- Based on the observation and interaction of the researcher with the participants, many employees verbally reported excitement and familiarity with the computer experience when attending second sessions.

4.3 LIMITATIONS

The limitations of the literature review and empirical study will be discussed below.

4.3.1 Limitations of the literature review

Whilst the topic of proctored versus unproctored testing is relatively new, the literature in the past few years has increased but in general research on mode of administration and UIT settings is ongoing and not yet extensively researched In terms of South African registered tests the relevance of specifically unproctored or online testing is under-researched and limited.

4.3.2 Limitations of the empirical study

One limitation of the study was the fact that the research did not meet the requirements for quasi-experimental design. Due to a protected strike when the study had officially commenced as well as the long hours and shifts individuals work in the hospitality industry, participants were not randomly assigned at any point to control groups. Quite a few managers were new to the organisation and despite the clear explanation about the aim and requirements for participation in the project, some scepticism existed from management and employee side about the testing, possibly even aggravated by the circumstances during the strike. Due to unforeseen circumstances as a result the intended design namely the quasi-experimental design did not meet the true technical requirements for a quasi-experimental design. Also employees attended testing based on considerations of both the convenience of managers and the employees' work load

or availability. In addition the unproctored session did involve initial contact with the test takers due to LPCAT being a simulation of online assessment and for therefore might not have reflected all the challenges encountered by test takers when there is no assistance when the test takers completed the tests.

This meant that the time lapse between testing sessions varied and therefore not exactly the same amount of time between the two test sessions for each participant was possible. Whilst in some instances participants had approximately a three month gap between testing sessions, it is uncertain to what extent this would have affected the constructs learning potential and interest respectively. Furthermore, based on the CAT assessment of the LPCAT, participants possibly received different test items and the computer adaptive technique minimised the chance of exact same items being presented. However, the possibility that participants might have remembered the information provided in the training part of the first test session was possible. In addition as the computer tests were designed to use computerised adaptive programs, indications are that if a person had responded differently in any way from the first testing session, then the following items would be different from the initial test, thereby providing items at similar levels but with different content on LPCAT.

Preference for career could possibly change in a short period due to specific exposure to people and/or other events or information, yet in general, more than two week rest period between testing sessions would be long enough for employees to not recall statements. For CPCAT the dimension ratings on factors might vary as self-report questionnaires could imply that participants' response might not be accurate or similar from one test session to another test session, but dramatic differences are unlikely without specific explanations for this.

The fact that a small sample of convenience (N=82 and N=81) was used impacts on the power of the statistical analysis and the generalisability of the findings. Whilst no inferences can be drawn for the larger hospitality industry or the South African population, the possible effect on LPCAT or CPCAT results or validity when used online is relevant. The approach of using proctored versus unproctored or internet-based

testing is still relevant in terms of current debates and future challenges. The findings are valuable in terms of researching the possible effects when computer tests are used in different modes of administration. The findings indicate that for both tests mode of administration did not significantly affect test results.

Even though participants were asked to report on their computer competency, their computer competency was not tested, and could therefore have been a perceived competency. Where some participants had never worked on a computer previously, they were able to use the few allocated keys to complete tests; however the anxiety that computer literacy could have had on participants was not measured.

4.4 RECOMMENDATIONS

Based on the findings, limitations and conclusions of the study the following recommendations are made.

- In light of the literature review it is important for South African tests to be researched in terms of the mode of administration. Also the fairness in terms of recruiting and making decisions based on high stakes testing through computer and internet accessibility is a challenge that remains in South African society.
- It is suggested that test providers in South Africa investigate the effect that the mode of administration could have in particular on psychological tests. More specifically this could be applied to those cognitive tests or tests that are used in situations where high stakes testing is imminent. Internet usage is growing in Africa and it would be beneficial for the practitioner to know how or to what extent a test is affected when the unproctored or online mode of administration is used. It would be important for practitioners to know how test scores that were taken in unproctored or non-traditional settings should be interpreted. Therefore, newly developed test, or tests being transformed from paper-and-pencil to computer or internet-based tests, could be considered for further research on the effect of the mode of administration.

