

**THE INCREMENTAL VALIDITY OF A
SITUATIONAL JUDGEMENT TEST (SJT)
RELATIVE TO PERSONALITY AND COGNITIVE
ABILITY TO PREDICT MANAGERIAL
PERFORMANCE.**

Siglind Fertig

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Commerce (Industrial Psychology)
at Stellenbosch University



Promotor: Mr F.S. de Kock

December 2009

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2009

Copyright © 2009 Stellenbosch University

All rights reserved

ABSTRACT

The last two decades have witnessed an upsurge in the research and use of psychometric tests to aid in the prediction of managerial performance. Currently the most prevailing predictor constructs of managerial performance are cognitive ability, personality, and experience. However, researchers and practitioners are still looking for ways in which to maximise the prediction of managerial performance. In recent years, Situational Judgement Tests (SJTs) have become an increasingly popular selection tool. SJTs are multidimensional psychometric instruments designed to assess an individual's judgement concerning work-related situations. Evidence to date indicates that SJTs are valid predictors of performance, especially for managerial positions in which interpersonal interactions are important. The main objective of this study was to examine whether SJTs significantly add to the prediction of managerial performance over other measures used for managerial selection, such as measures of cognitive ability and personality. Measures of specific cognitive abilities, personality and a SJT were administered to branch managers in a South African retail bank ($N = 124$) to investigate the ability of the measures to predict managerial performance. Managerial performance was measured using three measures; Performance Ranking, a Behavioural Observation Scale (BOS) and an Overall Performance Rating. Hierarchical multiple regression was used to investigate the relationship between the predictor composites and the managerial performance measures. Findings reveal different prediction patterns for the three criteria. A multiple correlation coefficient of .442 ($p > .05$) was obtained when predicting Performance Ranking measures, .308 ($p < .05$) was obtained for predicting the Behavioural Observation Scale (BOS) measure, and .318 ($p > .05$) was obtained when predicting the Overall Performance Rating measure. Therefore, only when predicting the BOS measure, the SJT provided incremental validity over cognitive ability and personality measures. Consequently, the average of the scores of the three criterion measures, i.e., the Managerial Performance Composite, was used to evaluate the *a priori* hypotheses. A multiple correlation of .366 ($p > .05$) was obtained for predicting the Managerial Performance Composite criterion. Results therefore indicate that the SJT did not exhibit meaningful or statistically significant incremental prediction over cognitive ability and personality to predict the composite managerial performance measure.

OPSOMMING

Die laaste twee dekades het 'n toename in die gebruik van psigometriese toetse in die voorspelling van bestuurdersprestasië waargeneem. Tans is kognitiewe vermoë, persoonlikheid en ervaring die mees algemene voorspellingskonstrukte vir bestuurdersprestasië. Navorsers en praktisyns is egter op soek na maniere om die voorspelling van bestuurdersprestasië te verbeter. 'n Onlangse verwikkeling is dat "Situational Judgement Tests" (SJTs) toeneem in gewildheid as seleksie-metode. SJTs is multi-dimensionele psigometriese toetse wat ontwerp is om 'n individu se oordeelsvermoë ten opsigte van werksverwante situasies te assesser. Navorsing tot op hede wys dat SJTs geldige voorspellers van prestasië is, veral vir bestuursposisies waarin interpersoonlike interaksies belangrik is. Die hoofdoel van hierdie studie was om te ondersoek of SJTs betekenisvolle waarde toevoeg tot die voorspelling van bestuurdersprestasië bo die gebruik van ander meetinstrumente wat vir bestuurskeuring gebruik word, soos metings van kognitiewe vermoë en persoonlikheid. Vir hierdie doel, is takbestuurders in 'n Suid Afrikaanse bank ($N = 124$) se kognitiewe vermoëns, persoonlikheid en situasionele beoordelingsvermoë getoets om die vermoë van die meetinstrumente om bestuurdersprestasië te voorspel, te ondersoek. Bestuurdersprestasië was deur middel van drie meetinstrumente bepaal; prestasië-rangordening ("Performance Ranking"), 'n gedragsobservasieskaal ("Behavioural Observation Scale") en 'n algehele prestasiëbeoordelingsmeting ("Overall Performance Rating"). Hiërargiese meervoudige regressie-ontleding was gebruik om die verhouding tussen die voorspellers en die bestuurdersprestasiëmetings te ondersoek. Verskillende voorspellingspatrone is vir die drie meetinstrumente gevind. 'n Meervoudige korrelasie koëffisiënt van .308 ($p < .05$) is vir die voorspelling van die BOS meting verkry, terwyl .442 ($p > .05$) en .308 ($p < .05$) onderskeidelik vir die voorspelling van die prestasië-rangordening en algehele prestasiëbeoordelingsmeting verkry is. Gevolglik het slegs die BOS meting inkrementele geldigheid getoon. Die gemiddeld van hierdie drie metings se tellings is gebruik om 'n bestuurdersprestasië-kombinasietelling "Managerial Performance Composite" te skep wat gebruik is om die finale besluit rakende die *a priori* hipoteses te maak. 'n Meervoudige korrelasie van .366 ($p > .05$) was gevind ten einde die bestuurdersprestasië-kombinasietelling te voorspel aan die hand van die voorspellers.

Die resultate dui dus aan dat die SJT nie betekenisvolle inkrementele geldigheid bo metings van kognitiewe vermoë en persoonlikheid vir die voorspelling van bestuurdersprestasie bied in die geval waar die kombinasietelling voorspel word nie.

ACKNOWLEDGEMENTS

This thesis is dedicated to my parents who, throughout all my life, have motivated and encouraged me to believe in myself. Their support, generosity and self-sacrifice have enabled me to come this far in life.

Writing a thesis is a journey... a journey one couldn't make on one's own. I would, therefore, like to convey my thanks and appreciation to the following people:

- My mentor, Leon Venter, for granting me the opportunity to conduct this study within the company. Your support, tolerance, and kindness have been an inspiration to me.
- Prof Hennie Kriek and Tina Joubert from SHL for their professional and knowledgeable statistical advice and assistance.
- My supervisor, Francois de Kock, for his assistance and patience and for believing that I could succeed in this study.
- Joy Dijkman for her statistical assistance and unbelievable support.
- My parents, Ammie and Eduard, for never doubting in my abilities, for listening to my frustrations, and for providing me with the encouragement to push through to the end.
- My special friend, Christine Bosman, for her sincere friendship, continuous support, and uninterrupted interest in all of the aspects and phases of my study.
- The love of my life, Adriaan van der Westhuizen, for his indescribable love and support. You have gone through all the highs and lows of this research endeavour with me and I couldn't have asked for anyone better to share them with. Many, many thanks.
- My Saviour for giving me the ability, strength and perseverance to complete my studies.

To those embarking on this journey: “Enjoy the ride!”

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT	iii
OPSOMMING	iv
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
ANNEXURES	xiv
ABBREVIATIONS	xv

CHAPTER ONE: INTRODUCTION AND OBJECTIVES OF THE CURRENT STUDY

1.1	INTRODUCTION AND JUSTIFICATION FOR THE STUDY	1
1.2	RESEARCH OBJECTIVES	5
1.3	OUTLINE OF THE STUDY	6

CHAPTER TWO: LITERATURE REVIEW

2.1	INTRODUCTION	7
2.2	RELEVANT PSYCHOMETRIC CONCEPTS IN PERSONNEL SELECTION	7
2.2.1	The Essential Logic Underlying Personnel Selection	8
2.2.2	Reliability	8
2.2.3	Validity	9
2.2.3.1	Content-related Validity	10
2.2.3.2	Construct-related Validity	10
2.2.3.3	Criterion-related Validity	11
2.3	THE CRITERION CONSTRUCT	12
2.3.1	Job Performance	12
2.3.2	Managerial Performance	16

2.4	PREDICTORS OF MANAGERIAL PERFORMANCE	19
2.4.1	Cognitive Ability	19
2.4.1.1	Theoretical Underpinnings and the Structure of Cognitive Ability	21
2.4.1.2	Empirical Findings regarding the Predictiveness of Cognitive Ability	22
2.4.2	Personality	27
2.4.2.1	The Structure of Personality	27
2.4.2.2	Empirical Findings on the Predictiveness of the Big Five Personality Factors	29
2.4.3	Situational Judgement Tests	32
2.4.3.1	Empirical Findings on the Psychometric Properties and Predictiveness of SJTs	33
2.4.3.1.2	Reliability	33
2.4.3.1.2	Face Validity	35
2.4.3.1.3	Construct-related Validity	35
2.4.3.1.4	Criterion-related Validity	39
2.4.3.1.5	An Integrated Model of the Construct and Criterion-related Validity Evidence for SJTs	41
2.4.3.1.6	Incremental Validity	43
2.5	PROPOSED RESEARCH MODEL	45
2.6	CONCLUSION: CHAPTER TWO	52
 CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY		
3.1	INTRODUCTION	53
3.2	RESEARCH DESIGN	53
3.3	HYPOTHESES	54
3.4	PARTICIPANTS AND PROCEDURE	57
3.5	MEASURING INSTRUMENTS	60
3.5.1	Managerial Performance	60
3.5.2	Situational Judgement Test	63
3.5.3	Cognitive Ability	69
3.5.4	Personality	71

3.6	STATISTICAL ANALYSIS	73
3.7	CONCLUSION: CHAPTER THREE	74
CHAPTER FOUR: RESEARCH RESULTS		
4.1	INTRODUCTION	75
4.2	DESCRIPTIVE STATISTICS	75
4.2.1	Assumptions underlying Multivariate Statistical Analyses	76
4.2.1.1	Accuracy of Data File and Missing Values	76
4.2.1.2	Ratio of Cases to Independent Variables	76
4.2.1.3	Outliers	76
4.2.1.4	Univariate Normality, Multivariate Linearity and Homoscedasticity	78
4.3	ITEM ANALYSIS	83
4.4	DIMENSIONALITY ANALYSIS	89
4.5	RESULTS	93
4.5.1	Inter-correlations	93
4.5.1.1	The Relationship between Verbal Evaluation and Managerial Performance (H ₁)	95
4.5.1.2	The Relationship between Numerical Interpretation and Managerial Performance (H ₂)	95
4.5.1.3	The Relationship between Basic Checking Ability and Managerial Performance (H ₃)	96
4.5.1.4	The Relationship between Conscientiousness and Managerial Performance (H ₄)	96
4.5.1.5	The Relationship between Extraversion and Managerial Performance (H ₅)	97
4.5.1.6	The Relationship between Conscientiousness and Performance on a Situational Judgement Test (H ₆)	97
4.5.1.7	The Relationship between Agreeableness and Performance on a Situational Judgement Test (H ₇)	97
4.5.1.8	The Relationship between Emotional Stability and Performance on a Situational Judgement Test (H ₈)	98

4.5.1.9	The relationship between Managerial Performance and Performance on a Situational Judgement Test (H ₉)	98
4.5.2	Corrections for Unreliability and Restriction of Range	103
4.5.3	Regression Results	106
4.5.3.1	Standard Multiple Regression of Managerial Performance Composite on all Predictors (H ₁₀)	107
4.5.3.2	Hierarchical Regression of Managerial Performance Composite on all Predictors (H ₁₁)	110
4.5.4	Do Situational Judgment Test (SJT) Scores Mediate the Influence of Personality Measures on Managerial Performance? (H ₁₂)	115
4.6	CONCLUSION: CHAPTER FOUR	116
CHAPTER FIVE: DISCUSSION OF RESULTS		
5.1	INTRODUCTION	118
5.2	GENERAL CONCLUSIONS	118
5.2.1	Inter-correlations	119
5.2.1.1	Relationship between Cognitive Ability and Managerial Performance	119
5.2.1.2	Relationship between Personality and Managerial Performance	123
5.2.1.3	Mediating Affect of Individual Differences on Performance on the Situational Judgement Test	127
5.2.1.4	Relationship between Performance on the Situational Judgement Test and Managerial Performance	129
5.2.2	Regression Results	131
5.3	LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH	133
5.4	CONCLUSIONS AND IMPLICATIONS	136
	REFERENCES	138
	APPENDICES	155

LIST OF TABLES

Table 2.1	Campbell's Taxonomy of Eight Major Performance Components	13
Table 2.2	Comparison of Three Managerial Performance Models	18
Table 2.3	Relationship between Competencies	19
Table 2.4	The Operational Validities (ρ) of General and Specific Cognitive Abilities for Job Performance (K = 250, N = 25,000)	23
Table 2.5	The Operational Validities (ρ) of General and Specific Cognitive Abilities for Job Performance (K = 283, N = 13,000)	24
Table 2.6	The Operational Validities (ρ) of Cognitive Ability for Job Performance over Eight Occupational Groups (K = 283, N = 13,000)	24
Table 2.7	Predictive Validity (r) for Overall Job Performance of General Mental Ability (GMA) Scores Combined with a Second Predictor Using (Standardised) Multiple Regression	26
Table 2.8	The Big Five Personality Factors	28
Table 2.9	Meta-Analytic Results of Correlations between Situational Judgement Tests and Cognitive Ability, Agreeableness, Conscientiousness, and Emotional Stability	37
Table 3.1	Demographic Profile of the Sample	59
Table 4.1	Analysis of Univariate Descriptives of all Variables	77
Table 4.2	Kolmogorov-Smirnov Test of Normality for the Cognitive Measures	79
Table 4.3	Kolmogorov-Smirnov Test of Normality for the Video-based Simulation	79
Table 4.4	Kolmogorov-Smirnov Test of Normality for the BOS Measure	80
Table 4.5	Kolmogorov-Smirnov Test of Normality for the OPQ Subscales	81
Table 4.6	Reliability Analysis of the Items Comprising the BOS Questionnaire	85
Table 4.7	Reliability Analysis of the Items Comprising the BOS Questionnaire Post-Deletion of Item 7	86
Table 4.8	Mean Inter-Item Correlations of the BOS Subscales	87
Table 4.9	Factor Loadings of All Items Comprising the BOS Questionnaire	90

Table 4.10	Factor Loadings of Items Comprising the BOS Questionnaire, Excluding Item 7	91
Table 4.11	Correlations between Predictors and Criteria	100
Table 4.12	Correlations between the Predictors and the Behavioural Observation Scale (BOS) corrected for Unreliability and Restriction of Range	105
Table 4.13	Standard Multiple Regression of all Managerial Performance Composite on all Predictors	109
Table 4.14	Hierarchical Regression of Managerial Performance on all Predictors	112
Table 4.15	Summarised Hierarchical Regression results of the BOS, Perf Rank and OPR Performance Measures	114
Table 4.16	Cross validity estimates for Standard Multiple Regression Models	116
Table 5.1	Correlations between Cognitive Ability Predictors and Criteria	121
Table 5.2	Correlations between Personality Measures and Criteria	126
Table 5.3	Correlations between Individual Differences and the VBS	128
Table 5.4	Correlations between Performance on the VBS and Criteria	131

LIST OF FIGURES

Figure 2.1	Campbell's Determinants of Job Performance	14
Figure 2.2	Framework for Relating the Multidimensional Nature of SJT to KSAOs and Job Performance	39
Figure 2.3	Conceptual Model of the Factors affecting the Validity of SJTs	42
Figure 2.4	Hypothesised Relationship between the Dependent and Independent Variables	46
Figure 3.2	Sample Item from the NCC2	71
Figure 4.1	Research Model indicating Significant Correlations between Predictors and Criteria	102

ANNEXURES

APPENDIX A: WPS Competency Profile for Branch Managers	155
APPENDIX B: Questionnaires	167
APPENDIX C: Fritz and MacKinnon's (2007) necessary sample sizes for six of the most common and the most recommended tests of mediation for various combinations of parameters	175
APPENDIX D: Ethics Form	177

ABBREVIATIONS

BOS	Behavioural Observation Scale
FFM	Five Factor Model
GMA	General Mental Ability
HR	Human Resources
I/O psychology	Industrial/Organisational Psychology
KSAO	Knowledge, skills, abilities, and other personal characteristics
MPerf Comp	Managerial Performance Composite
OPQ32	Occupational Personality Questionnaire
OPR	Overall Performance Rating
16PF	Sixteen Personality Factor Questionnaire
Perf Rank	Performance Ranking
SHL	Saville & Holdsworth Ltd.
SJT	Situational Judgement Test
SME	Subject Matter Expert
VBS	Video-based Simulation
WPS	Work Profiling System

CHAPTER ONE

INTRODUCTION AND OBJECTIVES OF THE CURRENT STUDY

1.1 INTRODUCTION AND JUSTIFICATION FOR THE STUDY

Selection research has enjoyed a rich history within the field of industrial/organisational (I/O) psychology, as decades of basic and applied research have been devoted to understanding the prediction of job performance. Accurate prediction of job performance is critical to the success of organisations. Schneider adequately summarises the rationale behind selection research in I/O psychology as:

...an approach to understanding organizational functioning and effectiveness by focusing first on individuals and relationships between individual attributes and individual job behaviour. The hallmark of I/O has been a concern for discovering what individual characteristics (abilities, needs, satisfactions) are useful for predicting work behaviour required for the organization to be effective (productivity in terms of quality and/or quantity, absenteeism, turnover, sales, and so forth). I/O work is based on the simple assumption that when accurate predictions about the effectiveness of individuals are made, then it follows that the organization will be more effective.

For example, I/O researchers assume that when assessments of individuals at the time of hire are significantly related to some performance standard on the job two or five years later, then utilization of the assessment technique for hiring people will yield a higher proportion of effective workers and the organization will be more effective. Issues surrounding the definition and measurement of effectiveness are a major focus for I/O psychologists because we believe that if we fail to grapple with what we want to predict, it will be terribly difficult to predict it. (1984, p. 206)

Thus, a central premise of our approach in this research relies on the individual difference model. Individual difference variables are those human attributes that set one individual apart from another and can be classified broadly into abilities

(cognitive and physical), personality, orientation (interests and values), knowledge and emotion (Landy & Conte, 2007). The individual differences model has a number of fundamental assumptions that guide personnel selection. The first assumption is that adults have a variety of attributes that are relatively stable over a period of time. People have a habitual way of dealing with others and events in their environment. Second, people differ with respect to such attributes. The attributes often form the basis for personnel decisions when they are relevant to a given job. Third, differences among people on the said attributes remain relatively constant, even after training or accrued professional experience. Specific knowledge and abilities may be enhanced by training and experience, but, in general, the relative rank order of an enlarged group of people will not change substantially. Fourth, different jobs require different attributes. Maximising the fit between the attributes of the candidate and the needs of the job is one of the fundamental principles of personnel selection. The fifth assumption is that such attributes can be measured. Various selection tools and techniques are available to determine which candidates are most suited for a particular job and organisation (Landy & Conte, 2007).

It is well established in the literature that certain individual differences, such as cognitive ability and personality, are some of the best predictors of job performance (Schmidt & Hunter, 1998). Recently, there has also been an upsurge in research into the prediction of managerial performance (e.g., Campbell, Dunnette, Lawler & Weick, 1970; Goffin, Rothstein & Johnston, 1996; Hunter & Hunter, 1984). The prediction of managerial performance is particularly important, since managers can uniquely affect the culture and productivity of an organisation, due to their influential positions (Young, Arthur & Finch, 2000).

Job performance is multidimensional in nature and managerial performance, in particular, involves a broad domain of required performance behaviours (Borman & Brush, 1993; Campbell, 1990). Young et al. (2000, p. 54) argue that “different aspects of managerial performance require different abilities and/or personality characteristics”. They continue by saying that “the explanatory power of a given individual differences variable should depend on the particular aspect of managerial performance being predicted” (p. 54). Many potential predictors of managerial performance have been examined of which the constructs of cognitive

ability (Cascio, 1991; Hunter & Hunter, 1984; Jensen, 1998; Schmidt & Hunter, 1998), personality (Barrick & Mount, 1991; Ones, Dilchert, Viswesvaran & Judge, 2007; Tett & Christiansen, 2007; Tett, Jackson & Rothstein, 1991), and experience (Borman, Hanson, Oppler, Pulakos & White, 1993) are currently the most prevalent. However, researchers and practitioners are looking for ways in which to maximise the prediction of managerial performance by investigating the use of new predictor constructs and measures of such constructs.

In recent years, Situational Judgement Tests (SJTs) have become an increasingly popular selection tool (Lievens, Peeters & Schollaert, 2008; Whetzel & McDaniel, 2009). SJTs are psychometric instruments designed to assess an individual's judgement concerning work-related situations (Chan & Schmitt, 2002; McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001). SJTs present respondents with a variety of scenarios, which they would typically encounter on the job, and ask them to indicate which of a set of possible responses is the most appropriate for a particular situation (Lievens et al., 2008). Such responses are often scored according to their relative level of effectiveness, rather than being indicated as simply right or wrong. Unlike many selection tests, SJTs are multidimensional. As job situations are complex, the situations presented in the SJTs are usually also complex (Chan & Schmitt, 2002).

Evidence, to date, indicates that SJTs are valid predictors of performance, especially in terms of managerial positions, in which interpersonal interactions are important (Motowidlo, Dunnette & Carter, 1990). They can assess job-related skills, such as those relating to personal initiative, conflict management, interpersonal communication, problem solving, negotiation, teamwork facilitation, and cultural awareness, that remain untapped by other measures (Bledow & Frese, 2009; Chan & Schmitt, 1997; Lievens & Sackett, 2007; McDaniel & Nguyen, 2001; McDaniel & Whetzel, 2005; O'Connell, Hartman, McDaniel, Grubb & Lawrence, 2007; Weekley & Jones, 1999).

The increased popularity of SJTs is due to a number of its positive features being proven. First, research indicates that SJTs can validly predict job performance incrementally over cognitive ability and personality tests (Chan & Schmitt, 2002;

Clevenger, Pereira, Wiechmann, Schmitt & Schmidt-Harvey, 2001; McDaniel et al., 2001; McDaniel, Hartman, Whetzel & Grubb, 2007; McDaniel, Whetzel, Hartman, Nguyen & Grubb, 2006; O'Connell et al., 2007; Weekley & Jones, 1997, 1999). Second, SJTs produce lower levels of adverse impact than do traditional ability tests (Clevenger et al., 2001; Hanson & Borman, 1995; Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996). Valid predictors with relatively low adverse impact are difficult to find, resulting in the search for such alternative predictors becoming increasingly important in most applied settings. Third, SJTs produce more favourable test-taker reactions than do tests of cognitive ability (Weekley & Ployhart, 2005). In fact, Rosen (1961) argues that, even if an SJT adds nothing to the prediction of success beyond what can be obtained by means of intelligence tests and biodata analysis, "... the instrument's high face validity makes it more desirable to use than some others" (p. 97).

A current debate in the literature revolves around whether situational judgement is a construct, or a measure of other constructs. McDaniel et al. (2006) argue that performance on SJTs is influenced by cognitive ability and personality, and that the extent to which SJTs measure such constructs varies, as it is moderated by the particular SJT's response instructions. They further propose that, although SJTs correlate with other constructs, such tests also allow for the coverage of individual differences not measured by the others. Therefore, job performance is predictable in terms of cognitive ability and personality, as well as by means of the use of SJTs. Weekley and Ployhart conclude that:

...Situational Judgement Tests have many characteristics that make them attractive predictor measures. However, research must be conducted to better understand these measures, particularly how they relate to other commonly used individual difference variables. Future research will help establish the degree to which SJTs are generalizable or specific to the job-organization in which they were developed. (2005, p. 101)

1.2 RESEARCH OBJECTIVES

In the light of the above discussion, it is evident that various individual differences directly and indirectly affect job performance, and that the use of SJTs makes a significant contribution to the prediction of managerial performance. The question then arises as to how SJT performance relates to managerial performance, in conjunction with measures of cognitive ability and personality. Therefore, the first research objective is to determine whether SJTs significantly add to the prediction of managerial performance over other measures used for managerial selection, such as those of cognitive ability and personality. McDaniel et al. (2006), on concluding that SJTs typically show incremental validity over cognitive ability tests, called for more research into such validity over both cognitive ability and personality tests. Furthermore, since the SJT used in the current research is video-based, the results should also address their call for more validity data on video-based SJTs (McDaniel et al., 2006).

Similarly, it is apparent that the meaning of SJT scores remains in dispute. Weekley and Ployhart urge that:

... more research into the relative contribution of cognitive and non-cognitive constructs as determinants of SJT performance is warranted. Although the evidence indicates that cognitive ability is related to nearly all SJTs (McDaniel et al., 2001), it is likely that noncognitive correlates will vary as a function of the item content of the SJT itself (as dictated by changes in the job). Future research should continue to explore the correlates of SJTs across different jobs and organizations. A particularly important point is that if SJTs are measurement methods, research should not attempt to identify the correlates assessed by SJTs in general, but rather focus on the correlates of SJTs in particular classes of jobs. What is clearly needed is some theoretical basis for understanding how and why personality traits might be differentially related to various SJTs. (2005, p. 100)

Therefore, the second research objective is to answer the above-stated question by investigating the SJT's relation with the other predictor measures for managerial positions, in terms of managerial performance. By so doing, we shall be able to

address an important gap in the current knowledge base, thereby promoting an understanding of the constructs measured by the SJT, as well as of the relationships that the SJT have with other constructs. To meet such research objectives, the current study is structured as described below.

1.3 OUTLINE OF THE STUDY

Chapter two provides an extensive review of the literature on job performance, focusing specifically on managerial performance. Thereafter, the relevant predictors namely cognitive ability, personality and situational judgement, are discussed in detail. Based on the literature review, an empirical model is proposed, which outlines the possible relationships between the variables.

In the light of the model proposed in chapter two, the research design and methodological approach to be adopted in the current study are discussed in chapter three. In addition, the composition and nature of the sample, as well as the measuring instruments used, are described. The statistical analysis used in the study is also explained, with the results of the data analysis being reported in chapter four. The findings of the statistical analysis, which was carried out in an attempt to determine whether to accept or reject the hypotheses stated in chapter three, are detailed in chapter four.

Chapter five presents the discussion and conclusions of the main findings regarding the two research questions and their hypotheses. Chapter five discusses the limitations and recommendations, based on the results of the current study.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

Chapter one explained that the current study is approached from the individual difference model perspective and argued the importance of understanding the incremental validity of SJTs over personality and cognitive ability tests as a predictor of managerial performance. Chapter two starts with a theoretical discussion of important concepts in the field of personnel selection. Thereafter, the literature on the criterion construct is reviewed, in the light of the objective of all selection research to explain variance in job performance. If selection research indicates that predictors do, in fact, predict variance in criterion performance, then those selection practices that are aimed at maximising criterion performance could be used, given that such practices would lead to selection utility (Schmidt & Hunter, 1998). Therefore, the first part of this chapter will examine the models of job performance in general, and managerial performance specifically.

The central premise of this study is that variance in performance is a result of the specific differences in individuals. Therefore, the series of individual differences that are used in this study, namely cognitive ability, personality and situational judgement, will be discussed. In respect of each difference, the definitions of the construct, the major theoretical perspectives, contemporary research about the psychometric properties, as well as evidence from selection research, will be presented. The discussion will culminate in a suggested model that will be used to test the hypotheses in order to answer the research question.

2.2 RELEVANT PSYCHOMETRIC CONCEPTS IN PERSONNEL SELECTION

Those concepts that are relevant to personnel selection and the discussion of the current study are briefly discussed below.

2.2.1 The Essential Logic Underlying Personnel Selection

A crucial element in the achievement of organisational goals is the selection of individuals with a high ability to perform their jobs. Campbell, Gasser and Oswald (1996) reviewed the findings on the value of high and low job performance, estimating that “the top 1% of workers produces a return about 3.3 times as great as the lowest 1% of workers” (Jensen & Nyborg, 2003, p. 270). Moreover, they estimated that “the value might be from 3 to 10 times the return of the lowest 1%, depending on the variability of job performance” (Jensen & Nyborg, 2003, p. 270). The ultimate challenge in personnel selection is, therefore, to maximise predictive efficiency by identifying and selecting individuals with the highest job-relevant ability. The ideal would have been for selection decisions to be based on measurements of the multidimensional final criterion, namely job performance. However, when selection decisions are being made, direct information about the applicants’ job performance is not yet available. The best alternative, therefore, is to make such decisions based on the estimates of job performance, namely predictor information, which are available at the time (Theron, 2007).

Predictor information provides accurate estimates of job performance, to the extent that (1) the predictor correlates with a measure of job performance; and (2) the nature of the relationship in the relevant applicant population is known. A wide variety of predictor measures is currently available to assist practitioners with making such selection decisions (Schmidt & Hunter, 1998). The measures are designed to reveal attributes, skills, and qualities of the individual that indicate their suitability for the job. This means that only appropriate measures ought to be used. The use of reliable and valid measures is more likely to lead to an appropriate selection decision and the appointment of a suitable candidate (Hunter & Hunter, 1984). More detail about the concepts of reliability and validity follows below.

2.2.2 Reliability

Reliability refers to the consistency of the scores obtained from a measure (Nunnally, 1978). A measure is considered reliable if the same results are achieved when the measure is repeatedly applied to the same group (Babbie & Mouton, 2001).

There are various forms of reliability, of which the following is the most important and used most often.

Test–retest reliability refers to the stability of test scores at different points in time. The weakness in this method is the added complication of sources of distortion that occur between the tests, such as the time gap, the level of difficulty of the items concerned, the specific subjects/sample, and the sample size. Internal consistency reliability refers to the relationship of the test items to each other, which is considered to be a random sample of a universe of items. The coefficient alpha is the best index of internal consistency and must be as high as possible. Parallel-form reliability is essentially the same as internal consistency reliability, except that, in the former case, the items are divided into two test versions instead of one (Kaplan & Saccuzzo, 2001).

Kaplan and Saccuzzo (2001) note that reliability, however, does not ensure accuracy; it only indicates the extent to which test scores are free from potential errors of measurement. Therefore, reliability is a necessary, though insufficient, condition for validity. Without reliability, the research results obtained by using the instrument are not replicable, with replicability being fundamental to the scientific method.

2.2.3 Validity

The *Standards for educational and psychological testing* define validity as “the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test” (cited in SIOPSA, 2005, p. 5). Stated otherwise, validity refers to the extent to which a selection test measures what it claims to measure, which impacts on the certainty with which inferences can be made. Neither the test nor the content of a test can be valid, but rather the inferences, which can be drawn from scores on the measure, might be so. Hence, validity is not an inherent property of the test, rather being the relationship between people’s performance on the test and their performance on the job, to the standard measured by the test (Gatewood & Feild, 1994).

In the past, validity was categorised according to content, criterion-related and construct validity. However, researchers have now started to move away from these

three separate aspects of validity evidence, in preference of viewing validity as “a unitary concept with different sources of evidence contributing to an understanding of the inferences that can be drawn from a selection procedure” (SIOPSA, 2005, p. 6). Therefore, the evidence concerning content relevance, criterion-relatedness, and construct meaning is integrated in terms of such a definition of validity. Such integration will be discussed in the following sections.

2.2.3.1 Content-related Validity

Content-related validity indicates that “the selection procedure adequately samples and is linked to the important work behaviours, activities, and/or worker KSAOs [knowledge, skills, abilities, and other personal characteristics] defined by the analysis of work” (SIOPSA, 2005, p. 22). A selection measure that is content valid exposes the job applicant to situations that are likely to occur on the job, and then tests whether the applicant currently has sufficient knowledge, skills, and abilities to handle such situations. The principles underlying the measure, therefore, are based on the notion that, if test items are reasonable samples of the actual job, then the relationship between the test scores and performance is clear (Landy, 1993). Such an approach can only be used when the sample sizes are small, since the demonstration of content validity is typically made by expert judgement and is, therefore, the only type of evidence that is logical rather than statistical (Kaplan & Saccuzzo, 2001).

2.2.3.2 Construct-related Validity

Construct-related validity refers to the evidence that an assessment actually measures the construct that it intends to measure. Such evidence is required to test hypotheses about the relationships between measures and their constructs (Schmitt & Chan, 1998). Since the inclusion of items in a selection procedure should be based primarily on their relevance to a construct or content domain, the researcher may “consider the relationships among items, components of the selection procedures, or scales measuring constructs” (SIOPSA, 2005, p. 22).

2.2.3.3 Criterion-related Validity

The purpose of any selection measure is to predict future performance or other work behaviour. Test scores should be interpreted in terms of expected job performance, and not in terms of the construct being assessed. Criterion-related validity demonstrates a relationship between the “results of a selection procedure (predictor) and one or more measures of work relevant behaviour of work outcomes (criteria)” (SIOPSA, 2005, p. 12).

Criterion-related validity studies are conducted in one of two ways. In a predictive criterion-related study, the data is collected over time. First, the test scores are collected from job applicants, with their performance being measured and the strength of the predictor–criterion relationship being evaluated at a later stage. Since the test motivation of the job applicants is, consequently, more realistic, it might positively influence the way in which the applicants complete the tests (Schmitt & Chan, 1998). Unfortunately, determining predictive validity is quite time consuming (Nunnally, 1978). In a concurrent criterion-related study, the test is administered to employees, with the test scores then being correlated with the existing measures of each person’s performance. The reason behind the adoption of such an approach is that, if the best performers currently in the job perform better on the test than do those who are struggling to master the job, the test has validity. However, since the respondents, in such a case, are already employees, their test motivation might be lower, resulting in its negatively influencing the way in which the tests are completed if the respondents do not take such testing seriously (Schmitt & Chan, 1998).

The observed correlation between the predictor and the criterion is called the validity coefficient. Such a coefficient expresses the extent to which a test is valid for making inferences about the criterion (Gatewood & Feild, 1994). There is no specific cut-off point at which a validity coefficient is considered more or less meaningful. Since validity coefficients larger than .60 are rarely seen, those ranging between .30 and .40 are usually considered high. A coefficient is statistically significant “if the chances of obtaining its value by chance alone are quite small: usually less than 5 in 100” (Kaplan & Saccuzzo, 2001, p. 138).

As discussed earlier, validity is viewed as a unitary concept representing all of the evidence that supports the intended inferences drawn from the selection measures. Another type of validity that is worth mentioning, based on the premise of this study, is incremental validity. Landy and Conte defines incremental validity as “the value in terms of increased validity of adding a particular predictor to an existing selection system” (2007, p. 148.) He continues by saying that the issue with assessments is therefore “not *which* tool to use, but what *combination* of tools to use for the greatest predictive ability and the lowest cost” (2008, p. 148). Those concepts already discussed will be briefly referred to later in the chapter.

2.3 THE CRITERION CONSTRUCT

2.3.1 Job Performance

As it is important to be able to predict performance, a clear understanding of what job performance entails is essential. Campbell’s (e.g., Campbell, 1990; Campbell et al., 1996; Campbell, McCloy, Oppler & Sager, 1993) theories of job performance are among those most commonly accepted. Campbell and colleagues developed a model of job performance that described its latent structure and determinants. Several key features define Campbell’s conceptualisation of job performance. First, job performance is defined as observable behaviours. Such performance is what people do, with it being reflected in the actions that people take, rather than in the results or outcomes of their behaviours. In other words, the performance is not influenced by factors that might be beyond the individual’s control. Second, job performance includes only those actions or behaviours relevant to the organisation’s goals. Finally, job performance is conceptualised as a multidimensional construct, consisting of more than one kind of behaviour.

Campbell’s most significant contribution was the development of a taxonomy of eight major performance components. The components are described in Table 2.1. Those that are subset describe the highest order latent variables for every job in the occupational domain, although some factors might not be relevant for all jobs. However, Campbell contends that three of the performance components, namely job-specific task proficiency, demonstrating effort, and maintaining personal discipline,

are important components of performance in virtually every job. Campbell et al. (1996) acknowledges that one additional performance component may be added to the taxonomy. The potential component identifies how well individuals adapt to new conditions or job requirements.

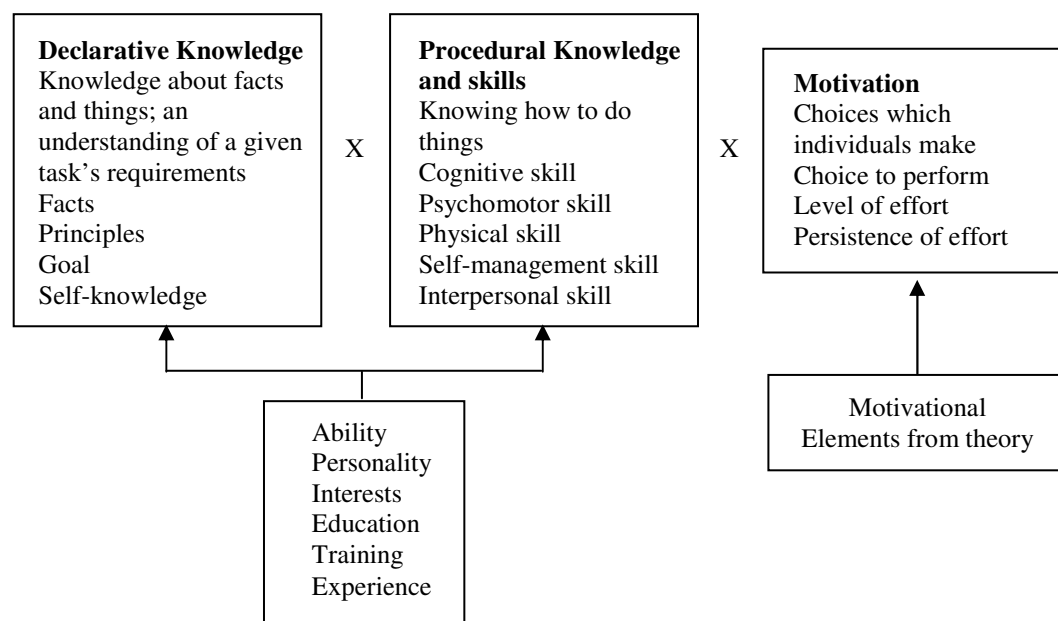
Table 2.1

Campbell's Taxonomy of Eight Major Performance Components

Performance component	Definition
Job-specific task proficiency	An individual's capacity to perform the core substantive or technical tasks central to the job.
Non-job-specific task proficiency	An individual's capacity to perform tasks or execute performance behaviours that are not specific to their particular jobs.
Written and oral communication task proficiency	An individual's proficiency in writing and speaking, independent of the correctness of the subject matter.
Demonstrating effort	The consistency of an individual's effort; the frequency with which people will expend extra effort when required; the willingness to keep working under adverse conditions.
Maintaining personal discipline	The extent to which an individual avoids negative behaviour, such as excessive absenteeism, alcohol or substance abuse, and legal or rule infractions.
Facilitating peer and team performance	The extent to which an individual supports peers, helps peers with problems, helps to keep a work group goal directed, and acts as a role model for peers and the work group.
Supervision/ leadership	Proficiency in influencing the performance of subordinates through face-to-face interpersonal interaction and influence.
Management/ administration	Behaviour directed at articulating for the unit, organising people and resources, monitoring progress, helping to solve problems that might prevent goal accomplishment, controlling expenses, obtaining additional resources, and dealing with other units.

Source: Adapted from Campbell, 1990, p. 708.

Campbell proposed three direct determinants of job performance, namely declarative knowledge (knowing what to do), procedural knowledge and skill (knowing how to do it), and motivation (choice, level of effort, and persistence). He refers to such determinants as the basic building blocks, or causes, of performance. Each of the eight performance components is a function of the three performance determinants, although their relative importance might differ across the eight dimensions. Furthermore, each of the performance determinants is affected by individual difference variables (e.g., ability, personality and interests), situational variables (e.g., education, training and experience), and their interaction. (See Figure 2.1.)



Source: Adapted from Campbell, 1990, p. 707.

Figure 2.1 Campbell's Determinants of Job Performance

Other researchers have expanded the criterion domain to consider dimensions of performance outside the technical proficiency or task performance elements of job performance. In particular, Borman and Motowidlo (1993) distinguished between task and contextual performance. They define task performance as “the proficiency with which job incumbents perform activities that are formally recognised as part of their job” (p. 73), as well as that which contributes to the organisation’s technical core. In contrast, contextual performance is more informal, contributing to organisational effectiveness in ways that shape the organisational, social, and psychological context

in which the technical core must function. The five categories of contextual performance are: (1) volunteering to carry out task activities, even if they are not part of the job; (2) persisting with extra effort, when necessary, to complete own work; (3) helping and cooperating with others; (4) following organisational rules and procedures; and (5) endorsing and supporting organisational procedures.

Borman and Motowidlo (1993) argue that both task and contextual performance are important contributors to organisational effectiveness. Further, Motowidlo and Van Scotter (1994) and Borman, White and Dorsey (1995), among others, have demonstrated that experienced supervisors weigh employee task and contextual performance approximately equally when making overall performance or effectiveness judgments of the employees.

Campbell et al. (1996) point out that the performance factors suggested by the abovementioned authors can easily be integrated into the eight-component taxonomy as sub-factors, forming a hierarchical description of the latent structure of performance. Campbell's (1990) job-specific and non-job-specific task proficiency components overlap with Borman and Motowidlo's (1993) task performance domain, whereas the demonstrating effort, maintaining personal discipline, and facilitating peer and team performance components are captured in the behaviours that Borman and Motowidlo (1993) describe as contextual performance.

Viswesvaran and colleagues (Viswesvaran, 1993; Viswesvaran & Ones, 2000; Viswesvaran, Schmidt & Ones, 2005), on the other hand, developed a three-level hierarchical model of job performance, with a general factor at the highest level. Their meta-analysis identified 25 conceptually distinct categories of job performance measures (e.g., quality of performance, communication skills, compliance and acceptance of authority). In addition, five main themes (productivity, conscientiousness, interpersonal skills, withdrawal, and measures of overall job performance) were identified to group the 25 measures. Thus, the five group factors are at the second level, with the 25 categories of performance measurement being at the lowest level, in their hierarchical model. Viswesvaran, therefore, argues that a general performance factor explains substantial variation in virtually all measures of job performance.

Campbell et al. (1996) acknowledge that the existence of a general factor is likely, due to the contribution of *g* and the element of conscientiousness to many performance components. However, they argue that the eight factors that they have identified describe the highest order latent variables that can usefully describe performance.

2.3.2 Managerial Performance

The performance domain of most jobs is complex (Campbell et al., 1993), and might even be more so for managerial jobs (Borman & Brush, 1993; Tett, Guterman, Bleier & Murphy, 2000). Numerous managerial performance taxonomies are available (e.g., Bartram, Robertson & Callinen, 2002; Borman & Brush, 1993; Kurz & Bartram, 2002; Tett et al., 2000) to expand on Campbell's (1990) supervision–leadership and management–administration components, two of which are exceedingly comprehensive.

Borman and Brush (1993) identified 246 potential dimensions of managerial performance from published and unpublished studies across a wide range of occupational settings. The dimensions are behaviourally based and, therefore, reflect what managers actually do, and not what they, or others, believe they do. The dimensions were compressed into 187 similar content domains, where after further psychometric methods were used to develop an 18-factor solution. The 18 factors (e.g., planning and organising, training, coaching, developing subordinates, technical proficiency) compare well with previous research efforts and are easily compared to the task or citizenship performance categories, or to one of Campbell et al.'s eight components (Borman & Brush, 1993).

Tett et al. (2000) developed a more specific managerial performance taxonomy that combined twelve models of managerial performance, including that of Borman and Brush (1993). They identified 53 dimensions of managerial performance, grouped in nine general areas: traditional functions; task orientation; person orientation; dependability; open-mindedness; emotional control; communication; developing self and others; and occupational acumen and concerns. Unlike general dimensions in

other models, the nine categories do not represent correlations among competencies. All the competencies refer to work behaviours attributable to the individual.

Consistent with the work of Tett et al. (2000) is that of Bartram and colleagues, who developed the generic Saville & Holdsworth Ltd. (SHL) competency framework (Bartram et al., 2002; Kurz & Bartram, 2002), which can also be used to model managerial performance. They analysed the structure of the universe of competencies, which they define as “sets of behaviours that are instrumental to the delivery of desired results” (Bartram et al., 2002, p. 7). They proposed the eight-factor competency framework, better known as the Great Eight model, which divides 112 specific competencies into eight broader categories; leading and deciding; supporting and cooperating; interacting and presenting; analysing and interpreting; creating and conceptualising; organising and executing; adapting and coping; and enterprising and performing (Bartram, 2005).

Table 2.2 consists of a comparison of the models, which indicates clear similarities between the competencies, such as decision making; planning; organising; interacting; problem solving; and leading, with the exception of certain variations, such as safety concern and self-development.

The in-house competency model of the host organisation in the current study is based on the competency framework of SHL. The competency profile for branch managers, which was generated by SHL’s work profiling system (WPS), can be found in appendix A. The essential components identified were: leading and deciding; interacting and presenting; organising and executing; and enterprising and performing. When comparing the components to those of Borman and Brush (1993) and Tett et al. (2000), as listed in Table 2.3, a clear relationship between the competencies is found.