- Practice effects related to cognitive and non-cognitive measures could be researched.
- Whilst LPCAT will be available in the near future on the internet platform, unproctored settings whilst possible should carefully be researched. The test provider could also consider that many people and test locations may not have internet access, therefore to discontinue the current standalone computerised program option for LPCAT would not be advised. Also in an effort to accommodate I/O Psychologists' preference for computer-or internet-based testing systems the option of providing both program and internet should be kept open.
- CPCAT could be used in unproctored UIT settings as only very few of the sub-dimensions showed statistically significant differences between the mean scores. Furthermore, it should be kept in mind that in the South African context where 11 official language are relevant a certain level of English proficiency may be needed.
- Internet online support and counselling could be researched in future.

4.5 CHAPTER SUMMARY

Tippins (2009) referred to the importance of identifying high stakes testing and indicated that cognitive tests should only be used in South Africa with supervised or managed mode. These findings confirm that test results for a cognitive test do not differ when mode of administration is manipulated. In line with good practice whilst the benefits of UIT were discussed and the future developments of computer-and internet-based assessments can add value in the workplace a level of caution should still be exercised when UIT is considered. Test providers need to investigate the mediums in which tests will be presented, anxiety related to first test sessions, and the equivalence of test results in UIT setting.

REFERENCES

Antal, M., Erös, L., Imre, A. (2010).Computerized adaptive testing: Implementation issues. *Sapientiae, Informatica, 2*(2), 168-183.

Arce-Ferrer, A.J., & Guzmán, E. M. (2009). Studying the equivalence of computer-delivered and paper-based administrations of the Ravens Standard Progressive Matrices Test. *Educational and Psychological Measurement*, 69, 855-867.

Arthur,jr. W., Glaze, R.M., Villado, A.J., & Taylor, J.E. (2009). Unproctored internet-based tests of cognitive ability and personality: Magnitude of cheating and response distortion. *Industrial and Organisational Psychology*, 2, 39-45.

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. *International Journal of Selection and Assessment*, *8*(4), 261-274.

Bartram, D. (2004). Assessment in organisations. *Applied Psychology an International Review*. *53*(2), 237-259.

Bartram, D. (2006). The internationalization of testing and new models of test delivery on the internet. *International Journal of Testing*, *6*(2) 121-131.

Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ31i scores. *International Journal of Selection and Assessment*, *12*(3) 278-284.

Beaty, J.C., Dawson, C.R., Fallaw, S.S.,& Kantrowitz, T.M. (2009). Recovering the scientist - practitioner model: How IO's should respond to unproctored internet testing. *Industrial and Organizational Psychology*, 2, 58-63.

Beaty, J.C., Nye, C.D., Borneman, M.J., Kantrowitz, T.M., Drasgow, F., &Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored non cognitive tests as predictive as job performance? *International Journal of Selection and Assessment*, *19*(1) 1-10.

Brokett, R., G. (1997) Humanism as an Instructional Paradigm. In C.R. Dills& A. J. Romiszowski (Eds.), *Instructional development paradigms* (pp245-256). Englewood Cliffs, NJ: Educational Technology Publications.

Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, 2, 35-38.

Christensen, L.B, (2001). *Experimental Methodology*. (8th ed.). Massachusetts, Pearson Education.

Cianciolo, A.T., & Sternberg, R.J. (2004) *Intelligence a brief history*. Blackwell Publishing.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*(1), 155- 159.

Coyne, I., &Bartram, D. (2006). Design and development of the ITC guidelines on computer based and internet delivered testing. *International Journal of Testing*, *6*(2), 133 -142.

Coyne, I., Warszta, T., Beadle, S., & Sheehan, N. (2005). The impact of mode of administration on the equivalence of a test battery: A quasi-experimental design. *International Journal of Selection and Assessment*, *3*(13) 220-224.

Davies, C., Foxcroft, C., Griessel, L., & Tredoux, N. (2009). Computer Based and Internet-delivered assessment. In C. Foxcroft, & G. Roodt (Eds.), *An introduction to psychological assessment* (3rd ed., pp. 185-200). Cape Town, Southern Africa: Oxford University Press:

De Beer, M. (2000). *Learning Potential Computerised Adaptive Test (LPCAT): User's Manual.* Pretoria: University of South Africa.

De Beer, M. (2005). Development of the Learning Potential Computerised Adaptive Test (LPCAT).*South African Journal of Psychology*, *35*(4), 717-747.

De Beer, M. (2006). Dynamic testing: Practical solutions to some concerns. *South African Journal of Industrial Psychology*, *32*(4) 8-14.

De Beer, M. (2010). A modern assessment psychometric approach to dynamic assessment. *Journal of Psychology in Africa*, *20*(2), 241-246.