Table 2.2

Comparison of Three Managerial Performance Models

SHL competencies	Borman & Brush (1993) competencies	Tett et al. (2000) competencies		
Leading and deciding	Planning and organising	Short-term planning	Coordinating	Self-development
Supporting and cooperating	Guiding, directing, and motivating subordinates and	Motivating by persuasion	Job enrichment	Developmental goal setting
Interacting and presenting	Training, coaching, and developing	Organisational awareness	Cooperation	Performance assessment
Analysing and interpreting	Communicating effectively and keeping	Decision delegation	Task focus	Political astuteness
Creating and conceptualising	Representing the organisation to	Motivating by authority	Compassion	Personal responsibility
Adapting and coping	Technical proficiency	Assertiveness	Directing	Seeking input
Organising and executing	Administration and paperwork	Problem awareness	Decision making	Public presentation
Enterprising and performing	Maintaining good working relationships	Developmental feedback	Goal setting	Oral communication
	Coordinating subordinates and other resources to get the job	Quantity concern	Cultural appreciation	Stress management
	Decision making/problem	Quality concern	Rule orientation	Strategic planning
	Staffing	Monitoring	Listening skills	Trustworthiness
	Persistence in reaching goals	Team building	Tolerance	Timeliness
	Handling crises and stress	Productivity	Customer focus	Professionalism
	Organisational commitment	Initiative	Sociability	Loyalty
	Monitoring and controlling resources	Technical proficiency	Financial concern	Written communication
	Delegating	Urgency	Orderliness	Adaptability
	Selling/Influencing	Decisiveness	Safety concern	Politeness
	Collecting and interpreting data	Creative thinking	Resilience	

Table 2.3

Relationship between Competencies

In-house competency model	Borman & Brush (1993) competencies	Tett et al. (2000) competencies	
Leading and deciding	Guiding, directing, and motivating subordinates and providing feedback	Directing	Decisiveness
	Decision making/problem solving	Decision delegation	Decision making
Interacting and presenting	Communicating effectively and keeping others informed	Oral communication	Public presentation
	Maintaining good working relationships	Professionalism	Sociability
Organising and executing	Coordinating subordinates and other resources to get the job done	Strategic planning	Coordinating
	Monitoring and controlling resources	Orderliness	
Enterprising and performing	Selling/Influencing	Creative thinking	Customer focus

2.4 PREDICTORS OF MANAGERIAL PERFORMANCE

2.4.1 Cognitive Ability

For quite some time, I/O psychologists believed that general cognitive ability (often referred to as general mental ability [GMA], *g*, or intelligence) was the single most important attribute that an individual possessed for successful job performance. The core of cognitive ability as a psychological construct is conceptualised as enabling people to solve problems, acquire knowledge, and apply reason to situations (Jensen, 1998). In an effort to reach consensus about the nature of intelligence, a group of 52 experts, including leading researchers in the field of psychological science, recently defined intelligence more meticulously as “a very general mental capacity that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience” (Gottfredson, 1997, p. 13). They continued to state that intelligence “is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capacity for comprehending our surroundings – ‘catching on,’ ‘making sense’ of

things, of ‘figure out’ what to do”. Ones, Dilchert, Viswesvaran and Salgado describe the impact of such a significant individual trait on individuals’ lives as follows:

Intelligence affects individuals’ lives in countless ways. It influences work lives of employees perhaps to a greater extent than any other individual differences trait. It determines whether an employee will be able to perform assigned tasks. It is the strongest predictor of learning and acquisition of job knowledge as well as overall job performance. It is remarkably relevant regardless of the occupation one holds. It even predicts extrinsic career success (i.e., earnings and promotions). As such, it is an exceedingly precious trait to include in employee selection systems. (2009, p. 2)

One of the most important qualities of cognitive ability is its information-processing skills that “can be applied to virtually any kind of content or context” (Gottfredson, cited in Ones et al., 2009, p. 7). The theory conceptually underpins Campbell’s (1990) statement that “general mental ability is a substantively significant determinant of individual differences in job performance for any job that includes information-processing tasks” (p. 56). Ones et al. (2009) agree with Campbell, explaining that “cognitive ability is an integral part in models of job performance due to its relation to knowledge and skill acquisition” (p. 7). The more cognitively demanding the knowledge to be acquired and the more complex the task to be performed, the greater is the relationship between cognitive ability and performance (Hunter & Hunter, 1984).

In all of the empirical findings presented above (e.g. Bertua, et al., 2005; Hunter & Schmidt, 1998) and both the general and specific cognitive abilities were validated against an overall training or job performance score. None of these studies include specific detailed measures of job performance and the performance of specific cognitive abilities relating to detailed performance measures is not determined. Would a specific cognitive ability such as numerical reasoning correlate higher with a person’s performance on the specialist knowledge requirements of his/her job or with the client relationship requirements? Murphy (2002) states that the implication of using specific abilities rather than the broader g is that people who do well on one ability might not exactly be the same as those who do well in other abilities. He

acknowledges that these individuals will overlap substantially due to the *g* factor subsuming the specific abilities.

Reeve and Hakel (2002, p.55) examine the importance of *g* and arguments for and against the *g* theory. One of these arguments is that for a given person, would *g* “differentially determine the development of performance capacities across domains?” They further state *g* has no within-person variance where these “intraindividual differences in the profile of specific cognitive abilities, interests, and personality may substantially influence one’s choice to allocate effort and cognitive resources in a specific domain.”

The nature of cognitive ability has been well defined and accepted. However, the structure of cognitive ability, which will be elaborated upon next, has not enjoyed such consensus.

2.4.1.1 Theoretical Underpinnings and the Structure of Cognitive Ability

The structure of cognitive ability has been the subject of much research. It was first studied by Spearman (1904), who distinguished between *g* (general cognitive ability) and *s* (specific cognitive abilities), holding that a general factor is measured in all cognitive ability tests, whereas one or more specific factors are unique to each test. Other views of the structure of cognitive ability have also been proposed (e.g., Cattell & Horn; Guilford; Sternberg; and Thurstone, cited in Ree, Carretta & Steindl, 2001), all of which reflect the dominant view of cognitive ability at a particular time in the 20th century.

The most recent and dominant contemporary approach to the structure of cognitive ability was presented by Carroll (2005). After reviewing and reanalysing hundreds of datasets, resulting from different groups of individuals taking multiple cognitive ability measures, Carroll (2005) hypothesised the three-stratum model. At the apex of the hierarchical structure is a general factor (*g*), which is believed to be responsible for some positive correlation among all the ability tests. At the second stratum are group factors, or broad abilities, such as prior acquisition of knowledge and the production of ideas. At the third level are the specific factors, or narrow abilities,

which represent each group factor. Individuals with similar general cognitive ability might differ with regard to their specific abilities, due to the differential ‘investment’ of their cognitive capacity in the narrow cognitive domains (Carroll, 2005). Carroll’s theory continues to serve as the reference point for research into the structure of cognitive ability.

2.4.1.2 Empirical Findings regarding the Predictiveness of Cognitive Ability

The central role that general cognitive ability plays in human functioning, as suggested earlier, has led many researchers to investigate the relationship between an individual’s cognitive ability and level of job performance (Landy & Conte, 2007).

Schmidt and Hunter (1998) reviewed over 85 years of research into the predictive validity of 19 different selection methods. Their meta-analysis concluded that cognitive ability tests have validity coefficients in the order of .51 for predicting job performance, which means that over 25% of the variance in performance across jobs is due to differences in cognitive ability. Their findings are summarised in Table 2.7.

In addition, Schmidt and Hunter (1998) also propose that cognitive ability is the most valid predictor of future job performance in cases where employees have no previous experience in a particular job. Much of the predictive power of cognitive ability is explained by the relationship between cognitive ability, the acquisition of job knowledge and job performance. Individuals with higher levels of cognitive ability tend to acquire new job knowledge more easily and quickly. They, therefore, develop a better understanding of how to do their jobs than do individuals with lower levels of cognitive ability.

It is evident that, for most jobs, general cognitive ability is the most important trait determinant of job and training performance (Schmidt & Hunter, 1998). The next question is whether such results are generalisable over cultures.

Salgado, Anderson, Moscoso, Bertua and De Fruyt (2003), using a large-scale meta-analysis of published research on the predictive validity of cognitive ability, reviewed over 250 European studies (combined N = 25 000), to determine the predictive

validity of general and specific cognitive abilities in predicting job performance. They found an operational validity of .62, noting that such a finding indicates that cognitive ability “is an excellent predictor of job performance” (p. 585). The validity of the five specific cognitive ability measures varied from .35 for verbal to .56 for memory (see Table 2.4).

Table 2.4

The Operational Validities (ρ) of General and Specific Cognitive Abilities for Job Performance (K = 250; N = 25 000)

	Job performance
General cognitive ability	.62
Verbal ability	.35
Numerical ability	.52
Spatial/Mechanical ability	.51
Perceptual ability	.52
Memory	.56

Source: Salgado et al., 2003.

In another meta-analytic study, Bertua, Anderson and Salgado (2005) analysed the predictive validity of general and specific cognitive abilities to predict job performance for seven occupations (managerial; professional; engineering; sales; clerical; operators; and drivers) (K = 283; combined N = 13 000). Their results also indicate that both cognitive and specific abilities are valid predictors of job performance for all occupations. They found an operational validity of .48 for cognitive ability, while the validity of the five specific ability measures varied from .35 for spatial to .50 for perceptual. In addition, the predictive validity for managerial performance (.69) is one of the highest (see tables 2.5 and 2.6).

Table 2.5

The Operational Validities (ρ) of General and Specific Cognitive Abilities for Job Performance ($K = 283$; $N = 13\ 000$)

	Job performance
General cognitive ability	.48
Verbal ability	.39
Numerical ability	.42
Perceptual ability	.50
Spatial ability	.35
Average	.42

Source: Bertua et al., 2005.

Table 2.6

The Operational Validities (ρ) of Cognitive Ability for Job Performance over Eight Occupational Groups ($K = 283$; $N = 13\ 000$)

	Job performance
Managerial	.69
Professional	.74
Engineering	.70
Sales	.55
Clerical	.32
Operators	.53
Drivers	.37
Miscellaneous	.40

Source: Bertua et al., 2005.

Once again, the conclusion is that general cognitive ability might be the best single predictor of job performance for a wide range of jobs and occupations, and even more so for jobs with a high level of complexity (Cartwright & Cooper, 2008; Ones et al., 2009; Salgado & Anderson, 2002; Schmidt & Hunter, 1998). Hunter and Hunter (1984) illustrated that the predictiveness of cognitive ability varies systematically as a function of job complexity. Using a sample of US studies, they estimated the validity of general cognitive ability for supervisor ratings of overall job performance to be .57

for high-complexity jobs, .51 for medium-complexity jobs, and .38 for low-complexity jobs. The results are consistent with those of Hunter and Schmidt (1996). When considering more specific roles, cognitive ability is much more predictive of a person's performance in professional/managerial roles (with aggregate validities of .58) than it is of a person's performance in unskilled jobs (with aggregate validities of .23). Up to 33% of managerial job performance was found to be accounted for by estimates of the manager's cognitive ability (Schmidt & Hunter, 1998).

Not only does cognitive ability predict managerial performance well, but also other outcomes, such as job mobility and promotion. Wilk, Desmarais and Sackett (1995, p. 84) noted that "individuals with higher cognitive ability move into jobs that require more cognitive ability and that individuals with lower cognitive ability move into jobs that require less cognitive ability". In Baydoun and Neuman's (1992) review of management selection techniques, they reported criterion-related validity of .53 for cognitive ability tests. They suggest that, of the various methods, cognitive ability testing provides the strongest criterion-based validity when used for managerial selection.

Even though cognitive ability has very important advantages in terms of its predictiveness of job performance, its use has a number of limitations. For instance, the use of cognitive ability tests presents a serious problem regarding adverse impact (Ones et al., 2009). Adverse impact leads to the selection of fewer minority group members, because minorities tend to obtain lower average scores on such tests. Differences are normally found in the mean ability scores of different racial and ethnic groups, with the scores being large enough to affect selection outcomes (Cartwright & Cooper, 2008; Hunter & Hunter, 1984). Consequently, I/O psychologists have moved away from using only cognitive ability tests to predict job performance towards also assessing other individual differences, such as physical abilities, personality, interests, knowledge, and emotion, when examining the behaviour of people in work settings.

Schmidt and Hunter's (1998) meta-analysis investigated the validity of, and utility for, various combinations of predictors. Their findings are summarised in Table 2.7. The three combinations with the highest multivariate validity and utility for job

performance were found to be cognitive ability, plus a work sample test (with a mean validity of .63); cognitive ability, plus an integrity test (with a mean validity of .65); and cognitive ability, plus a structured job interview, which partly measures conscientiousness and related personality traits, such as emotional stability and agreeableness (with a mean validity of .63).

Table 2.7

Predictive Validity (r) for Overall Job Performance of General Mental Ability (GMA) Scores Combined with a Second Predictor Using (Standardised) Multiple Regression

Personnel measures	Validity (r)	Multiple R	Gain in validity from adding supplement	% increase in validity	Standardised regression weights	
					GMA	Supplement
GMA tests	.51					
Work sample tests	.54	.63	.12	24%	.36	.41
Integrity tests	.41	.65	.14	27%	.51	.41
Conscientiousness tests	.31	.60	.09	18%	.51	.31
Employment interviews (structured)	.51	.63	.12	24%	.39	.39
Employment interviews (unstructured)	.38	.55	.04	8%	.43	.22
Job knowledge tests	.48	.58	.07	14%	.36	.31
Job try-out procedure	.44	.58	.07	14%	.40	.20
Peer ratings	.49	.58	.07	14%	.35	.31
Training & Experience behavioural consistency method	.45	.58	.07	14%	.39	.31
Reference checks	.26	.57	.06	12%	.51	.26
Job experience (years)	.18	.54	.03	6%	.51	.18
Biographical data measures	.35	.52	.01	2%	.45	.13
Assessment centres	.37	.53	.02	4%	.43	.15
T & E point method	.11	.52	.01	2%	.39	.29
Years of education	.10	.52	.01	2%	.51	.10
Interests	.10	.52	.01	2%	.51	.10
Graphology	.02	.51	.00	0%	.51	.02
Age	-.01	.51	.00	0%	.51	-.01

Source: Schmidt & Hunter, 1998, p. 265.

From Table 2.7, it is also clear that cognitive ability tests assess only one aspect of the many skills, abilities and aptitudes that determine job performance. Such a conclusion corresponds with Murphy's suggestion that different attributes of people serve different demands of the job (cited in Landy & Conte, 2007, p. 90). It could, therefore, be meaningful to use other primarily non-cognitive selection measures (such as personality and Situational Judgement Tests), in conjunction with cognitive ability tests, in order to assess the other skills and abilities that contribute to successful job performance. The predictive validity of such measures will be explored in the following section.

2.4.2 Personality

Personality can be defined as “a relatively stable set of characteristics, tendencies, and temperaments that have been significantly formed by inheritance and by social, cultural, and environmental forces” (Ivancevich & Matteson, 1993, p. 121). In the simplest terms, personality is the typical way in which an individual responds to events and people (Landy & Conte, 2007). Personality is considered a trait, because it is fairly stable, even though situations and circumstances might lead a person to behave differently.

2.4.2.1 The Structure of Personality

For many decades, researchers did not view personality as a valid predictor of job performance, because no well-accepted taxonomy existed for classifying personality traits. Without a well-accepted taxonomy it is not possible to determine whether there are consistent, meaningful relationships between particular personality constructs and performance criteria in different occupations. However, significant advances in defining the structure of personality have been made over the past few decades.

Cattell used a lexical approach to explore the factors of personality, entailing the examination of the most common adjectives descriptive of personality. Using factor analysis, he developed a relatively complex taxonomy of individual differences, consisting of 16 primary factors and 8 second-order factors. His taxonomy formed the basis of his well-known and widely used assessment test, the Sixteen Personality

Factor Questionnaire (16PF). As Cattell's results could not be duplicated, his method of organisation was discarded. Nevertheless, he provided the theoretical groundwork for much of the current research in the measurement of personality (cited in Barrick & Mount, 1991, p. 2).

Costa and McCrae (1992) reanalysed Cattell's 16PF and developed the Five Factor Model (FFM) with five higher order personality factors. The Big Five personality factors and their definitions are presented in Table 2.8. The FFM is the most widely accepted structure of personality variables, with nearly all recent research using the FFM as the taxonomy for organising personality factors (e.g., Barrick & Mount, 1991; Hough & Dilchert, 2009; La Grange & Roodt, 2001; Tett & Christiansen, 2007; Tett et al., 1991). Evidence of the model has been growing over the past few decades, with Salgado et al. (2003) recently finding that personality measures based on the FFM produce higher validity coefficients for predicting job performance than do other measures.

Table 2.8

The Big Five Personality Factors

Factor	Definition
Neuroticism (low emotional stability)	A tendency to easily experience unpleasant emotions, such as anxiety, anger, or depression.
Extraversion	Energy, surgency, and the tendency to seek stimulation and the company of others.
Openness to experience	Appreciation for art, emotion, adventure, and unusual ideas; a tendency to be imaginative and curious.
Agreeableness	A tendency to be compassionate and cooperative, rather than suspicious and antagonistic, towards others.
Conscientiousness	A tendency to show self-discipline, act dutifully, and aim for achievement.

Source: Adapted from Costa & McCrae, 1992.

Hough and Dilchert (2009), however, believe that the FFM is “an inadequate taxonomy of personality variables for I/O psychology to build knowledge and understand the determinants of work behaviour and performance” (p. 3). After an extensive review of the literature, they proposed a nomological-web clustering approach to the development of personality constructs that are conceptually and empirically similar. According to the approach, the cluster analysis of the patterns of relationships of target variables with other variables (including variables outside the personality domain, such as job performance criteria) are needed to identify the homogeneous personality constructs that are characterised as having similar nomological nets. Hough and Dilchert (2009) called for other researchers to refine the taxonomy to gain a better understanding of the interrelationships between personality and performance.

While the FFM has, and will continue to have, its critics, it has done a great deal to advance our understanding of the personality construct. The next section examines the predictiveness of the Big Five personality factors.

2.4.2.2 Empirical Findings on the Predictiveness of the Big Five Personality Factors

The predictive validity of personality measures has been investigated by numerous researchers. Recent meta-analytic studies of personality and job performance have confirmed that the Big Five can serve as a useful tool in personnel selection, although some of the factors have proven to be more useful than others (e.g., Barrick & Mount, 1991; Barrick, Mount & Judge, 2001; Borman, Penner, Allen & Motowidlo, 2001; Rothmann, Meiring, Van der Walt & Barrick, 2002; Tett et al., 1991).

Conscientiousness has been found to be the strongest and most consistent predictor of performance across occupational groups (Barrick & Mount, 1991; Barrick et al., 2001; Salgado, 1997; Schmidt & Hunter, 1998; Tett, Jackson, Rothstein & Reddon, 1994). A conscientious person is described as being responsible, prudent, goal-oriented, organised, persistent, self-disciplined, and achievement-oriented (Costa & McCrae, 1992). Corrected correlations between conscientiousness and job performance have been found to be .31 (Schmidt & Hunter, 1998) .22 (Barrick &

Mount, 1991) and .34 (Ones, Viswesvaran & Schmidt, 1993). Conscientiousness is beneficial for performance in most job settings for two reasons. First, the variable that links conscientiousness and job performance is job knowledge. Individuals who are highly conscientious presumably put time and effort into acquiring high levels of job knowledge, and hence tend to perform better than those who are less conscientious (Schmidt & Hunter, 1998). Second, individuals who are highly conscientious exhibit a greater tendency for setting performance-related goals, which, in turn, facilitates higher levels of job performance than is the case with those who are less conscientious (Barrick, Mount & Strauss, 1993).

Another personality trait that has been quite consistently, although not as strongly, correlated with job performance is emotional stability (i.e., low neuroticism) (Barrick et al., 2001). The contribution of emotional stability to job performance is attributed to the self-confidence, resilience and calmness of emotionally stable individuals (Costa & McCrae, 1992). Other researchers suggest that the effects of emotional stability on job performance might be as general as those of conscientiousness (Salgado, 1997). Conscientiousness and emotional stability have also been shown to have incremental validity over cognitive ability measures (Salgado & Anderson, 2002; Salgado et al., 2003).

The other personality factors are related to job performance, but only for some occupations or certain criteria. Openness to experience was found to be related to training proficiency across all occupations, though neither for job proficiency nor for personnel data. Barrick and Mount (1991) argue that being broad-minded, curious and cultured leads to positive attitudes toward learning experiences in general, and might, therefore, be useful in identifying those individuals who are most likely to benefit from training programmes. Results for agreeableness indicate that it is not an important predictor of job performance, even for jobs with a large social component, such as sales and management. Extraversion, however, was found to be a valid predictor (across the criterion types) for both such occupations. Personal characteristics, such as energy, talkativeness, assertiveness, and the tendency to seek stimulation and the company of others, therefore lead to effective performance in such jobs. For both occupations, however, the estimated true score correlations are less than .20 (Barrick & Mount, 1991).

Therefore, only two of the Big Five personality dimensions, namely conscientiousness and extraversion, have proven to be valid predictors of managerial performance (Barrick & Mount, 1991; Barrick et al., 2001).

Hogan and Holland (2003) found that all Big Five personality dimensions predict relevant job performance criterion variables, with estimated true validities of .43 (neuroticism/emotional stability), .35 (extraversion), .34 (agreeableness), .36 (conscientiousness), and .34 (openness to experience). The meta-analytic results indicate that, if the traits and specific behaviour at work that are theoretically related are measured reliably, the correlation between the two is $r = .20$ and $r = .50$. When specific demographic factors and measures of cognitive ability are added to this, up to two-thirds of the variance on any work measures may be accounted for (Furnham, 2008).

Not only do personality measures appear to be valid predictors of job performance, but the meta-analytic evidence suggests that the personality assessments have significantly less adverse impact on minority groups than do measures of cognitive ability variables. When Hough, Oswald and Ployhart (2001) summarised studies that examined the mean-score differences between Whites and various other ethnic minorities, they found essentially no differences for most personality variables.