De Beer, M (2011). Initial review of psychometric properties of a computerised career preference test for career guidance assessment. *Journal of Psychology in Africa*, *21*(2), 311-314.

De Beer, M. (2012). The Learning Potential Computerised Adaptive Test in South Africa In S.Laher, &Cockroft, K, (Eds.), *Psychological assessment in South Africa: Research and applications 2000 – 2010. (pp.137-157).*

Do, B. (2009). Research on unproctored internet testing. *Industrial and Organizational Psychology*, 2, 49- 51.

Drummond, R,J. (2004). *Appraisal procedures for counsellors and helping*

professionals (5th ed.). Upper Saddle River: Pearson Prentice Hall.

Eid, M., &Diener, E. (2006). *Handbook of multimethod measurement in psychology*. Washington: United Book Press.

Employment Equity Act, 55 of 1998.*Government Gazette*, 400 (19370). Cape Town, South Africa, 19 October 1998.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.).Oxford: The Alden Press.

Foster, D.F. (2009). Secure, online, high-stakes testing: Science fiction or business reality? *Industrial and Organizational Psychology*, 2, 31-34.

Foster, D.F. (2010). Worldwide testing and test security issues: Ethical challenges and solutions. *Ethics & Behavior*, *20*(3-4), 207 - 228. DOI: 10.1080/10508421003798943.

Foxcroft, C.D., & Davies, C. (2006). Taking ownership of the ITC's guidelines for computer-based and internet-delivered testing: A South African Application, *International Journal of Testing*, *6*(2), 173-180.

Foxcroft, C., Paterson, H., le Roux, N., &Herbst, D. (2004) *Psychological assessment in South Africa: A needs analysis.* Pretoria, South Africa: Human Sciences Research Council.

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

Foxcroft, C., Roodt, G., & Abrahams, F. (2009).The practice of psychological

assessment: controlling the use of measures, competing values, and ethical practice standards. In C. Foxcroft, & G. Roodt, (Eds.), *An introduction to psychological assessment (3rd ed., pp. 9-26*). Cape Town, South Africa: Oxford University Press.

Gibby, R.E., Ispas, D., McCloy, R.A., Biga, A. (2009). Moving beyond the challenges to make unproctored internet testing a reality. *Industrial and Organizational Psychology*, 2, 64-68.

Gilmore, N., (2008). *The relationship between learning potential and job performance.* (Unpublished Master's thesis). University of South Africa.

Green, S.B. & Salkind, N.J. (2008). *Using SPSS for windows and macintoch: Analyzing and understanding Data* (5thed.). Upper Saddle River, NJ: Pearson International.

Gregory,R.,J. (2007). *Psychological testing: History, principles, applications* (5th ed.). Boston: Pearson Educational Inc.

Griessel, L., Jansen, J., & Stroud, L. (2009) Administering psychological assessment measures In C. Foxcroft, & G. Roodt, (Eds.), *An introduction to psychological assessment (3rd ed., pp.107-124*). Cape Town, Southern Africa: Oxford University Press.

Haynes, S.N., & O'Brien, W. H. (2000). *Principles and practice of behavioural assessment.* Dordrecht: Kluwer Academic.

Health Professions Act, 56of 1974.*Government Gazette*, 511 (30674). Cape Town,

South Africa, 17 January 2008.

Hense, R., Golden, J.H., & Burnett, J. (2009). Making a case for unproctored internet testing: Do the rewards outweigh the risks? *Industrial and Organizational Psychology*, 2, 20-23.

IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

International Test Commission (2005).*International guidelines on computer-based and internet delivered testing. Version 2005*.[http://www.intestcom.org]

International Test Commission (2011).*International guidelines for quality control in scoring, test analysis, and reporting of test scores*.Version 2011. [http://www.intestcom.org]

Joubert, T., & Kriek, H.J. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *South African Journal of Industrial Psychology*, *35*(1), 78-88.

Kaminsky, K.A., & Hemingway, M.A. (2009).To proctor or not to proctor? Balancing business needs with validity in online testing. *Industrial and Organizational Psychology*, 2, 24-26.

Kanjee, A. & Foxcroft, C. (2009). Cross cultural test adaptation, translation and test in multiple languages. In C. Foxcroft, & G. Roodt, (Eds.), *An introduction to psychological assessment (3rd ed., pp.107-124*). Cape Town, Southern Africa: Oxford University Press.

Kantrowits, T.M., &Dawson, C.R. (2011). Computer adaptive Testing: (CAT): A faster, smarter and more secure approach to pre-employment testing. *Journal of Business Psychology*, 26, 227-232.

Kozulin, A., Gindis, B., Ageyev, V.S., & Miller, S.M. (2003). *Vygotsky's Educational Theory in Cultural Context.* Cambridge: Cambridge University Press.

Krauss, S.E. (2005). Research paradigm and meaning making: A primer. *The Qualitative Report*, *10*(4), 758-770.

Lee, J., Morena, K.E., &Sympson, J.B. (1986). The effects of mode of test administration on test performance. *Educational and Psychological Measurement*, 46, 467- 474.

Leeson, H.V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, *6*(1) 1-24.

Lievens,F., & Burke, E. (2011). Dealing with the threats inherent in unproctored internet testing of cognitive ability: Results from a large scale operational test program. *Journal of Occupational and Organizational Psychology*, 84, 817-824.

Lievens, F., & Harris M.M. (2003). Research on internet recruiting and testing: current status and future directions. In C.L. Cooper & I.T. Robertson (Eds.) *International Review of Industrial and Organizational Psychology,* vol.16 (pp. 131-165). Chicester: John Wiley & Sons Ltd.

Murphy, R., (2002). A review of South African research in the field of dynamic assessment. Unpublished Master's thesis. University of Pretoria, South Africa.

Murphy, R., & Maree, D.J.F. (2009).Revisiting core issues in dynamic assessment. *South African Journal of Psychology*, *39*(4), 420-431.

Naglieri, J.A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the internet: New problems, old issues. *American Psychologist*, 59, 150-162.

Pearlman, K. (2009). Unproctored internet testing: Practical, legal, and ethical concerns. *Industrial and Organizational Psychology*, 2, 14-19.

Reynolds, D.H., Wasko, L.E., Sinar, E.F., Raymark, P.H., & Jones, J.A. (2009). UIT or not to UIT? That is not the only question. *Industrial and Organizational Psychology*, 2, 52-57.

Salgado, J.F., &Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, *11*(2/3), 194-205.

Salkind, N.J. (2009). *Exploring research* (7th ed.). Upper Saddle River, New Jersey: Pearson Education Incorporated.

Statistics South Africa. (2012). Census 2011 key results. Retrieved from http://www.statssa.gov.za

Sternberg, R.J., & Kaufman, S.B. (2011).*The Cambridge Handbook of Intelligence.* Cambridge: Cambridge University Press.

Styles, I., &Andrich, D. (1993). Linking the standard and advanced forms of the Raven's Progressive Matrices in both the pencil-and-paper and computer-adaptive testing

formats. *Educational and Psychology Measurement,* 53, 902-924.

Tanizaki,H. (1997). Power comparison of non- parametric tests: Small-sample properties from Monte Carlo experiments. *Journal of Applied Statistics, 24*(5), 603-632.

Terre Blanche, M., Durrheim.,& Painter, D. (2006). *Research in Practice: Applied Methods for the Social Sciences* (2nd ed.). University of Cape Town: UCT Press.

Tredoux, C., &Durrheim, K. (2002). *Numbers, Hypothesis & Conclusions: A course in statistics for the social sciences.* Cape Town: UCT Press.

Tippins, N. T., Beaty, J., Drasgow F., Gibson W. M., Pearlman, K. Segall, D.O., & Shepherd, W.J. (2006) Unproctored internet testing in employment settings. *Personnel Psychology*, 59, 189-225.

Tippins, N.T. (2009). Internet alternatives to traditional proctored testing: Where are we now. *Industrial and Organizational Psychology*, 2, 2-10.

Van Eeden, R., & De Beer, M., (2009) Assessment of cognitive functioning. In C. Foxcroft, & G. Roodt, (Eds.), *An introduction to psychological assessment (3rd ed., pp.127-147*). Cape Town, Southern Africa: Oxford University Press.

Vispoel, W.P. (2000). Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. Revision of a paper presented at the April 1997 meeting of the National Council on Measurement in Education in Chigago.

Weiner, J.A., & Morrison, Jr, J.D. (2009). Unproctored online testing: Environment

conditions and validity. *Industrial and Organizational Psychology*, 2, 27-30.

Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, *6*(4), 473- 492.

Whitley, B.E. (2002). *Principles of Research in Behavioral Sciences* (2nd ed.). New York: McGraw-Hill.

Willis, I.J.W. (2007). World views, Paradigms, and the Practice of Social Science] Research. *Sage Publications, Inc 1.*