Despite evidence of the apparent advantages of personality measures in personnel selection, the use of such measures is not without its critics. Concerns have been raised about the extent of honesty with which job applicants answer personality inventories. Hogan, Hogan and Roberts (1996) have argued that personality inventories represent self-presentations, in terms of which responses to items indicate how the respondent would like to be seen. The research suggests that applicants, whether they are conscious of being so or not, are motivated to present the image most likely to be viewed positively by the decision makers concerned. The researchers, therefore, recommend that personality tests are used only in addition to the use of other selection instruments (Fletcher, 1991; Hough, 1998).

2.4.3 Situational Judgement Tests

SJTs are designed to assess an applicant's judgement regarding a situation encountered in the workplace (Weekley & Ployhart, 2006). The typical SJT presents test-takers with a series of job-related scenarios in written, verbal, or visual form. They are asked to indicate, from a set of options, how they would handle the scenario in question (Clevenger et al., 2001; Lievens et al., 2008; Weekley & Ployhart, 2005). The applicants are either required to select the action that they would most likely, or least likely, perform, or to rank order, or rate, the effectiveness likely to emanate from taking alternative courses of action (Cullen, Sackett & Lievens, 2006; McDaniel & Nguyen, 2001; McDaniel et al., 2007; Motowidlo, Hooper & Jackson, 2006). Applicants' responses are interpreted as direct indicators of how they would handle the situation if it were actually to occur on the job (Motowidlo et al., 1990; Weekley & Ployhart, 2005). Because applicants are neither placed in a simulated work setting, nor are asked to perform the task or behaviour described (as would be the case in an assessment centre or with a work sample), the SJTs are classified as low-fidelity simulations (Clevenger et al., 2001; Motowidlo et al., 1990), although video-based SJTs are considered to have higher fidelity than have written SJTs (McDaniel et al., 2006). Asking applicants to respond to realistic scenarios is an attractive form of testing, in that it seeks out information about judgement in context, in contrast with the decontextualised nature of many standardised tests (Cullen et al., 2006).

Although SJTs have a long history, dating back to the 1920s, they have recently increased in popularity as a predictor of job performance (Bergman, Drasgow, Donovan & Henning, 2006; Clevenger et al., 2001; Lievens et al., 2008; McDaniel et al., 2001; Weekley & Ployhart, 2006; Weekley, Ployhart & Harold, 2004; Whetzel & McDaniel, 2009). The increased popularity of SJTs is due to research showing that such tests have significant criterion-related validity and possess incremental validity over and above cognitive ability and personality tests (Chan & Schmitt, 2002; Clevenger et al., 2001; McDaniel et al., 2001; McDaniel et al., 2006; McDaniel et al., 2007; O'Connell et al., 2007; Weekley & Jones, 1997, 1999), while usually producing lower levels of adverse impact and more positive test-taker reactions (Clevenger et al., 2001; Hanson & Borman, 1995; Landy & Conte, 2007; McDaniel et al., 2006; Motowidlo et al., 1990; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996).

In contrast, researchers have raised a few concerns about the use of SJTs. First, the answers that are provided often describe 'best practice' behaviour, rather than what the respondent would actually do, given the real-life situation (Creighton & Scott, 2006). Second, some SJTs do not present a sufficient range of responses that might be assumed to the scenario in question. The respondents might, therefore, be forced to select actions or responses that do not necessarily fit their behaviour. If they find such a limitation of choice to be frustrating, the validity of the measures might be adversely affected (e.g., Chan & Schmitt, 2005; Ployhart & Harold, 2004). Third, the description of the scenarios in many SJTs tends to be brief, with the result that candidates do not become fully immersed in them. Such a limitation removes some of the intended realism of the scenario, reducing the quality and depth of assessment that can be obtained by such means.

Consensus clearly needs to be reached about whether SJTs are, in fact, useful personnel selection tools. The next section deals in more detail with the psychometric properties and the predictiveness of SJTs.

2.4.3.1 Empirical Findings on the Psychometric Properties and Predictiveness of SJTs

This section provides a brief discussion of the research conducted into the psychometric properties and predictiveness of SJTs. (Refer to section 2.2 for a more detailed discussion of the concepts.)

2.4.3.1.1 Reliability

Due to the considerable variance in the findings of different studies, McDaniel et al. (2001) conducted a meta-analysis to provide an objective and precise summary of the SJT literature. They found that the internal consistency reliability (Cronbach's alpha) of SJTs varied between .43 and .94. Their research identified various factors that moderate such variability in the internal consistency reliability of SJTs. Ployhart and Ehrhart (2003) examined the impact of six different types of instructions on the psychometric characteristics of SJTs. They found that instructions involving rating

(e.g., 'Rate how likely...' or 'Rate how effective...') led to the highest internal consistency, ranging from .65 to .73. Instructions with two response alternatives ('most/least likely to do' or 'most/least effective') had somewhat lower internal consistency, ranging from .30 to .65. Instructions with only one choice ('have done in past' and 'should do') had the lowest internal consistency, ranging from .24 to .65.

In Ployhart and Ehrhart's (2003) study, as the construct domain was relatively homogeneous, internal consistency was found to be the appropriate form of reliability. However, for SJTs of a multidimensional nature, estimating the reliability is problematic, because the scale and item heterogeneity makes Cronbach's alpha an inappropriate reliability index (Whetzel & McDaniel, 2009). Test-retest reliability has been suggested as a better measure for assessing reliability (Hough & Dilchert, 2009; McDaniel & Whetzel, 2005). Ployhart, Porr and Ryan (cited in Lievens et al., 2008) reported test-retest reliability of .84, with earlier studies by Bruce and Learner (cited in Lievens et al., 2008) and Richardson, Bellows, Henry and Co (cited in Lievens et al., 2008) finding test-retest reliabilities that ranged from .77 to .89. Such studies indicate that the test-retest reliability of SJTs is adequate when using the criterion of acceptable reliability of .8 of Nunnally (1978).

Parallel-form reliability is another option, but is often infeasible, because it requires the use of different item content to measure the same constructs. Because it is difficult to identify the particular constructs assessed using SJTs, construct equivalence across forms can be problematic (Whetzel & McDaniel, 2009). One exception is that of Chan and Schmitt (2002), who estimated parallel-form reliability at .76.

Due to such test development and data collection problems, many researchers continue to provide internal consistency estimates, while acknowledging that they are underestimating the reliability of SJTs. More research is needed on the appropriate methods by means of which to assess the reliability of SJTs (McDaniel & Whetzel, 2005).

2.4.3.1.2 Face Validity

Face validity refers to how well a test ‘on the face of it’ appears to measure what it is supposed to measure (Murphy & Davidshofer, 2005). Applicants generally prefer selection tools which they perceive to be job-related. It is, therefore, not surprising that applicants perceive SJTs to have high face validity (Clevenger et al., 2001; Lievens et al., 2008).

The extent to which an SJT’s face validity is affected by the medium in which the assessment is delivered is not clear. Research, in laboratory contexts, comparing applicants’ reactions to a video-based assessment with reactions to a paper-and-pencil version of the same assessment, indicates that the video-based version is perceived to be more face valid than the paper-and-pencil version (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan & Drasgow, 2000). Lievens and Sackett (2006), however, conducted the same comparison, but in an actual selection context, and found no significant difference between the two formats.

2.4.3.1.3 Construct-related Validity

Construct-related validity tests hypothesise about the relationships existing between measures and their constructs (Schmitt & Chan, 1998). Some selection measures can be clearly defined with respect to the constructs that they measure. For example, cognitive ability tests measure the construct of general cognitive ability, and conscientiousness tests measure the construct of conscientiousness. However, other selection measures, such as interviews and assessment centres, are best classified as measurement methods that simultaneously measure multiple constructs. Distinguishing between methods and constructs is critical to the understanding of SJT research (Whetzel & McDaniel, 2009). Some researchers argue that SJTs measure a single construct, while others have shown that SJTs measure multiple constructs, such as personality and cognitive ability.

McDaniel and Nguyen (2001) meta-analytically explored the evidence concerning the relationship between SJTs and the Big Five personality dimensions. They found that SJTs correlated with agreeableness (observed mean: $r = .25$), conscientiousness (observed mean: $r = .26$), and emotional stability (observed mean: $r = .31$). However, the SJTs did not correlate with extroversion (observed mean: $r = .06$) and openness to experience (observed mean: $r = .09$).

McDaniel et al. (2001), in contrast, meta-analytically explored the relationship between SJTs and cognitive ability ($K = 79$, $N = 16984$). The results indicated that SJTs had a mean correlation of .46 with cognitive ability. There was, however, substantial variance around this estimate, such as where SJTs based on job analysis had higher correlations with cognitive ability (.50) than those not based on job analysis (.38). In addition, SJTs with less detailed questions had higher correlations with cognitive ability (.56) than those with more detailed questions (.47). Other researchers similarly found that written SJTs had higher correlations with cognitive ability than did video-based SJTs (Lievens & Sackett, 2006; McDaniel et al., 2006; Weekley & Jones, 1997)

McDaniel et al. (2007) conducted another meta-analysis that expanded on the work of both McDaniel and Nguyen (2001) and McDaniel et al. (2001) with more data and the inclusion of a response–instruction moderator. The purpose of the meta-analysis was to determine whether the SJT response instructions operate as a moderator of the construct and the criterion-related validities of SJTs. Their results, as indicated in Table 2.9, indicate that an SJT’s response instructions do influence the constructs measured by the test.

Table 2.9

Meta-Analytic Results of Correlations between Situational Judgement Tests and Cognitive Ability, Agreeableness, Conscientiousness, and Emotional Stability

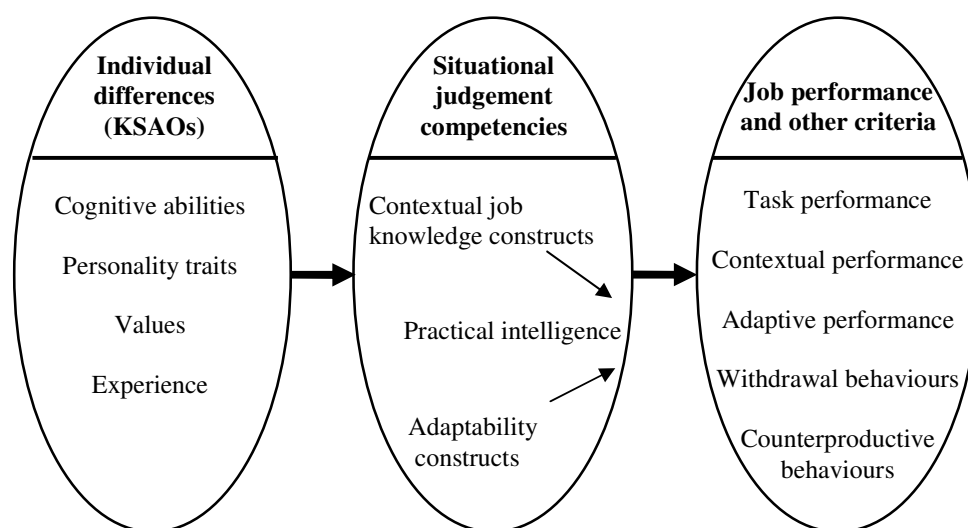
Distribution of correlations with SJTs	<i>N</i>	<i>No. of rs</i>	<i>ρ</i>
Cognitive ability	22 553	62	.39
Behavioural tendency instructions	5 263	21	.23
Knowledge instructions	17 290	41	.43
Agreeableness	14 131	16	.33
Behavioural tendency instructions	5 828	11	.53
Knowledge instructions	8 303	5	.20
Conscientiousness	19 656	19	.37
Behavioural tendency instructions	5 902	11	.51
Knowledge instructions	13 754	8	.33
Emotional stability	7 718	14	.41
Behavioural tendency instructions	5 728	10	.51
Knowledge instructions	1 990	4	.11

Source: McDaniel et al., 2006, p. 191.

SJTs with knowledge instructions (which asked the respondents to evaluate the effectiveness of possible responses to a given situation) had higher correlations with cognitive ability (.43 versus .23). In contrast, SJTs with behavioural tendency instructions (which asked the respondents to identify how they would likely behave in a given situation) had higher correlations with personality constructs, such as conscientiousness (.51), emotional stability (.51) and agreeableness (.53), than did SJTs with knowledge instructions (.33, .11, and .20 respectively). From such results, McDaniel et al. (2006) concluded that SJTs with knowledge instructions assess primarily cognitive ability, with some personality, and that SJTs with behavioural tendency instructions assess primarily personality, with some cognitive ability. The findings suggest that the construct validity of a SJT can be changed by altering the response instructions (McDaniel et al., 2006; McDaniel et al., 2007; Whetzel & McDaniel, 2009).

While such researchers seek to find a correlation between SJTs and specific constructs, other researchers argue that SJTs are merely measurement methods that can be designed to measure a variety of constructs. For example, to measure interpersonal constructs, a test with various interpersonal situations can be developed (Bledow & Frese, 2009; Chan & Schmitt, 1997; Hough & Dilchert, 2009; Lievens et al., 2008; McDaniel & Nguyen, 2001; McDaniel & Whetzel, 2005; O'Connell et al., 2007; Weekley & Jones, 1999; Whetzel & McDaniel, 2009). Based on such a premise, Lievens, Buyse and Sackett (2005) examined the use of an SJT aimed at testing interpersonal skills for making college admission decisions. The aim was to measure skills other than cognitive ability. They found positive correlations ($r = .21, p < .05$) between the students' SJT scores and their performance on the interpersonal course, which was part of their curriculum. Lievens et al.'s (2005) study demonstrated that SJTs are not designed to replace traditional cognitive ability tests, but, rather, to increase the coverage of skills not measured by traditional predictors.

Chan and Schmitt recently extended their research about what SJTs measure by adding adaptability to their hypothesis (cited in Landy & Conte, 2007, p. 144). They believe that various KSAOs generate competencies related to tacit knowledge and adaptability or practical intelligence, which, in turn, lead to higher levels of job performance. When studying their model in Figure 2.2, recognition of the relationship between KSAOs and practical intelligence assists in clarifying why SJT, cognitive ability, and personality scores correlate with one another. Furthermore, the model also serves to clarify why SJTs add incrementally above any one or combination of those attributes. This is because "the attributes support the development of tacit knowledge and adaptability but are different from any of those supporting KSAOs" (Landy & Conte, 2007, p. 144). McDaniel and Nguyen's (2001) research supports the model, in that they have found that SJT scores increase with increasing years of experience and that it is reasonable to presume that tacit knowledge and adaptability increase with experience.



Source: Landy & Conte, 2007, p. 145.

Figure 2.2 Framework for Relating the Multidimensional Nature of SJTs to KSAOs and Job Performance

In summary, because of its multidimensional nature, the extent to which SJTs measure different constructs seems to vary greatly. Despite a recent trend to develop SJTs to measure specific constructs, such as personal initiative (Bledow & Frese, 2009), researchers also agree that SJTs measure broad job knowledge, and even practical intelligence (Landy & Conte, 2007; Weekley & Ployhart, 2005). Lievens et al. (2008) argue that such a finding is no surprise, as “SJT items may refer to a wide range of situations and include different types of content to which applicants must attend when making a decision” (p. 433).

2.4.3.1.4 Criterion-related Validity

Criterion-related validity refers to the degree of effectiveness with which performance on a test predicts performance in a real-life situation (Murphy & Davidshofer, 2005). Such predictiveness is of utmost importance, since tests that are highly predictive of job performance serve to improve the accuracy of selection decisions and lead to enhanced selection utility (Boudreau, 1984).

Based on the basic tenet of behavioural consistency; that past performance is the best predictor of future performance, Wernimont and Campbell (cited in Motowidlo et al., 1990) argue that job simulations, such as SJTs, can be very useful in predicting job performance, probably even more so than measures of cognitive ability and personality.

Motowidlo et al. (1990) argue that high-fidelity simulations should be superior predictors of future work performance, since such simulations more closely resemble actual work conditions than do low-fidelity simulations. In order to determine how much fidelity is necessary for a simulation to be sufficiently predictive, they developed and validated a low-fidelity simulation test to predict managerial performance. Correlations between the test and various job performance criteria ranged from .20 to .40, which indicates that even low-fidelity simulations can predict performance.

McDaniel et al. (2001) conducted a meta-analysis of 102 validity coefficients to determine the criterion-related validity of SJTs. They found that SJTs have substantial validity ($p = .34$) for the prediction of job performance that was generalisable. On reanalysing and updating their 2001 data, McDaniel et al. (2003) found that the knowledge-response instructions yielded higher validity (.33) than did the behavioural tendency instructions (.27). Such results compare well with the validity of the best predictors of job performance, such as cognitive ability.

Similarly, Chan and Schmitt (2002) hypothesised that an SJT would positively predict task performance, contextual performance and overall job performance. Their results showed that the correlations between the SJT and the relevant performance criteria were all positive and statistically significant ($p < .05$).

In another study, Becker (2005) developed an SJT of employee integrity and examined whether scores on the SJT predicted integrity-relevant outcomes. In validating the SJT, he found the scores to be valid predictors of integrity-relevant outcomes in real-world settings.

Lievens and Sackett (2006) examined the difference in predictive validity between video-based, versus written, SJTs. On changing an existing video-based SJT to a written one, while keeping the verbal content constant, they found that the video-based version had higher predictive validity.

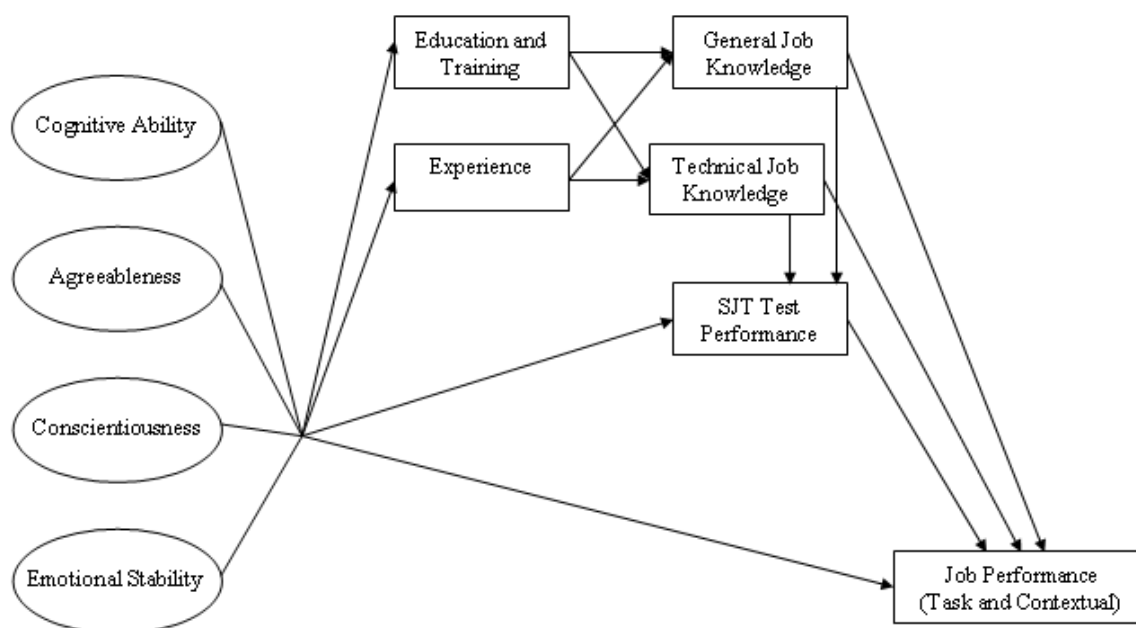
In summary, the criterion-related validity of SJTs in predicting performance seems well established, both in terms of the meta-analysis conducted by McDaniel et al. (2001) and in terms of the various primary studies conducted since Motowidlo et al. (1990) revived interest in the method. The next section describes McDaniel et al.'s (2006) model, which integrates both the construct and criterion-related validities for SJTs.

2.4.3.1.5 An Integrated Model of the Construct and Criterion-related Validity Evidence for SJTs

McDaniel et al. (2006) summarised the aforementioned construct and criterion-related validity evidence of SJTs in the light of the following model (see Figure 2.3). Their model illustrates that performance on SJTs is influenced by four individual differences: cognitive ability; emotional stability; conscientiousness; and agreeableness. The extent to which SJTs and measures of such constructs correlate varies, as such correlation is moderated by the SJT's response instructions. The individual differences influence the level of job knowledge that is obtained through education and training, as well as experience. Job knowledge is divided into general and technical job knowledge, in order to separate specific job/industry knowledge from the knowledge applicable to most jobs.

McDaniel et al. (2006) describes general job knowledge as "composed of basic knowledges common to most jobs and [which] might be viewed as work socialisation knowledge" (p. 194). The knowledges to which they refer include, for example, the value of showing up for work, being nice to colleagues, and refraining from inappropriate behaviour. McDaniel et al. (2006) continue by saying that general job knowledge may also include knowledge needed at a management level, such as how to deal with subordinates. Some SJTs are specifically designed to measure general supervisory

knowledge. Technical knowledge, in contrast, is “specific to a job or an industry and is gained through education, training, and experience” (McDaniel et al., 2006, p. 194). Some SJTs are designed to measure technical knowledge, such as selling skills. Although little research has, so far, reported on the knowledge correlates of SJTs, they are generally assumed to measure knowledge, whether tacit or explicit (McDaniel et al., 2006).



Source: McDaniel et al., 2006, p. 195.

Figure 2.3 Conceptual Model of the Factors affecting the Validity of SJTs

Furthermore, the model suggests that test performance on SJTs is a function of the four individual differences, both directly and through their influence on job knowledge, as mediated by education, training and experience. In addition, general and technical job knowledge also seems to influence test performance on SJTs directly. Such a deduction corresponds with Lievens et al.’s (2005) and Weekley and Ployhart’s (2005) findings that, although SJTs correlate with other constructs, they also increase the coverage of skills not measured in such a way. Job performance is, therefore, predicted by cognitive ability, emotional stability, conscientiousness, agreeableness, SJTs, and job knowledge.

McDaniel's (2006) model also holds that, since SJTs measure individual differences to varying degrees, the incremental validity of a specific SJT will vary, according to its correlates. For example, when an SJT correlates significantly with cognitive ability, it will probably provide little incremental validity over cognitive ability but substantial incremental validity over personality and vice versa. Therefore, SJTs with significant cognitive and non-cognitive correlates will probably provide little incremental validity over a battery of both cognitive ability and personality tests.

The model, furthermore, suggests that job performance can be measured with varying emphasis on task and contextual performance. McDaniel et al. (2006) argue that knowledge-based SJTs better predict task performance than contextual performance. In contrast, behavioural tendency SJTs are expected to better predict contextual performance. Such an assertion is consistent with the research of Chan and Schmitt (2002), who studied the correlations between SJTs and task performance, contextual performance and overall job performance.

2.4.3.1.6 Incremental Validity

Landy and Conte (2007) define incremental validity as "the value in terms of increased validity of adding a particular predictor to an existing selection system" (p. 148). They continue by explaining that an important principle underlying incremental validity is that "in assessment is that it is not which tool to use, but what combination of tools to use for the greatest predictive ability at the lowest cost" (Landy & Conte, 2007, p. 148). Therefore, a key question concerning the usefulness of SJTs concerns their incremental prediction ability over that of other selection tests. As a measurement method of a multidimensional nature, SJTs can assess various constructs relating to job performance. They can, therefore, be expected to measure other aspects of performance, which cannot be measured by specific tests, such as cognitive ability and personality, therefore adding incrementally to such tests.

Weekley and Jones (1997, 1999) examined whether SJTs have incremental validity over job experience and cognitive ability. Results for three different SJTs indicated that they accounted for 2.5%, 5.7% and 9.6% of the variance in performance. In two additional studies, they found the SJTs to provide incremental validity of 3.3% and 1.1%.

Clevenger et al. (2001) extended Weekley and Jones' (1997) work by including measures of job knowledge and conscientiousness. For three different samples, the results indicated that, although SJT scores correlated strongly with cognitive ability, the SJT added incrementally to the prediction in two of the three samples. The incremental validity afforded by the SJT in the two samples was statistically significant ($R^2 = .016 - .028$).

The efforts by Clevenger et al. (2001) were extended, by Chan and Schmitt (2002), by including the remaining four of the Big Five personality traits. In addition, their study examined the validity of SJTs in the prediction of overall job performance, as well as in three performance dimensions: core technical proficiency, job dedication, and interpersonal facilitation. They found the relative incremental validities of the SJT across the set of eight predictors to be .24 for predicting core technical proficiency, .30 for job dedication, .17 for interpersonal facilitation, and .21 for overall job performance.

In addition, McDaniel et al. (2007) found in their recent meta-analysis, that SJTs provide 3% to 5% incremental validity over cognitive ability, 6% to 7% over personality and 1% to 2% over both cognitive ability and personality. Such increments in predictive validity are meaningful, since they represent likely improvements in the accuracy of selection decisions, as well as in selection utility (Boudreau, 1984).

From the literature review, it is evident that SJTs predict job performance, that the observed validities of situational judgment tests compare well with those of cognitive ability and personality tests, and that their use is reported to result in significantly lower adverse impact. Most importantly it is evident that SJTs provide incremental prediction over and above cognitive ability and personality. The next section proposes a research

model and explains the hypothesised relationships between the various predictors chosen for the current study, as well as between the predictors and criterion measures.

2.5 PROPOSED RESEARCH MODEL

Considering the empirical research findings and arguments noted above, it is evident that various individual differences directly and indirectly affect managerial job performance. However, various questions remain unanswered within the existing body of knowledge and, therefore, require investigation in the present research study. The research initiating question, then, which arises is, “How does SJT performance affect managerial performance when used in conjunction with measures of cognitive ability and personality?”

The present research attempts to develop and test a model of various individual differences that theoretically relate to managerial performance. The focus of the present research differs from previous published studies since it investigates managerial performance specifically as a criterion construct, instead of using a broad measure of general job performance.

In the same vein, this study attempts to shed light on the construct validity of SJT measures. It is apparent from research literature that the meaning of SJT scores remains in dispute. Therefore, we address this question by developing the nomological network surrounding SJT performance, through investigating the relationship of SJT performance with scores obtained from other predictor measures, i.e. cognitive ability and personality. For this purpose, a model was constructed to investigate the interrelationships among the variables in the study. The hypothesised relationships between the various predictors, as well as between the predictors and criterion measures, are depicted in Figure 2.4.

The proposed research model was constructed after careful consideration of (a) the published literature findings on the construct and criterion-related validity of predictors in the model, as well as (b) the results of the job analysis process, which outlined certain

key person characteristics required for successful managerial job performance in the present research context.

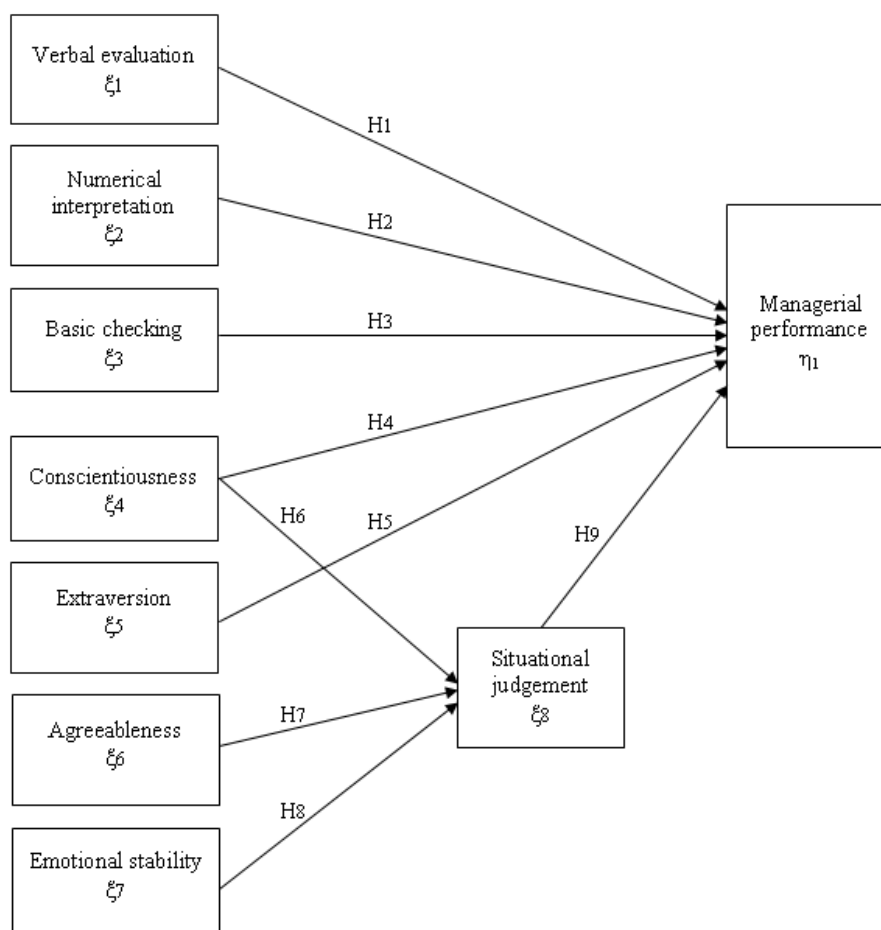


Figure 2.4 Hypothesised Relationships between the Dependent and Independent Variables.

In line with the research models reviewed earlier (i.e., Bartram et al., 2002, Borman & Brush, 1993, Tett et al., 2000) managerial job performance is deemed to be a direct function of various groups of individual differences, namely cognitive ability factors (H1 to H3), as well as personality traits, specifically those of conscientiousness and extraversion (H4 to H5). In addition, the criterion-related validity evidence on SJT (cited earlier) strongly suggests that SJT performance also directly affects managerial performance (H9).

Cognitive ability influences on managerial job performance: It is well accepted in the literature that cognitive ability creates variance in job performance in general, as well as in managerial performance in particular. In fact, Schmidt and Hunter (1998) found that up to 33% of managerial job performance is accounted for by estimates of managers' cognitive ability. Baydoun and Neuman (1992) support such a conclusion by stating that cognitive ability testing provides the strongest criterion-based validity when used for managerial selection.

Even though the bulk of prior research into cognitive ability predictors of managerial job performance used general cognitive ability measures, an alternative would be to investigate *specific cognitive abilities* as predictors. The cognitive measures used in the present research were limited to measures of specific cognitive abilities by default, since these were the measures used for selection. These specific abilities were identified as being critical to job success during the job analysis and competency modelling process that was followed. In the present research, the job analysis identified areas of managerial job performance, such as *interacting and presenting*, and *analysing and interpreting*, that require managers to have a minimum level of verbal and numerical abilities. A second reason why selected specific abilities were chosen in the present selection research was because of the clear links between the constructs measured and the task at hand: managers routinely use their ability to evaluate verbal (written) information, to interpret financial reports and to execute a control function by overseeing the correctness of client records, legal and fiduciary documentation and others. Therefore, three measures of specific cognitive abilities were used for selection that represented core personal requirements for effective branch managers. *Verbal evaluation* refers to the ability to understand the meaning of words and to use them effectively. *Numerical interpretation* refers to the ability to understand numerical relations and to use numbers effectively and *basic checking* refers to the speed and accuracy of checking at a basic level (refer to section 3.5.3 for a detailed description of the measurement of the specific cognitive abilities).

It was, therefore, hypothesised that the three specific cognitive abilities – *verbal evaluation* (H1), *numerical interpretation* (H2) and *basic checking* (H3) – impact directly on managerial performance. Since these measures represent specific abilities which will be included as single measures (and not as a composite score) in this research, we did not expect the obtained correlations with the overall job performance measure to be large. Also, we would not expect within higher levels of managerial performance that basic cognitive abilities such as basic checking would be related to managerial performance.

Personality influences on managerial job performance: Apart from cognitive ability, personality factors are likely to influence managerial performance, since they influence a person's work-related behaviour. Both the research literature and in-house job analysis process suggest that only two of the Big Five factors might be expected to influence managerial job performance directly. *Conscientiousness* (H4) is expected to be positively related to managerial performance, because the former assesses personal characteristics (such as responsibility, carefulness, the ability to plan, persistence, the capacity to work hard and the ability to follow instructions and procedures) which are important attributes for accomplishing managerial tasks. Due to the frequent interactions of managers with other employees and customers, as well as their responsibility for influencing others in their leadership and supervisory roles, another personality trait that is expected to be positively related to managerial performance is that of *extraversion* (H5). The quality of extraversion relates to personal characteristics, such as energy, talkativeness, assertiveness, and the tendency to seek stimulation and the company of others. Such proposed relationships are supported by Barrick and Mount's (1991) research, which found that both conscientiousness and extraversion directly impact on managerial performance.

The excluded Big Five personality factors are not expected to be related to managerial performance. Since executing existing policies and procedures represents a critical part of a bank manager's core task, such execution does not allow for much independent thinking, creativity, or open-mindedness. Therefore, neither openness to experience nor agreeableness is expected to be related to managerial performance. Agreeableness

represents an individual's interpersonal orientation of likeability and cooperativeness, neither of which is an ideal characteristic for managers to have. Managers are required to lead and make decisions based on what is the best in a certain situation, rather than to be influenced by their need for acceptance. This argument is supported by the results of various meta-analyses that investigated the relationship between agreeableness and managerial performance. These studies found very low correlations (Barrick & Mount, 1991; McDaniel & Nguyen, 2001; Tett & Christiansen, 2007) between agreeableness and performance. Emotional stability, in contrast, can be argued to relate to managerial performance, since the characteristic is related to the ability to form and maintain positive relationships in one's work environment. Viewed from the negative pole, neurotic individuals are nervous, highly strung, stress prone, moody and insecure. Such traits tend to inhibit positive motivational tendencies. That is, individuals who spend time worrying about their performance, doubt their abilities, require assurance from others, and are depressed are generally unable to develop adequate coping strategies and cannot focus attention on the tasks at hand. The former traits do not facilitate effective job performance. Barrick and Mount (1991) argue that such individuals are, over time, often selected out of the labour pool altogether. They found very low correlations for emotional stability and managerial performance, arguing that, as long as an individual possesses sufficient emotional stability, the predictive value of any differences is minimised.

Mediating affect of individual differences on performance on the SJT: Furthermore, the model suggests that *Conscientiousness* (H6), *Agreeableness* (H7) and *Emotional Stability* (H8) influence managerial performance indirectly through its effects on performance on the SJT. It is argued that personality influences the judgements that people make about appropriate and inappropriate courses of action, and that SJTs, at least in part, capture such personality constructs (Clevenger et al., 2001). Therefore, it is expected that such personality constructs indirectly impact on managerial performance. Such an expectation is consistent with Weekley and Ployhart's (2005) findings that SJTs partially served to mediate the effects of personality on performance. In addition, McDaniel and Nguyen's (2001) meta-analytical exploration of the relationship between SJTs and the Big Five personality dimensions found that SJTs correlated with

agreeableness, conscientiousness, and emotional stability, which also pointed to the possibility of mediating effects on performance.

No significant relationship is expected between the specific cognitive abilities and performance on the SJT. It is evident from the literature that, depending on the response instruction, general cognitive ability which represents higher order cognitive functioning, such as abstract thinking, deductive reasoning and general cognitive complexity, correlates highly with SJTs (McDaniel & Nguyen, 2001). However, the cognitive ability tests used in the current study measured lower level basic specific abilities (*verbal evaluation, numerical interpretation and basic checking*) and not general intelligence. The nature of the SJT used in the present study measures knowledge of appropriate interpersonal behaviour and problem solving in daily managerial scenarios, so that it does not provide the opportunity to measure verbal, numerical or checking abilities. In other words, due to the absence of a clear conceptual link between the predictor construct and the constructs measured by the SJT, we also do not expect a statistical relationship between these variables.

Influences of performance on the SJT on managerial job performance: In the suggested model, we propose that performance on the SJT (H9) is expected to directly affect managerial performance. The first reason why we expect this effect relates to the bandwidth of the SJT measure, i.e. the ‘broadness’ with which it measures the underlying construct or constructs. Although SJTs are believed to measure personality and general cognitive ability to varying degrees (McDaniel et al., 2006), it is argued that SJTs as a measurement method also measure a variety of other constructs, such as broad job knowledge or practical intelligence (Landy & Conte, 2007). SJTs can assess job-related skills that remain untapped by other measures, ranging from skills relating to conflict management, interpersonal communication, problem solving, negotiation, and teamwork facilitation (Chan & Schmitt, 1997; Lievens & Sackett, 2007; McDaniel & Whetzel, 2005; O’Connell et al., 2007; Weekley & Jones, 1999). Evidence to date indicates that SJTs are valid predictors of performance, especially for managerial positions in which

interpersonal interactions are important (Motowidlo et al., 1990), perhaps due to the broad 'bandwidth' with which it measures person attributes.

Apart from "bandwidth" as a reason why SJT performance could predict managerial performance, "fidelity" may also play a role. The second reason why we expect the hypothesised H9-path (i.e, the correlation between performance on the SJT and managerial job performance) is due to the high fidelity of the SJT used in the current study. Essentially, the SJT was developed to represent a content- and face-valid measure of managerial job performance, almost to the extent that it represents a 'work-sample' of managerial judgment in interpersonal situations. Furthermore, since the SJT is presented in video format, it enables the respondents to see the scenario unfold, including the behaviour of the individuals involved, as if they were making real-life observations. Presentation formats with high fidelity are more reflective of actual job experiences, and are, therefore, likely to be more predictive of on-the-job behaviour. For example, an SJT presented in video format was found to have higher predictive validity than the same SJT presented in written format (Lievens & Sackett, 2006). For these reasons, we expect a significant and positive relationship between SJT performance and managerial job performance measures.

Incremental validity of the SJT, relative to personality and cognitive ability, in the prediction of managerial performance: Since the SJT conceptually represents constructs not tapped by the other predictors in the model, as well as its high degree of conceptual overlap with the criterion measure, we expect SJT performance to increment the validity of basic cognitive ability and personality measures in predicting managerial performance. Furthermore, consistent with McDaniel et al.'s (2006) model, the current model suggests that the incremental validity of the SJT measure's scores will vary according to its correlates. In other words, if the SJT correlates significantly with the personality factors, though probably providing little incremental validity over personality, SJT scores will most likely provide substantial incremental validity over cognitive ability, and vice versa. Therefore, we expect the addition of SJT scores to regression models already containing measures of basic cognitive abilities and personality to significantly

add to the ability of the model to explain variance in criterion performance, i.e. managerial job performance.

2.6 CONCLUSION: CHAPTER TWO

Chapter two dealt with the most important theories of job performance and managerial performance. Thereafter, the literature on, and empirical findings of, cognitive ability, personality and situational judgement for predicting managerial performance was examined. Such an examination provided the foundation for the research model and research hypothesis. Chapter three provides an explanation of the research methodology used for empirically testing the plausibility of the hypothesised model.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1 INTRODUCTION

Chapter three outlines the specific research process that was followed in examining the correlations and incremental validity of an SJT with alternative predictors in the prediction of managerial performance. The chapter includes the research design, sample, measuring instruments, and the statistical analysis.

3.2 RESEARCH DESIGN

A correlative *ex post facto* design was used to obtain the objective of the study. The design observes the independent and dependent variable across individuals to establish the extent to which they co-vary (Babbie & Mouton, 2001), and involves the prediction of a dependent/criterion variable (Shavelson, 1981). The term *ex post facto* means ‘after the fact’ and is used to describe the relationship between variables as they naturally occur, without any intervention (Gravetter & Forzano, 2003). Gravetter and Forzano (2003) suggest that the design is most suitable when variables, which are impossible or unethical to manipulate, are measured. In the current study, the independent variables are situational judgement, cognitive ability and personality. The manifestations of these variables are already evident in the branch managers who were the subjects of this study. Therefore, the existing levels of the variables are not manipulable.

Some important concerns about the use of the design should be noted. Huysamen (1980) warns about the risk of inappropriate interpretation. His concern is that, although it may be possible to establish causality in some situations, the results might only support associations. Furthermore, Christensen (1985) cautions that the correlation between two variables can exist due to the existence of a third variable, which caused both of them, and not because of their interpreted causal relationship. It is, therefore, recommended that

the results of the data obtained by means of the application of such a design be interpreted with caution (Kerlinger, 1992).

The specific research design is the concurrent validity design. Concurrent validity is established when a new instrument and an established valid instrument, which measure the same phenomenon, are administered to the same subjects at the same time, with the results of the new instrument correlating with the results of the established valid instrument (McKenzie, Mitchell & Oliver, 1995). The information collected is used to describe the population at that time. Concurrent validation is simpler, less expensive, and not as time consuming as predictive validation (Groth-Marnat, 2003). Concurrent validity is especially useful when there are poor tests of the same criterion, on which the new test attempts to improve. In such cases, the results will probably indicate significant, though modest, correlations (Kline, 1986). Concurrent validity, on its own, is clearly not entirely satisfactory. To accept a test as valid, further evidence is needed, in the form of other validity evidence (Kline, 1986).

3.3 HYPOTHESES

In the light of the literature review and the proposed relationships between the variables, the following hypotheses have been formulated:

Hypothesis 1:

A significant positive relationship exists between a manager's verbal evaluation ability and managerial performance.

$$H_01: \rho[X1, Y1]=0$$

$$H_a1: \rho[X1, Y1]>0$$

Hypothesis 2:

A significant positive relationship exists between a manager's numerical interpretation ability and managerial performance.

$$H_{02}: \rho[X2, Y1]=0$$

$$H_{a2}: \rho[X2, Y1]>0$$

Hypothesis 3:

A significant positive relationship exists between a manager's basic checking ability and managerial performance.

$$H_{03}: \rho[X3, Y1]=0$$

$$H_{a3}: \rho[X3, Y1]>0$$

Hypothesis 4:

A significant positive relationship exists between a manager's conscientiousness and managerial performance.

$$H_{04}: \rho[X4, Y1]=0$$

$$H_{a4}: \rho[X4, Y1]>0$$

Hypothesis 5:

A significant positive relationship exists between a manager's extraversion and managerial performance.

$$H_{05}: \rho[X5, Y1]=0$$

$$H_{a5}: \rho[X5, Y1]>0$$

Hypothesis 6:

A significant positive relationship exists between a manager's conscientiousness and his/her performance on a Situational Judgement Test.

$$H_{06}: \rho[X4, X8]=0$$

$$H_{a6}: \rho[X4, X8]>0$$

Hypothesis 7:

A significant positive relationship exists between a manager's agreeableness and his/her performance on a Situational Judgement Test.

$$H_{07}: \rho[X6, X8]=0$$

$$H_{a7}: \rho[X6, X8]>0$$

Hypothesis 8:

A significant positive relationship exists between a manager's emotional stability and his/her performance on a Situational Judgement Test.

$$H_{08}: \rho[X7, X8]=0$$

$$H_{a8}: \rho[X7, X8]>0$$

Hypothesis 9:

A significant positive relationship exists between a manager's performance on a Situational Judgement Test and his/her managerial performance.

$$H_{09}: \rho[X8, Y1]=0$$

$$H_{a9}: \rho[X8, Y1]>0$$

Hypothesis 10:

H₀10: Managerial job performance cannot be reliably predicted from a battery consisting of measures of specific cognitive abilities (verbal evaluation, numerical interpretation and basic checking), personality (conscientiousness and extraversion) and an SJT.

H_a10: Managerial job performance can be reliably predicted from a battery consisting of measures of specific cognitive abilities (verbal evaluation, numerical interpretation and basic checking), personality (conscientiousness and extraversion) and an SJT.

Hypothesis 11:

H₀11: Measures of situational judgement will not provide incremental validity in the prediction of managerial performance relative to the predictability provided jointly by cognitive ability and personality measures (conscientiousness and extraversion).

H_a11: Measures of situational judgement provide incremental validity in the prediction of managerial performance relative to the predictability provided jointly by cognitive ability and personality (conscientiousness and extraversion).

Hypothesis 12:

H₀12: The SJT will not mediate (i.e., partially or fully) the effects of conscientiousness (H6), agreeableness (H7), and emotional stability (H8) on managerial performance.

H_a12: The SJT will mediate (i.e., either fully or partially) the effects of conscientiousness (H6), agreeableness (H7), and emotional stability (H8) on managerial performance.

3.4 PARTICIPANTS AND PROCEDURE

A sample can be described as “...a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005, p. 120). From the various sampling techniques that are available (Babbie & Mouton, 2001), the non-probability convenience sampling method was chosen for the current study, as the researcher sought to obtain as many respondents as possible. According to Gravetter and Forzano (2003), convenience sampling relies on the availability of potential subjects and their willingness to respond. They warn that, even though the method employed is the most commonly used sampling method in psychological research, caution must be exercised in generalising from the data. Babbie and Mouton (2001) urge researchers to alert their readers of the risks associated with such a method and to provide detailed information about the sample. Even though convenience sampling was used, the sample represented a substantial portion of the population of branch managers of the bank in question (38%), and, therefore, the results of the study can safely be generalised to all branch managers in the participant organisation.

The current study was conducted in a South African retail bank. The sample consisted of 124 branch managers, who were geographically spread throughout the country. The branch managers managed the daily operations of the branch, as well as their branch staff. The data were collected by the bank's human resources (HR) personnel and a supervising industrial psychologist during an annual regional branch manager conference. The head of operations allowed all branch managers the time to complete the SJT. All supervisors (regional managers) were also given the opportunity to complete three different measures of managerial performance (see section 3.5.1 for more detail) for each of the branch managers reporting to them. As a result, the SJT and BOS data were collected concurrently. All the participants were assured that their responses would be kept confidential, that the responses would be used only for research purposes, and that such responses would not be recorded in the branch managers' personnel files or used for any personnel decision purpose. The branch managers' cognitive ability and personality data were retrieved from the HR database, as these scores were placed on file as part of the selection process.

For a detailed description of the sample, see Table 3.1. The sample consisted of a group of 124 branch managers. Nine (7.3%) of the branch managers did not complete their demographic profile. Of the group, 69 (55.6%) were male and 46 (37.1%) were female. The ethnic composition of the group consisted of 65 (52.4%) Black African, 3 (2.4%) Indian, 28 (22.6%) Coloured and 19 (15.3%) were White. The composition of the sample in terms of first language preference was Afrikaans (26.6%), English (16.1%), and Tswana (13.7%). The educational level of the sample ranged from Grade 12 (62.1%) to a degree (8.9%). The age of the branch managers varied between 24 and 48 ($M = 32.89$, $SD = 5.474$) and their months of service ranged between 5 and 150 ($SD = 30.454$).

Table 3.1

Demographic Profile of the Sample

GENDER		
RESPONSES	Frequency	Percentage
Female	46	37.1
Male	69	55.6
Unreported	9	7.3
RACE		
RESPONSES	Frequency	Percentage
Black	65	52.4
Indian	3	2.4
Coloured	28	22.6
White	19	15.3
Unreported	9	7.3
HOME LANGUAGE		
RESPONSES	Frequency	Percentage
Afrikaans	33	26.6
English	20	16.1
Ndebele	1	0.8
North Sotho	6	4.8
Sotho	1	0.8
South Sotho	6	4.8
Swazi	1	0.8
Tsonga	2	1.6
Tswana	17	13.7
Venda	5	4.0
Xhosa	10	8.1
Zulu	13	10.5
Unreported	9	7.3
EDUCATION		
RESPONSES	Frequency	Percentage
Grade 12 (Matric)	77	62.1
Post-Matric Certificate	12	9.7
Diploma	15	12.1
Degree	11	8.9
Unreported	9	7.3

Table 3.1 Continued

MONTHS OF SERVICE		
VARIABLE	Mean (years)	Standard deviation
Months of Service	50.34	30.454
AGE		
VARIABLE	Mean (years)	Standard deviation
Age	32.89	5.474

Note. $N = 124$.

3.5 MEASURING INSTRUMENTS

The following measuring instruments were chosen for the purposes of this quantitative empirical study. All non-propriety measures, such as the demographic questionnaire and Behaviour Observation Scale, are provided in appendix B.

3.5.1 Managerial Performance

The criterion measure for managerial job performance in the current study was constructed by formulating a definition of the construct of managerial job performance, followed by operationalising its measurement by constructing subjective assessment rating measures to measure the latent construct.

In order to measure managerial performance, three different measures were developed and administered, ultimately being combined into a single composite measure. The three measures that were used included a Behaviour Observation Scale (BOS) completed by supervisors, Overall Performance Ratings (OPR) completed by supervisors, as well as a Performance Ranking (Perf Rank) assignment of branch managers by the supervisor within each region.

The performance ratings were collected by the bank's HR personnel. The questionnaires were administered during rating sessions with groups of supervisors (10 to 15 supervisors in each group) at the regional branch manager conferences referred to above. Ratings were included in the research only if the supervisor (i.e., regional manager) rating the participant (i.e., branch manager) had had sufficient opportunity (i.e., at least three months) in which to observe the branch manager's performance. Rater training was provided prior to each rating session. The rater training included: (a) explanation and discussion about the scales and the competency model on which it is based (b) the provision of strategies by means of which to avoid common rating errors; (c) granting the assurance that the ratings, which would be used for research purposes only, would be kept entirely confidential; and (d) the holding of a discussion of the importance of the project to the organisation as a whole. In summary, we expected the rater training to impart a common frame-of-reference for rating the behaviour of branch managers.

Behaviour Observation Scale (BOS): Firstly, a self-constructed Behaviour Observation Scale (BOS) was developed to establish the subjective ratings by the supervisors (see appendix B). The BOS was used to assess the eight competencies required by the branch managers, as identified in the WPS (see appendix A for the WPS report). After being developed by the researcher, the BOS was reviewed by a panel of subject matter experts within the organisation in respect of the use and clarity of language and the content validity of the scale. After changes had been made to incorporate their inputs, consensus was reached about the representativeness of the items concerned. The BOS consisted of 16 items, rated on a Likert scale, ranging from 1 (*almost never*) to 5 (*almost always*). Examples of the behavioural competencies involved were provided under each item to guide the raters. For example, under the item *Makes effective decisions* the following behavioural indicators are given: *Makes well-informed, objective decisions; Uses principles, values and sound business sense to make decisions; Takes time to fully understand the implications before making a decision; Takes responsibility for decisions.* A participant's final score on the BOS measure was calculated by determining the mean score for all BOS items.

Overall Performance Rating (OPR): In addition to rating each branch manager on the BOS, the supervisors (regional managers) were also requested to assign an Overall Performance Rating (OPR) to each of their branch managers, by responding to the following statement ‘Rate this individual’s overall job performance as a branch manager, by assigning a rating from 1 (*highest possible score*) to 10 (*worst possible score*)’. The overall ratings were linearly transformed (inverted) so that the participant’s score for OPR ranged from 1 (lowest possible score) and 10 (highest possible score).

Performance Ranking (Perf Rank): Since the supervisory ratings of job performance are prone to low variability in higher level positions (Guion, 1998), we also requested the supervisors to provide a Performance Ranking assignment (*from top performer to lowest performer*) of branch managers under their supervision. It was also included as an alternative performance measure, since Bartram (2007) indicated how validity can be increased by used forced-choice criterion measures. A participant’s ranking score was calculated by processing a decimal (RankingDec) to allow for the fact that the supervisors have varying numbers of branch managers reporting to them. Thus, if a branch manager was rank-ordered to be position 2/9, his/her ranking was processed to a score of 0.222. Thereafter, the resulting figure was inverted by subtracting the decimal from 1 ($1 - 0.222 = 0.778$) to allow the candidates with higher rankings to receive reciprocal higher values.

Managerial Performance Composite (MPerf Comp): In order to provide a single overall measure of job performance for the regression analyses, it was decided to form a composite of the three separate ratings explained above, since the provision of such a composite would provide the best single overall indication of managerial performance while, at the same time, avoiding potential limitations inherent in each type of rating (e.g., see Gatewood & Feild, 1994). The Managerial Performance Composite was calculated by forming an unweighted linear composite of all three job performance measures, using an additive combination of total scores from each measure. Prior to this composite formation, we confirmed the interrelationships between the variables by inspecting their intercorrelations (all were high) and conducting linear z-score transformations to equate scale ranges and make scores comparable.

It was decided to include the specific performance measures in conjunction with the composite measure in order to explore the effect of using different performance measures on predictive validity of the predictor set, especially considering Bartram's (2007) research indicating that validity can be increased by using forced-choice criterion measures over normative measures.

The psychometric properties of job performance measures have received considerable research attention. For example, Hunter and Schmidt (1996) explain that criterion measures should exhibit reliability in order to allow for their appropriate use. In their research, Chan and Schmitt (2002) assessed the reliability of scores of measures of three dimensions of job performance, using subjective supervisory rating items. They found internal consistency reliabilities (Cronbach's alpha) of .82 for task performance, .78 for motivational contextual performance and .76 for interpersonal contextual performance.

Moreover, the factorial validity of job performance measures have received considerable research attention, as explained earlier in the literature review. We based the formulation of a single composite measure of job performance on the finding that a single general factor of job performance can normally be observed to underlie various job performance measures (Viswesvaran, Schmidt, & Ones, 2005). An assessment of this assumption, as well as detailed results obtained for the psychometric qualities assessed by the BOS and other job performance criterion measures used in the study, are reported in chapter four.

3.5.2 Situational Judgement Test (SJT)

The specific Situational Judgement Test (SJT) used in the current study was a managerial judgement assessment called the Video-based Simulation (VBS). The VBS was developed for a large telecommunications company in order to determine the managerial potential of thousands of applicants for first-line managerial positions (Kriek, 1994). The VBS consists of 25 video vignettes of situations to which managers are typically exposed on a daily basis. An example of the transcribed dialogue from one such situation from the test, lasting about 40 seconds, follows:



Colleague: Would you please sign this invoice for us, Mr. Jacobs?
Manager: R2 100?
Colleague: Yes, we need cartridges for the laser printers.
Manager: What for?
Colleague: This is for an important job that must go out by the end of the week.
Manager: I'm sorry, Sydney, I can't do that. I can only approve amounts of up to R2 000.
Colleague: Please, we need your help.
Manager: I can't really do it, you know... I need the permission of the manager of the section.
Colleague: This is very important and very urgent.
Manager: He will only be back next week... Oh what the heck!... Let's get the job done and take the flak later!

(The scenario stops here.)

The questions to be answered, based on the scenario, were:

(1) I can identify myself with the way in which the manager handled the situation and see myself acting the same in similar situations:

(i) Yes (ii) Possibly (iii) Uncertain (iv) Improbably (v) No

(2) The manager took a good decision in this situation:

(i) Yes (ii) Possibly (iii) Uncertain (iv) Improbably (v) No

(3) In a similar situation, I would:

- (i) act the same way as the manager did.
- (ii) thoroughly examine the document first.
- (iii) look for a responsible person to give approval.
- (iv) not sign the document.
- (v) ask for two order forms.

It is clear from this example that the items resort into three different categories within each vignette situation. The first type of item (Item 1) requires the respondent to indicate the degree to which his or her behavioural intention would be similar to that depicted by the actor portraying the manager in the video vignette. The second type of item (Item 2) requires a judgment of the quality of the decision taken by the actor and, therefore, requires the respondent to apply knowledge and experience related to the situation to evaluate the adequacy of the depicted response. The third type of item (Item 3) requires the respondent to choose two best alternatives from a list of possible reactions that describe the respondent's likely behavioural response. Therefore, the item types assess different aspects of judgment related to the scenario. It is clear from this description that the item types correspond well with the distinction between SJT items that measure *behavioural intention* (e.g., Item type 1 and 3) or *knowledge instruction* (Item type 2) (Whetzel & McDaniel, 2009). McDaniel and Nguyen (2001) identified two classes of SJT response items: knowledge and behavioural tendency. They explain that knowledge response instructions request respondents to choose the correct or best possible response

or, alternatively, to rate the effectiveness of responses. The second type, i.e., behavioural tendency response instructions require the respondent to select the option that indicates what the respondent would likely do or to rate the likelihood that they would perform an action. Since they found that these response instructions moderated the validity of SJTs, we opted to combine scores from both types of items contained in the VBS SJT-measure, since we expect combined scores to mitigate the moderating effect of using either.

The SJT measure (VBS) was developed by identifying a general task list and typical critical incidents that were likely to be encountered by first-line managers, based on a detailed job and task analysis utilising integrated data from quantitative and qualitative sources. The task analysis identified the following nine generic task dimensions for managers: maintenance of discipline; conflict management; negotiation skills; time management; decision making; prioritisation; participative management; motivating of staff; and sensitivity. The ability to handle the generic task dimensions successfully was, according to the subject matter experts (SMEs) used in development of the measure, the key behavioural skill needed for success as a first-line manager (Kriek, 1994).

The validity of the SJT measure (VBS) was considered throughout its development process. Based on the task dimensions and typical critical incidents listed above, typical scenarios and work situations were developed. Relevant questions relating to content and process were formulated for each work situation, with the assistance of subject matter experts and focus groups. The content validity of the VBS was increased by the involvement of managers as work specialists in determining the content of the instrument. Qualitative feedback received from the participants was also positive with regard to the face validity of the VBS. Criterion-related validity of the VBS was examined with the use of a concurrent validity design. For this purpose, the job performance levels of first-line managers was determined with the use of a supervisor rating questionnaire which was developed from the job analysis results. Statistical analysis of the concurrent validity indicated a correlation of $r = .48$ between VBS scores and the criterion measure of supervisor ratings of job performance ($p < 0,001$). Internal consistency reliability estimates were calculated and indicated a coefficient alpha of .87 (Kriek, 1994). Such

results compare well with the findings of Motowidlo et al. (1990) and McDaniel et al. (2001), who reported correlations of between .28 and .37 with important criterion measures.

The SJT measure (VBS) is usually scored by means of a computer programme that assigns scores according to correspondence between responses and ‘correct’ responses, as determined by SMEs (Kriek, 1994). Since a fair amount of time had elapsed between initial development and the present administration, we opted to revise the scoring key as an additional control measure. The scoring key used by the computer programme was developed in a different company and could have also become outdated since its initial development in 1994. We revised the scoring key by presenting the VBS scenarios to an expert panel consisting of regional managers, organisational development consultants, as well as one recruitment consultant specialising in branch manager recruitment (N = 11). The panel was requested to study the SJT instructions, items, and response options, and indicate the ‘correct’ responses for each of the VBS scenarios, using their own experience and judgment. Interrater agreement¹ across the scale type items (i.e., Items 1 and 2 within each of the 25 scenarios) approached moderate agreement (> .50; LeBreton et al., 2008) (mean $r_{WG} = .48$). The r_{WG} index of interrater agreement normally assumes values ranging from 0 (perfect lack of agreement) to 1 (perfect agreement). However, this value could have been attenuated by the relatively small number of judges (LeBreton et al., 2008) as well as diversity of judges, resulting in differing views about the most desirable response option. The percentage agreement for the items that requested respondents to choose an appropriate course of action was not calculated, since two options could be chosen instead of only one correct option. However, the frequencies of response options chosen as the correct option within each scenario were inspected and results showed that the majority of items showed either one or two of the alternatives drawing the majority (> 50%) of responses.

¹ James, Demaree, and Wolf’s (1984, 1993) interrater agreement index r_{WG} is arguably the most frequently used index of interrater agreement (LeBreton et al., 2008). This index defines agreement in terms of the proportional reduction in error variance, using the ratio of observed variance taken over K different judges or raters and variance expected when there is a complete lack of agreement.

The test items used to develop the coding scheme were the same as the items used for the branch manager sample, i.e. each scenario consisted of two sections, namely (a) two items requesting both respondents' agreement with the actor's behaviour and an expression of the degree to which they would follow the same course of action as that of the actor, and (b) choosing two alternative courses of action from a list of four to five options. For the first two items that measured the respondent's degree of agreement with the behaviour enacted by the 'manager' in the vignette (Item 1: *behavioural intention*, and Item 2: *knowledge instruction*) we determined the ideal level of agreement (i.e., correct response) by calculating the mean score of subject matter experts on each item. Since response options on both items (Item 1 and 2, within each vignette) varied from 1 (*high agreement*) to 5 (*high disagreement*), we expect that respondents that chose responses closely matching those of experts would be deemed to be better judges of the correct response. Therefore, absolute deviation scores (i.e., similar to LeBreton, Senter, & Jenell, 2008) were calculated for each set of two items (in each vignette) and summed to a total score for each respondent, resulting in a total maximum score of 150 (25 vignettes X 2 items X 4²) for this part (i.e., *behavioural intent* and *knowledge instruction* agreement) of the SJT. For the second part of each scenario that related to the choice of alternative course of action we determined the correct response alternatives by calculating the frequencies of alternatives indicated as being the best response options. The two response options that received the highest frequencies were assumed to reflect the best choices amongst the alternatives. Each respondent's total score on the SJT was calculated by forming a weighted linear composite of 'accuracy' indices across item types, i.e. for the first two items within each scenario (agree-disagree responses) accuracy scores were assigned that were the average of the absolute of the deviation score across items, as well as for "choose an option identified as correct option" or "0 = did not choose a correct option", summed across the 25 scenarios.

Since SJTs typically exhibit a multidimensional nature (Chan & Schmitt, 2002), the internal consistency reliability is expected to be low. In such cases, the test-retest reliability has been suggested as a better measure for assessing the reliability involved

² The maximum deviation score per item is four scale points, when rounded. For example, $5-1 = 4$.

(Hough & Dilchert, 2009; McDaniel & Whetzel, 2005). However, due to the limitations of the current research, measuring the test–retest reliability was not possible in this case.

3.5.3 Cognitive Ability

To assess cognitive ability, a set of three specific cognitive ability tests, which were developed by SHL, were used. The tests, rather than measuring general cognitive ability, measure specific abilities, according to the competency profile for branch managers. Such specific abilities include their ability to evaluate verbal and written information, to interpret financial reports and to execute a control function, by overseeing the correctness of client records, legal and fiduciary documentation and others. These widely used measures were completed as part of each branch manager’s selection process over the past five years. Their scores were retrieved from the branch managers’ personnel files. They completed the following SHL cognitive ability tests:

Verbal evaluation was assessed using the 60-item multiple choice format VC1.1. The VC1.1 is designed to assess the ability to understand the meaning of words and to use them effectively. This test can be used for those with Grade 12 or higher, up to graduate level. The test consists of a series of passages, each of which is followed by several statements. The respondents are required to evaluate each statement in the light of the passage and to indicate whether the statements are true, false or need to be supplemented with additional information. An example passage is given below:

The successful company, Lane Finances, is pinning a good deal of its investment faith on property. Of the company’s investment, 40% is in property, which, according to the company account, is worth R120 million, but has a much higher current market value of R200 million. It is largely concentrated in high rental Johannesburg office property, and altogether, according to the Chairman, the total market value of all the company’s investments is R600 million.

This statement is followed by four statements, which the respondent is required to evaluate:

- (1) Over half of the company's investments are in areas other than property.
- (2) The company has invested heavily in oil exploration activities.
- (3) The company accounts do not indicate the current market value of the property.
- (4) Much of Lane Finances' property is in Johannesburg.

Reliability estimates of internal consistency ranging from .87 to .91 have been reported for mixed occupational positions for a sample of 788 respondents in the finance industry (SHL, 2003a), and for a sample of 5 508 respondents in the mixed industry sector (SHL, 2003b), respectively. A criterion-related validity coefficient of .51 was reported for 72 middle managers in the public service industry (SHL, 1999).

Numerical interpretation was assessed using the 35-item multiple choice format NCC2. The NCC2 is designed to measure the ability to understand and use numerical data in order to answer questions. Respondents are required to use the tables provided in order to solve basic numerical principles. An example item follows:

According to the chart, how many houses were sold in the quarter July to September inclusive?				
A	B	C	D	E
86	194	202	208	210

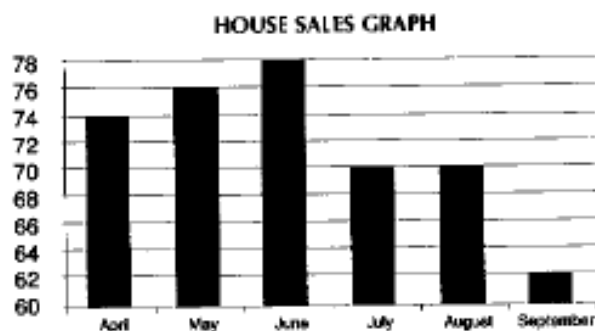


Figure 3.2 Sample Item from the NCC2

Internal consistency reliabilities of .86 and .87 have been reported for mixed occupational positions for a sample of 2 966 respondents in the finance industry (SHL, 2003c), and for a sample of 12 211 respondents in the mixed industry sector (SHL, 2003d), respectively.

Basic checking was assessed using the 80-item multiple choice format CP7.1, which measures the speed and accuracy of checking at a basic level. The format is predominantly used for those positions that require routine checking. Routine checking is an important part of the branch managers' job, because such a position entails checking multiple financial reports. Internal consistency reliabilities of .94 and .93 have been reported for mixed occupational positions for a sample of 1 379 respondents in the finance industry (SHL, 2003e) and for a sample of 9 665 respondents in the mixed industry sector (SHL, 2003f), respectively. A criterion-related validity coefficient of .14 was reported for 51 air traffic controllers (SHL, 2004).

3.5.4 Personality

To measure the Big Five personality factors, the applicable scales were selected from SHL's (2006) Occupational Personality Questionnaire (OPQ32). The OPQ32 is a self-report personality questionnaire that is widely used in the selection and development of employees. The OPQ32 model is an occupational model of personality, which describes 32 dimensions of people's preferred or typical style of behaviour at work. It is

particularly appropriate for use with professional and managerial groups, although the model deals with personality characteristics that are important in a wide variety of roles (SHL, 2006).

The OPQ32 model follows the general OPQ model of personality, which breaks personality down into three domains: *relationships with people*, *thinking styles* and *feelings and emotions*. The three domains are joined by a potential fourth, the *dynamism domain*, which is composed of scales such as Vigorous, Achieving and Competitive, which relate to sources of energy. The OPQ model provides its users with a clear framework for interpreting complex patterns of personality (SHL, 2006).

There is both a normative and ipsative version of the OPQ32. The ipsative version was used in the current study, as it is more resistant to faking (Martin, Bowen & Hunt, 2002) and is used most widely throughout the world (Bartram & Brown, 2004). The OPQ32i has 416 items, arranged in 104 blocks of four, in a forced choice format. Respondents have to choose one item that is “most like me” and one item that is “least like me” in each block.

The OPQ was “mapped onto the Great Eight factor level of the SHL Universal Competency Framework” (Bartram, 2007, p. 265) and was not designed to measure the Big Five. The second order combination of the Big Five factors explains only 50% of the total variance of the OPQ. Furthermore, only 25 of the 32 scales are conceptually related to the Big Five factors. This indicates that a second order factor model (the Big Five factors) and the OPQ is not a good fit (SHL, 2006).

Even though the OPQ was not designed to measure the five-factor model, various studies have shown that the OPQ32 subscales can be translated to the five-factor model (Beaujouan, 2000; Ferguson, Payne & Anderson, 1994; Matthews & Stanton, 1994; SHL, 1999). Therefore, measures of the Big Five personality factors were constructed, by creating a weighted linear composite for each Big Five personality factor from the relevant subscales, based on earlier published evidence of the factor loadings of OPQ32

scales on Big Five dimensions from concurrent validity studies (Bartram & Brown, 2005; SHL, 2006).

Evidence supporting the criterion-related validity of the OPQ has been reported in a number of studies across a range of industry sectors and job types (e.g., Robertson & Kinder, 1993; Saville, Sik, Nyfield, Hackston, & MacIver, 1996). The use of the OPQ32 is, furthermore, supported by reports of the adequacy of its internal consistency reliability. In one study, the alpha coefficients of 21 scales exceeded .70 (SHL, 2000a), while, in another, the alpha coefficients of 26 scales exceeded .70 (SHL, 2000b).

The psychometric properties (i.e., the composite reliability) of the Big Five composites are reported in chapter four.

3.6 STATISTICAL ANALYSIS

A combination of univariate and multivariate statistical techniques was used to analyse the data and to investigate the research hypotheses. Descriptive statistics were used specifically to investigate the means and standard deviations (SDs), in order to shed light on the central tendency and variability of variables in the study, namely for exploratory data analysis (EDA). Internal consistency reliability analysis and exploratory factor analysis were used mainly to investigate the psychometric measurement properties of the instruments concerned.

Inferential statistics (consisting of correlation and hierarchical regression analysis) were used to investigate the hypotheses (Field, 2005). The Statistical Package for the Social Sciences (SPSS) (Field, 2005) was used to analyse the data.

The Pearson product-moment correlation coefficients were computed to determine the degree of linear relationship between cognitive ability (H1 to H3), personality (H4 to H5) and SJT (H9) and managerial performance scores, as well as between personality (H6 to H8) and SJT scores. Pearson's correlation ranges from +1 to -1. A correlation of +1

means that there is a perfect positive linear relationship between the variables; whereas, a correlation of -1 means that there is a perfect negative linear relationship between the variables. A correlation of 0 means there is no linear relationship between the two variables (Kaplan & Saccuzzo, 2001).

Multivariate analysis techniques were used to examine the predictive validity of the instruments. Standard multiple regression analysis was used to predict the managerial performance scores from the full set of predictor variables, namely cognitive ability, personality and situational judgment test variables. Standard multiple regression is a statistical technique used to analyse the relationships between a single dependent (criterion variable) and several predictor variables (Tabachnick & Fidell, 1996). To examine the extent to which the SJT provides incremental validity in the prediction of managerial performance (H011), hierarchical multiple regression analysis was conducted. The focus in hierarchical regression is on “the change in predictability associated with predictor variables entered later in the analysis over and above that contributed by predictor variables entered earlier in the analysis” (Petrocelli, 2003, p. 11). Therefore, the first step of the hierarchical regression analysis entailed including cognitive ability, the second step entailed including the personality variables, with the third step consisting of the addition of the situational judgement measure. The partial correlations for each of the predictors in the regression were subsequently reported. Partial correlation is a measure of relative incremental prediction across all predictors in the regression (O’Connell et al., 2007).

3.7 CONCLUSION: CHAPTER THREE

Chapter three outlined the research design and methodology that were used to test the hypotheses stated in chapter two. Specifically, a description of the sample and measuring instruments was provided, after which the statistical analysis was explained. The results of the data analysis will be discussed in chapter four.

CHAPTER FOUR

RESEARCH RESULTS

4.1 INTRODUCTION

The objective of the current study was to investigate whether a situational judgment test (SJT) significantly adds to the prediction of managerial performance over other measures used for managerial selection, such as measures of cognitive ability and personality. It further aimed to investigate the relationship between the SJT and such predictor measures. A theoretical overview of the predictor and criterion constructs was provided in chapter two, whereas the statistical techniques used in the study were described in chapter three. The purpose of the following chapter is to present and discuss the statistical results that were obtained from the data analysis. The chapter begins with an overview of the univariate descriptive statistics, item analysis and dimensionality analysis of the measures, as well as an investigation of the assumptions underlying the multivariate analysis techniques used in this study. The chapter concludes with a detailed discussion of the results of the inferential statistical analyses, i.e. the correlational analyses and multiple regression analyses.

4.2 DESCRIPTIVE STATISTICS

Descriptive statistics are used to describe and summarise the basic characteristics of the data (sample and measures) under investigation (Tabachnick & Fidell, 1996). Such a description and summary comprise the first step in the statistical analysis, which is presented in Table 4.1. The descriptive statistics are analysed in light of the assumptions underlying multivariate statistical analyses. The assumptions are described in subsection 4.2.1 below.

4.2.1 Assumptions underlying Multivariate Statistical Analyses

In order to make accurate inferences about an actual population based on analyses performed on a sample, several assumptions for the data should be met in order to be able to apply a valid regression model (Field, 2005). The assumptions underlying regression are discussed in this section.

4.2.1.1 Accuracy of Data File and Missing Values

The minimum and maximum values, means and standard deviations of each of the variables were inspected for plausibility and no problems were detected.

4.2.1.2 Ratio of Cases to Independent Variables

According to Tabachnick and Fidell (1996), a sample size of $N \geq 50 + 8m$ is required for testing multiple correlation and $N \geq 104 + m$ is required for testing individual predictors, where m is the number of independent variables. As there are five predictors in the given sample, the adequate sample size of $N = 90$, in the case of the first equation, and $N = 109$, in the case of the second equation, was estimated. In the light of such criteria, the sample size in the current study, $N = 124$, was deemed sufficient to ensure adequate statistical power for the main regression analyses that were used to test the hypotheses.

4.2.1.3 Outliers

Univariate outliers: The presence of univariate outliers was explored by means of box plots, with a number of possible outliers being identified. The following possible outliers were found: one (case: 49) on the Verbal Evaluation measure; one (case: 2) on the Basic Checking measure; three (cases: 50, 102, 54) on the extraversion subscale; one (case: 54) on the emotional stability subscale; three on the agreeableness subscale (cases: 79, 20, 105); and two on the conscientiousness subscale (cases: 78, 90). In order to decide whether to delete possible outliers, we used a cut-off value of standardised score values

(Z-values) greater than 3.29 (Field, 2005). Based on this guideline, the researchers decided to delete the outlier on the Basic Checking measure (case: 2), as it had a z-value in excess of the cut-off of 3.29 and it also significantly skewed the distribution of the variable. When the outlier in question was removed, the effect on the mean and standard deviation of the variable was minimal.

Table 4.1

Analysis of Univariate Descriptives of all Variables

	Statistics		
	Mean	Std. Deviation	N
Perf Rank	.528	.271	109
OPR	6.278	1.904	122
BOS	3.758	.628	116
MPerf Comp	.027	.888	103
VC1.1	31.98	8.229	124
NCC2	13.15	3.915	124
CP7.1	61.85	7.492	123
Extraversion	5.733	.816	122
Openness	4.954	.945	122
Emotional Stability	5.859	.920	122
Agreeableness	5.247	.986	122
Conscientiousness	5.636	.910	122
SJT	60.667	4.725	120

Note. BOS – Behavioural Observation Scale

Perf Rank – Performance Ranking

OPR – Overall Performance Rating

MPerf Comp – Managerial Performance Composite

4.2.1.4 Univariate Normality, Multivariate Linearity and Homoscedasticity

Univariate normality: To assess normality of the distribution of the variables in the model, the researchers conducted one-sample Kolmogorov–Smirnov tests. The one-sample Kolmogorov–Smirnov test assesses the degree to which an observed pattern of categorical frequencies differs from the pattern that would be expected on the null hypothesis (Tabachnick & Fidell, 1996). Findings, as shown in tables 4.1 to 4.4, indicate that all variables are normally distributed, with the exception of the Basic Checking cognitive measure. The skewness z -scores for this measure of -2.361 ($-.513/.217$) indicate a significantly negative skewed distribution at $p < .05$ and the kurtosis z -scores (3.656) ($1.557/.431$) indicate significant kurtosis at $p < .0001$. As mentioned in section 4.2.1.3, the outlier on this measure was excluded from further analysis, thus significantly reducing the skewness ($-.022$) and kurtosis ($-.645$).

Five subscales on the OPQ indicated positive skewness, with four (Outgoing, Modest, Conceptual and Worrying) being significant at $p < .01$ (values greater than the absolute value of 2.58), and one (Emotionally Controlled) being significant at $p < .05$ (with values greater than the absolute value of 1.96). Three subscales were negatively skewed, two at $p < .05$ (Rule Following and Competitive) and one subscale (Achieving) just below the cut-off for significance at $p < .0001$. Kurtosis for all subscales were below the (1.96) cut-off for significance at $p < .05$ (Field, 2005). It was decided not to perform transformations of the variables to reduce skewness, since the scores in question were not used in the subsequent analyses. Instead, the OPQ-facet scores were eventually used to calculate the Big Five personality dimension scores, with these ultimately showing no evidence of skewness or kurtosis.

Table 4.2

Kolmogorov-Smirnov Test of Normality for the Cognitive Measures

One-Sample Kolmogorov-Smirnov Test				
		Verbal evaluation	Numerical interpretation	Basic checking
N		124	124	123
Normal parameters ^a	Mean	31.98	13.15	61.85
	SD	8.229	3.915	7.492
	Absolute	.088	.079	.061
Most extreme differences	Positive	.088	.079	.060
	Negative	-.061	-.065	-.061
Kolmogorov-Smirnov Z		.975	.881	.679
Asymp. sig. (2-tailed)		.298	.420	.745

Note. Test distribution is normal.

Table 4.3

Kolmogorov-Smirnov Test of Normality for the Video-based Simulation

One-Sample Kolmogorov-Smirnov Test		
		Video-based Simulation Total
N		119
Normal parameters ^a	Mean	308.4076
	SD	12.35097
	Absolute	.046
Most extreme differences	Positive	.046
	Negative	-.042
Kolmogorov-Smirnov Z		.499
Asymp. sig. (2-tailed)		.964

Note. Test distribution is normal.

Table 4.4

Kolmogorov-Smirnov Test of Normality for the BOS Measure

		One-Sample Kolmogorov-Smirnov Test																	
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	OPR	Perf Rank
N		122	122	122	122	122	122	122	121	121	122	122	122	122	121	121	120	122	110
Normal parameters ^a	Mean	3.84	3.85	3.73	3.70	3.73	4.07	3.32	3.64	3.49	3.79	3.76	3.87	4.02	3.91	3.33	3.74	3.72	.526
	SD	.866	.985	.988	1.018	.962	.977	1.100	.845	1.017	1.085	.872	.953	1.004	.957	1.012	.939	1.904	.270
Most extreme differences	Absolute	.313	.289	.255	.237	.324	.277	.224	.341	.296	.258	.288	.301	.256	.249	.242	.283	.172	.104
	Positive	.228	.178	.163	.148	.209	.169	.180	.238	.175	.135	.212	.191	.164	.156	.187	.192	.172	.072
	Negative	-.313	-.289	-.255	-.237	-.324	-.277	-.224	-.341	-.296	-.258	-.288	-.301	-.256	-.249	-.242	-.283	-.101	-.104
Kolmogorov-Smirnov Z		3.455	3.193	2.821	2.618	3.576	3.054	2.470	3.747	3.256	2.852	3.178	3.321	2.825	2.734	2.660	3.104	1.902	1.094
Asymp. sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.182

Note. Test distribution is normal.

Note. Abbreviations: S = Statement; OPR = Overall Performance Rating; Perf Rank = Performance Ranking

Table 4.5

Kolmogorov-Smirnov Test of Normality for the OPQ Subscales

		One-Sample Kolmogorov-Smirnov Test																		
		Pers	Cont	Outs	IM	Outg	Affil	SC	Mod	Dem	Car	DR	Eval u	Beha v	Conve n	Conce p	Inno v	VS	Adap	FT
N		122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122
Normal parameters ^a	Mean	6.90	6.69	6.13	4.42	5.37	4.48	5.55	4.64	5.27	5.02	6.47	5.47	5.24	5.63	5.18	5.37	3.90	4.70	5.65
	SD	1.55 6	1.75 4	1.80 9	1.76 7	1.75 0	1.91 7	1.84 1	2.10 5	1.71 5	1.80 9	1.86 8	1.581	1.725	1.759	1.532	1.622	1.77 4	2.11 5	1.92 8
Most extreme differences	Absolut e	.148	.134	.152	.151	.198	.124	.109	.160	.140	.136	.115	.173	.129	.124	.186	.151	.129	.137	.122
	Positive	.139	.120	.119	.151	.198	.124	.109	.160	.122	.105	.115	.139	.129	.105	.186	.127	.129	.125	.091
	Negativ e	-.148	-.134	-.152	-.128	-.089	-.089	-.105	-.087	-.140	-.136	-.106	-.173	-.101	-.124	-.117	-.151	-.101	-.137	-.122
Kolmogorov-Smirnov Z		1.63 6	1.48 2	1.67 6	1.66 7	2.19 0	1.37 2	1.20 5	1.77 1	1.54 7	1.49 9	1.27 2	1.910	1.420	1.370	2.057	1.672	1.42 2	1.51 8	1.34 4
Asymp. sig. (2-tailed)		.009	.025	.007	.008	.000	.046	.110	.004	.017	.022	.079	.001	.035	.047	.000	.007	.035	.020	.054

Note. Test distribution is normal.

Note. Abbreviations: Pers = Persuasiveness; Cont = Controlling; Outs = Outspoken; IM = Independent-minded; Outg = Outgoing; Affil = Affiliative; SC = Socially confident; Mod = Modest; Car = Caring; DR = Data-rational; Eval = Evaluative; Behav = Behavioural; Conven = Conventional; Concep = Conceptual; Innov = Innovative; VS = Variety-seeking; Adap = Adaptable; FT = Forward-thinking

Table 4.5 Continued

		One-Sample Kolmogorov-Smirnov Test																		
		DC	Consc	RF	Rel	Worry	TM	Opt	Trust	EC	Vig	Compet	Ach	Dec	Consist	E	O	ES	A	C
N		122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122	122
Normal parameters ^a	Mean	5.88	5.68	7.22	5.80	4.56	5.84	5.57	4.72	4.87	4.78	6.19	6.38	5.10	4.66	173.87	61.72	98.06	43.20	152.20
	SD	1.497	1.663	1.956	1.621	1.725	1.647	1.715	1.855	1.532	1.834	1.668	1.712	1.820	2.015	31.019	26.455	25.771	24.654	24.592
Most extreme differences	Absolute	.189	.126	.147	.131	.192	.177	.123	.158	.157	.148	.187	.183	.153	.132	.086	.054	.049	.064	.064
	Positive	.189	.126	.094	.131	.192	.139	.123	.121	.157	.148	.117	.120	.153	.091	.086	.052	.049	.064	.055
	Negative	-.188	-.109	-.147	-.123	-.095	-.177	-.123	-.158	-.121	-.122	-.187	-.183	-.134	-.132	-.055	-.054	-.036	-.062	-.064
Kolmogorov-Smirnov Z		2.083	1.392	1.618	1.444	2.124	1.955	1.360	1.746	1.738	1.635	2.062	2.025	1.687	1.456	.945	.602	.541	.707	.705
Asymp. sig. (2-tailed)		.000	.042	.011	.031	.000	.001	.049	.005	.005	.010	.000	.001	.007	.029	.333	.862	.932	.699	.702

Note. Test distribution is normal.

Note. Abbreviations: DC = Detail-conscious; Consc = Conscientious; RF = Rule-following; Rel = Relaxed; Worry = Worrying; TM = Tough-minded; Opt = Optimistic; Trust = Trusting; EC = Emotionally controlled; Vig = Vigorous; Compet = Competitive; Ach = Achieving; Dec = Decisive; Consist = Consistency; E = Extraversion; O = Openness; ES = Emotional stability; A = Agreeableness; C = Conscientiousness.

Multivariate linearity: The assumption of multivariate linearity is that there are straight-line relationships between pairs of variables. Multivariate analyses based on correlation capture only the linear relationships among variables, so nonlinear relationships among variables are ignored unless specifically added into the analysis by the researcher (Tabachnick & Fidell, 1996). Multivariate linearity was assessed by visual inspection of bivariate scatterplots and no compelling evidence for the violation of this assumption was found.

Homoscedasticity: The assumption of homoscedasticity for ungrouped data is that the variability in scores for one continuous variable is roughly the same at all values of another continuous variable. Heteroscedasticity, the failure of homoscedasticity, occurs because one of the variables is not normally distributed because there is greater error of measurement of one variable at some levels, or because one of the variables is spread apart at some levels by its relationship to a third variable (Tabachnick & Fidell, 1996). Homoscedasticity was assessed by visual inspection of bivariate scatterplots and no compelling evidence for the violation of this assumption was found.

In general, the results of our analyses of the characteristics of the data show that the data set broadly adhered to the assumptions underlying the multivariate analysis techniques we employed, with some exceptions. In these cases, appropriate remedies were consistently applied, thus rendering the data suitable for further analyses.

4.3 ITEM ANALYSIS

Item analysis is typically used to determine the effectiveness of each test item (Kaplan & Saccuzzo, 2001). Due to the fact that data were available at scale score total level only, item level analyses were precluded. However, we performed item analysis on the criterion measure which was constructed for this study. The items comprising the **Behaviour Observation Scale (BOS)** Questionnaire, which was designed to be a broad measure of managerial job performance, were analysed using the SPSS RELIABILITY procedure. The purpose of this analysis was to eliminate those items adversely affecting the internal consistency of the measure. The measure of reliability of measurement used in the current study was Cronbach's coefficient alpha, which

provides an indication of the average correlation among all of the items comprising the scale (Field, 2005; Nunnally, 1978; Tabachnick & Fidell, 1996).

The item analysis enabled the identification of Item 7 as an item to be considered for deletion for several reasons: (1) the item, once deleted, increases the coefficient alpha; (2) the item has consistently low inter-item correlations especially between Items 6 and 7 (.279), which were both intended to measure *relating and networking*; (3) the factor loadings of Item 7 are all below the value of .4, thus supporting the decision to delete the item. The analyses were rerun post deletion of Item 7, which increased the Cronbach coefficient alpha value to .899. The results are shown in Table 4.7.

According to Nunnally (1978), optimal mean inter-item correlation values range from .2 to .4. The correlations between the items within each subscale (with two items per scale) are generally moderate to high, ranging between .380 and .636 (as shown in Table 4.8). Such correlations suggest that the items generally reflect their designated latent variable with reasonable success, except in the case of Item 7, as discussed earlier. Furthermore, all correlations are statistically significant at the p (one-tailed) < .01 level.

Table 4.6

Reliability Analysis of the Items Comprising the BOS Questionnaire

Item-total statistics					
	Scale mean if item deleted	Scale variance if item deleted	Corrected item – total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
Statement 1	55.84	85.269	.617	.569	.887
Statement 2	55.83	83.779	.618	.598	.886
Statement 3	55.95	84.449	.574	.448	.888
Statement 4	55.97	81.512	.716	.609	.882
Statement 5	55.94	82.857	.689	.568	.883
Statement 6	55.61	85.179	.538	.515	.889
Statement 7	56.37	88.827	.288	.225	.899
Statement 8	56.03	86.399	.564	.487	.888
Statement 9	56.20	86.021	.471	.618	.892
Statement 10	55.88	80.959	.705	.595	.882
Statement 11	55.92	85.133	.627	.704	.886
Statement 12	55.81	87.059	.459	.320	.892
Statement 13	55.64	85.276	.523	.506	.890
Statement 14	55.73	86.580	.479	.416	.891
Statement 15	56.34	86.990	.421	.496	.894
Statement 16	55.91	83.669	.669	.740	.884

Reliability coefficients:

N of cases = 116.

N of items = 16.

Alpha = .895.

Table 4.7

Reliability Analysis of the Items Comprising the BOS Questionnaire Post-Deletion of Item 7

Item-total statistics					
	Scale mean if item deleted	Scale variance if item deleted	Corrected item – total correlation	Squared multiple correlation	Cronbach's alpha if item deleted
Statement1	52.54	78.337	.626	.567	.891
Statement2	52.53	77.068	.616	.598	.891
Statement3	52.66	77.758	.569	.432	.893
Statement4	52.67	74.796	.720	.609	.887
Statement5	52.65	76.091	.693	.568	.888
Statement6	52.32	78.584	.525	.509	.895
Statement8	52.74	79.706	.553	.483	.894
Statement9	52.91	78.991	.482	.614	.897
Statement10	52.59	74.366	.703	.594	.888
Statement11	52.63	78.409	.622	.698	.891
Statement12	52.52	80.287	.452	.312	.897
Statement13	52.34	78.350	.530	.504	.895
Statement14	52.44	79.222	.510	.367	.895
Statement15	53.04	80.163	.418	.484	.899
Statement16	52.61	76.883	.672	.738	.889

Reliability coefficients:

N of cases = 116.

N of Items = 15.

Alpha = .899.

Table 4.8

Mean Inter-Item Correlations of the BOS Subscales

		Correlations															
		Statement 1	Statement 2	Statement 3	Statement 4	Statement 5	Statement 6	Statement 8	Statement 9	Statement 10	Statement 11	Statement 12	Statement 13	Statement 14	Statement 15	Statement 16	
Statement 1	Pearson correlation	1.000	.543**	.440**	.432**	.522**	.238**	.333**	.469**	.464**	.407**	.274**	.222**	.435**	.356**	.490**	
	Sig. (1-tailed)		.000	.000	.000	.000	.004	.000	.000	.000	.000	.001	.007	.000	.000	.000	
Statement 2	Pearson correlation	.543**	1.000	.417**	.533**	.472**	.483**	.570**	.219**	.543**	.478**	.279**	.161*	.259**	.158*	.436**	
	Sig. (1-tailed)	.000		.000	.000	.000	.000	.000	.008	.000	.000	.001	.038	.002	.042	.000	
Statement 3	Pearson correlation	.440**	.417**	1.000	.528**	.435**	.207*	.335**	.504**	.408**	.308**	.164*	.304**	.298**	.365**	.355**	
	Sig. (1-tailed)	.000	.000		.000	.000	.011	.000	.000	.000	.000	.036	.000	.000	.000	.000	
Statement 4	Pearson correlation	.432**	.533**	.528**	1.000	.626**	.510**	.527**	.365**	.564**	.386**	.360**	.425**	.355**	.298**	.460**	
	Sig. (1-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	
Statement 5	Pearson correlation	.522**	.472**	.435**	.626**	1.000	.459**	.431**	.328**	.506**	.474**	.357**	.415**	.351**	.210*	.569**	
	Sig. (1-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.010	.000	
Statement 6	Pearson correlation	.238**	.483**	.207*	.510**	.459**	1.000	.421**	.068	.489**	.377**	.453**	.412**	.237**	.104	.345**	
	Sig. (1-tailed)	.004	.000	.011	.000	.000		.000	.231	.000	.000	.000	.000	.004	.129	.000	
Statement 8	Pearson correlation	.333**	.570**	.335**	.527**	.431**	.421**	1.000	.211*	.416**	.339**	.312**	.163*	.207*	.341**	.377**	
	Sig. (1-tailed)	.000	.000	.000	.000	.000	.000		.010	.000	.000	.000	.037	.012	.000	.000	
Statement 9	Pearson correlation	.469**	.219**	.504**	.365**	.328**	.068	.211*	1.000	.253**	.216**	.112	.438**	.299**	.610**	.185*	
	Sig. (1-tailed)	.000	.008	.000	.000	.000	.231	.010		.003	.009	.112	.000	.000	.000	.022	
Statement 10	Pearson correlation	.464**	.543**	.408**	.564**	.506**	.489**	.416**	.253**	1.000	.636**	.316**	.436**	.317**	.262**	.625**	
	Sig. (1-tailed)	.000	.000	.000	.000	.000	.000	.000	.003		.000	.000	.000	.000	.002	.000	

Statement 11	Pearson correlation	.407**	.478**	.308**	.386**	.474**	.377**	.339**	.216**	.636**	1.000	.360**	.354**	.282**	.203*	.786**
	Sig. (1-tailed)	.000	.000	.000	.000	.000	.000	.000	.009	.000		.000	.000	.001	.013	.000
Statement 12	Pearson correlation	.274**	.279**	.164*	.360**	.357**	.453**	.312**	.112	.316**	.360**	1.000	.305**	.324**	.132	.380**
	Sig. (1-tailed)	.001	.001	.036	.000	.000	.000	.000	.112	.000	.000		.000	.000	.075	.000
Statement 13	Pearson correlation	.222**	.161*	.304**	.425**	.415**	.412**	.163*	.438**	.436**	.354**	.305**	1.000	.442**	.264**	.331**
	Sig. (1-tailed)	.007	.038	.000	.000	.000	.000	.037	.000	.000	.000	.000		.000	.002	.000
Statement 14	Pearson correlation	.435**	.259**	.298**	.355**	.351**	.237**	.207*	.299**	.317**	.282**	.324**	.442**	1.000	.324**	.355**
	Sig. (1-tailed)	.000	.002	.000	.000	.000	.004	.012	.000	.000	.001	.000	.000		.000	.000
Statement 15	Pearson correlation	.356**	.158*	.365**	.298**	.210*	.104	.341**	.610**	.262**	.203*	.132	.264**	.324**	1.000	.264**
	Sig. (1-tailed)	.000	.042	.000	.000	.010	.129	.000	.000	.002	.013	.075	.002	.000		.002
Statement 16	Pearson correlation	.490**	.436**	.355**	.460**	.569**	.345**	.377**	.185*	.625**	.786**	.380**	.331**	.355**	.264**	1.000
	Sig. (1-tailed)	.000	.000	.000	.000	.000	.000	.000	.022	.000	.000	.000	.000	.000	.002	

Note. $N = 119 - 122$, due to missing values.

** . Correlation is significant at the 0.01 level (1-tailed).

* . Correlation is significant at the 0.05 level (1-tailed).

From the results of the item analysis provided in Table 4.6, it is evident that the alpha value (.895) approaches the highly satisfactory criterion of .9 for reliability measures (Nunnally, 1978). Overall, the item analysis of the BOS-measure of managerial performance points towards an internally consistent measure of job performance. Since the additional measures of job performance (i.e., Performance Ranking and Overall Performance Rating measures) consisted of single-item measures using a single rating source, reliability estimates for these measures could not be computed.

4.4 DIMENSIONALITY ANALYSIS

We conducted dimensionality analysis for the criterion measure, which was the only measure for which item-level data was available. Due to the fact that data were available at scale score total level only, item level analyses were precluded, although the dimensionality of these measures have widely been supported in similar settings, as discussed in the *Measuring Instruments* section (see section 3.5).

Dimensionality analysis is normally conducted by employing a choice of various exploratory factor analysis (EFA) techniques. The two most widely used forms of factor analysis are principal components analysis and principal axis factoring (Tabachnick & Fidell, 1996). Both the methods are concerned with describing the variance that is shared by the scores of people on three or more variables. The key difference between the two is how they handle unique variance. In principal components analysis, the total variance of a variable is analysed, including its unique variance. In other words, it is assumed that the test used to assess the variable is perfectly reliable and without error. In principal axis factoring, only the variance which is common to, or shared by, the tests is analysed. Therefore, an attempt is made to exclude unique variance from the analysis (Bryman & Cramer, 2001). For that reason, the principal axis factoring technique was used. Principal axis factoring with Varimax rotation was performed on the items comprising the BOS Questionnaire. The Kaiser criterion (i.e., eigenvalue-greater-than-unity) and scree plots were used to determine the number of factors to extract (Tabachnick & Fidell, 1996).

The EFA results (shown in tables 4.8 and 4.9) reflect the multidimensional, interrelated nature of managerial job performance, since a complex factor structure

emerges. From tables 4.8 and 4.9, it is evident that the factor loadings are moderate to high. Moreover, the BOS items tend to cross-load on more than one factor, meaning that there is a lack of simple structure, since the items reflect more than one latent performance dimension. Thus, it seems that the eight factors which the instrument was designed to measure, when factor- analysed, do not emerge clearly from the analysis. The results point to four broader dimensions, which are possibly second-order factors, underlying the eight facets of managerial job performance.

Table 4.9

Factor Loadings of All Items Comprising the BOS Questionnaire

	Rotated factor matrix^a			
	Factor			
	1	2	3	4
Statement 1	.354	.513	.366	-.014
Statement 2	.701	.202	.294	-.023
Statement 3	.349	.539	.161	.098
Statement 4	.626	.355	.179	.296
Statement 5	.508	.313	.358	.230
Statement 6	.640	-.077	.143	.470
Statement 7	.327	.050	.113	.048
Statement 8	.657	.221	.151	.003
Statement 9	.030	.878	.048	.158
Statement 10	.475	.219	.499	.278
Statement 11	.287	.125	.778	.179
Statement 12	.374	.054	.250	.281
Statement 13	.095	.339	.187	.771
Statement 14	.160	.392	.228	.305
Statement 15	.107	.617	.076	.088
Statement 16	.299	.188	.846	.140

Extraction method: Principal axis factoring.

Rotation method: Varimax with Kaiser normalisation.

a. Rotation converged in 7 iterations.

Table 4.10

Factor Loadings of Items Comprising the BOS Questionnaire, Excluding Item 7

Rotated factor matrix^a				
	Factor			
	1	2	3	4
Statement 1	.383	.500	.362	-.020
Statement 2	.722	.176	.301	-.016
Statement 3	.350	.529	.168	.107
Statement 4	.638	.334	.186	.311
Statement 5	.515	.297	.362	.238
Statement 6	.604	-.084	.168	.479
Statement 8	.639	.202	.172	.030
Statement 9	.046	.886	.049	.151
Statement 10	.459	.209	.510	.289
Statement 11	.256	.124	.793	.184
Statement 12	.357	.047	.260	.289
Statement 13	.080	.343	.191	.774
Statement 14	.201	.378	.219	.293
Statement 15	.100	.621	.085	.091
Statement 16	.283	.184	.853	.142

Extraction method: Principal axis factoring.

Rotation method: Varimax with Kaiser normalisation.

a. Rotation converged in 7 iterations.

Since the aim of the EFA was to assess the unidimensionality of the managerial job performance construct, as measured by the BOS, we ventured only a tentative interpretation of the observed factor structure. The factors, as per the EFA, could be interpreted as follows (see appendix B for a full description of the items):

Factor One: Influencing (people and resources): The items comprising the first factor (statements 2 and 8 designed to measure *persuading and influencing*, and statements 4 and 5 designed to measure *leading and supervising*) all seem to have an influencing component, whether consisting of people or resources, which might explain why they load onto the same factor. Item 10 can be interpreted as a complex

item, as it loads onto various factors strongly, though, theoretically, the item does not fit with the remainder of the items in the factor, seemingly better suited to the third factor.

Factor Two: Analysing and decision-making: Items 1 and 3 are both designed to measure *deciding and initiating action*, with Items 9 and 15 both being designed to measure *analysing*. The behavioural indicators seem to share a common underlying aspect of analysis, which might explain why they load onto one factor.

Factor Three: Achievement orientation: The items loading onto the third factor all seem to relate to reaching objectives/goals, whether these are personal or work-related. Items 10 and 11 are intended to measure *achieving personal work goals and objectives*. Item 16, which measures *delivers results and reaches targets* also relates to reaching objectives. Despite Item 10 being a complex item, theoretical interpretation of the items/factors suggest that such an item is better suited to this factor.

Factor Four: Maintaining order and relationships: Items 13 and 14 are intended to measure *following instructions and procedures*, with Item 6 relating to *developing and maintaining positive relationships with others*. Therefore, it seems that the common theme emerging among the items, concerns the maintenance or preservation of order through maintaining good working relationships, following instructions without challenging authority and carrying out policies and procedures, and so on.

The eigenvalues for the four factors were as follows: Factor 1 (6.363), Factor 2 (1.671), Factor 3 (1.087) and Factor 4 (1.014), thereby contributing 42.4%, 11.1%, 7.2% and 6.8% of the variance in the scales, respectively. The cumulative variance contributed by the four factors is 67.57%. Furthermore, the determinant (.000) is greater than .00001, indicating that multicollinearity was not problematic (Field, 2005).

In conclusion, although the reliability of the scale scores (see section 4.3 for detailed analysis) shows strong evidence that the BOS measure has high internal consistency ($\alpha = .899$), the dimensionality analysis shows that the factor structure is complex,

reflecting the theoretical complexity of the job performance construct. There is, however, considerable conceptual agreement regarding the comparison of the observed factor structure with the suggested theoretical models of managerial performance (e.g., Borman & Brush, 1993; Tett et al., 2001). It is suggested that future studies build on the results of the current study and develop a factorially pure job performance measure, with additional items measuring each facet of job performance, as suggested by the competency model developed by means of job analysis.

Based on the results of the item analysis and exploratory factor analysis, such as high internal consistency and consistently high item-total correlations, it was deemed acceptable to create an unweighted linear composite score for managerial performance from the constituent items (after the deletion of one problematic item, i.e. Item 7). Even though the items did not load clearly onto a single dimension of job performance, it is commonly deemed appropriate to linearly combine items that do not load clearly onto the same factor, since jointly, the linear combination would provide a representative sample of behaviours from the multifaceted nature of the broad construct domain of job performance. In other words, the combination of items that measure distinct parts of the broad construct domain of job performance yield a valid measure of the underlying single job performance latent construct (see Viswesvaran et al., 2005) of managerial job performance.

4.5 RESULTS

After it was confirmed that the data set met the assumptions for the chosen statistical analyses, the *a priori* hypotheses (refer to section 3.3) that were formulated were tested. Firstly the inter-correlation results are discussed, followed by the presentation of the multiple regression results.

4.5.1 Inter-correlations

Correlational analysis is designed primarily to examine linear relationships between variables. Pearson's product-moment correlation coefficient, r , was used to establish the nature of the various relationships between the variables. According to

Tabachnick and Fidell (1996), this statistic is the most frequently used measure of association and the basis of many multivariate calculations. Pearson's correlation ranges from +1 to -1. A correlation of +1 means that there is a perfect positive linear relationship between the variables; whereas, a correlation of -1 means that there is a perfect negative linear relationship between the variables. A correlation of 0 means there is no linear relationship between the two variables (Kaplan & Saccuzzo, 2001).

The criterion construct (i.e., managerial performance) was measured in four ways, using: (1) the Behavioural Observation Scale (BOS), (2) Performance Ranking (Perf Rank), (3) an Overall Performance Rating (OPR), and (4) a Managerial Performance Composite (MPerf Comp) obtained by calculating the mean of the z-scores of the first three measures. The Managerial Performance Composite (MPerf Comp) was chosen as the criterion against which to decide whether to reject or confirm the hypotheses, since this measure provides the best overall operationalisation of the managerial performance construct. As explained in the literature review, each individual type of job performance measure (e.g., supervisory BOS-ratings, relative Performance Ranking of subordinates) generally suffers from methodological weaknesses (see Guion, 1998) and by integrating the three into a composite score, we intended to overcome such weaknesses.

From the analyses it appears that the correlations between the four measures of managerial performance (i.e., criterion measures) are moderate to high and range from $r = .613$ to $r = .928$, p (one-tailed) $< .01$. The strength of the correlations between the measures is thus strong evidence of construct validity for the criterion measure. The Managerial Performance Composite was found to correlate highly with the Overall Performance Rating ($r = .928$, p (one-tailed) $< .01$); BOS ($r = .900$, p (one-tailed) $< .01$) and Performance Ranking ($r = .857$, p (one-tailed) $< .01$). The Overall Performance Rating and Performance Ranking also correlated highly with each other ($r = .684$, p (one-tailed) $< .01$). In this regard it appears that the supervisor raters had similar target behaviour in mind when assigning ratings to managers. The correlation results also suggest that although the BOS correlated highly with the Overall Performance Rating ($r = .823$, p (one-tailed) $< .01$) and the Performance Ranking ($r = .613$, p (one-tailed) $< .01$), the measures are not perfectly aligned and therefore not redundant. For this reason, it was decided to report the correlation and regression

results separately for the four alternative measures of job performance for illumination, although the Managerial Performance Composite score would be used for decision-making in terms of rejection or support of the list of hypotheses.

Having discussed the intercorrelations between the criterion measures, the testing of hypothesised relationships between predictor and criterion construct measures will be consequently be presented.

4.5.1.1 The Relationship between Verbal Evaluation and Managerial Performance (H₁)

The relationship between a manager's verbal evaluation ability (as measured by the Verbal Evaluation ability test) and managerial performance (as measured by the four separate indicators of performance) was investigated using Pearson's product-moment correlation coefficient. From the correlation matrix (Table 4.11) it can be seen that the first hypothesis, stating that a significant positive relationship (i.e., a unidirectional hypothesis) exists between a manager's verbal evaluation ability and managerial performance, is not supported. In this regard, an insignificant relationship exists between verbal evaluation and the Managerial Performance Composite, $r = .153$, p (one-tailed) $> .05$. The same applies to the relationship of this predictor with the BOS measure $r = .056$, p (one-tailed) $> .05$. In contrast, however, a weak, positive and significant relationship exists between verbal evaluation and both the Performance Ranking of managers (Perf Rank), $r = .247$, p (one-tailed) $< .01$ and the Overall Performance Rating criterion, $r = .171$, p (one-tailed) $< .05$. In conclusion, based on the insignificant relationship found with the composite of managerial performance, the null hypothesis (H₀1) cannot be accepted.

4.5.1.2 The Relationship between Numerical Interpretation and Managerial Performance (H₂)

A significant, positive relationship was found between numerical interpretation ability (as assessed with the Numerical Interpretation ability test) and managerial performance (on all the criterion measures except for the Behavioural Observation Scale, i.e. the Performance Ranking, Overall Performance Rating and the Managerial

Performance Composite) (see Table 4.11). The correlations for the three measures were as follows: between numerical interpretation and the Performance Ranking, $r = .288$, p (one-tailed) $< .01$; between numerical interpretation and the Overall Performance Rating, $r = .188$, p (one-tailed) $< .05$ and between numerical interpretation and the Managerial Performance Composite, $r = .179$, p (one-tailed) $< .05$. However, an insignificant relationship was found between numerical interpretation and the BOS measure, $r = .113$, p (one-tailed) $> .05$. Despite the weak correlation with one of the measures, the second hypothesis, stating that a significant positive relationship exists between a manager's numerical interpretation ability and managerial performance, is confirmed based on the significant correlation with the Managerial Performance Composite.

4.5.1.3 The Relationship between Basic Checking Ability and Managerial Performance (H₃)

The correlation results, as shown in Table 4.11, reveal that in most instances, the relationship between a manager's basic checking ability (as measured by the Basic Checking ability test) and managerial performance was found to be insignificant. The relationship between basic checking and the BOS measure was furthermore found to be negative, $r = -.012$, p (one-tailed) $> .05$. The relationship between basic checking and the Overall Performance Rating, $r = .087$, p (one-tailed) $> .05$ was insignificant. In contrast, a weak but significant positive relationship was found between basic checking and the Performance Ranking measure, $r = .262$, p (one-tailed) $< .01$. The relationship between basic checking ability and managerial performance can, however, not be confirmed due to the insignificant relationship found with the Managerial Performance Composite, $r = .143$, p (one-tailed) $> .05$.

4.5.1.4 The Relationship between Conscientiousness and Managerial Performance (H₄)

There was insufficient evidence to support hypothesis four as an insignificant relationship was found between conscientiousness (as measured by the composite formed from the Occupational Personality Questionnaire (OPQ) subscales) and all four measures of managerial performance (as shown in Table 4.11). H₀₄ can therefore

not be accepted. For the BOS, Performance Ranking, Overall Performance Rating and Managerial Performance Composite the correlation coefficients were, $r = .070$; $.026$; $-.003$ and $-.002$ respectively, ps (one-tailed) $> .05$.

4.5.1.5 The Relationship between Extraversion and Managerial Performance (H₅)

Hypothesis five, stating that there is a significant, positive relationship between a manager's extraversion (as measured by the composite formed from the OPQ subscales) and managerial performance, could not be corroborated. A negative and insignificant relationship was found between extraversion and the BOS, $r = -.036$, p (one-tailed) $> .05$, between extraversion and the Overall Performance Rating, $r = -.029$, p (one-tailed) $> .05$, between extraversion and the Performance Ranking measure, $r = -.007$, p (one-tailed) $> .05$ and lastly between extraversion and the Managerial Performance Composite, $r = -.015$, p (one-tailed) $> .05$.

4.5.1.6 The Relationship between Conscientiousness and Performance on a Situational Judgement Test (H₆)

The analysis of the relationship between conscientiousness (as measured by the composite formed from the OPQ subscales) and performance on a Situational Judgement Test (SJT) (as measured by the VBS) revealed a positive, significant relationship, $r = .243$, p (one-tailed) $< .01$. H₀₆ is therefore rejected in favour of H_{a6} thus confirming hypothesis six. In other words, increased levels of conscientiousness are associated with increased levels of performance on a Situational Judgement Test.

4.5.1.7 The Relationship between Agreeableness and Performance on a Situational Judgement Test (H₇)

There is insufficient evidence to support hypothesis seven. A positive but insignificant relationship exists between agreeableness and a manager's performance on the SJT, $r = .10$, p (one-tailed) $> .05$ (Table 4.11). H₀₇ can therefore not be rejected.

4.5.1.8 The Relationship between Emotional Stability and Performance on a Situational Judgement Test (H_8)

Emotional stability was assessed using the OPQ composite scale and findings indicated that the relationship between this construct and a managers performance on a Situational Judgement Test was found to be insignificant, $r = .083$, p (one-tailed) $> .05$ as shown in Table 4.11. H_{8} is thus rejected as there is insufficient evidence to support this alternative hypothesis.

4.5.1.9 The relationship between Managerial Performance and Performance on a Situational Judgement Test (H_9)

The correlation analyses (as shown in Table 4.11) indicate that a positive, moderate relationship was found between the BOS and the SJT, $r = .234$, p (one-tailed) $< .01$. However, insignificant relationships were found between the SJT and the other measures of managerial performance, using the Performance Ranking ($r = .081$, p (one-tailed) $> .05$), Overall Performance Rating ($r = .126$, p (one-tailed) $> .05$) and the Managerial Performance Composite ($r = .136$, p (one-tailed) $> .05$). However, hypothesis nine must be rejected since an insignificant correlation was found with the composite measure (MPerf Comp) of managerial performance, despite the significantly positive correlation with the BOS measure. This contradictory finding is noteworthy and will receive greater attention in the Discussion section.

From inspection of the correlation matrix in Table 4.11, it is evident that various other statistically significant correlations between scores of measures were obtained. However, these were not discussed since no other a priori relationships were hypothesised beyond those listed above. It can be argued that drawing conclusions from these significant findings could constitute ‘data dredging’, i.e. interpreting significant findings as if they had been expected. However, reference is made to these observed correlations later in order to illuminate plausible rival hypotheses that arose from the discussion of the results.

In summary, the results of our correlational analysis confirmed the hypothesised relationships, with the exception of a few cases. For the sake of clarity, Figure 4.1

summarises the obtained findings, with observed (uncorrected) correlation coefficients indicated on the paths between variables. It is evident from inspection of this figure that the criterion variable (managerial performance) was decomposed into its constituent measures, due to the lack of congruence between predictor-criterion relationships between the various criterion measures.

Table 4.11

Correlations between Predictors and Criteria

		Correlations												
Variable		Verbal evaluation	Numerical interpretation	Basic checking	Extraversio n	Openness to Experience	Emotional Stability	Agreeablen ess	Conscienti ousness	SJT	OPR	Perf Rank	BOS	MPerf Comp
Verbal evaluation	Pearson Correlation	1.000	.441**	.278**	.122	.131	.089	.093	-.054	.058	.171*	.247**	.056	.153
	Sig. (1-tailed)		.000	.002	.101	.086	.176	.165	.287	.275	.036	.007	.283	.070
Numerical interpretation	Pearson Correlation	.441**	1.000	.476**	.116	.041	.176*	-.056	.005	.135	.188*	.288**	.113	.179*
	Sig. (1-tailed)	.000		.000	.113	.334	.032	.278	.477	.083	.024	.002	.122	.042
Basic checking	Pearson Correlation	.278**	.476**	1.000	.096	-.030	.026	.104	-.023	-.030	.087	.262**	-.012	.143
	Sig. (1-tailed)	.002	.000		.158	.377	.394	.140	.404	.378	.182	.005	.451	.084
Extraversion	Pearson Correlation	.122	.116	.096	1.000	.171*	.446**	-.048	-.174*	.104	-.029	.007	-.036	-.015
	Sig. (1-tailed)	.101	.113	.158		.036	.000	.307	.034	.144	.382	.471	.356	.444
Openness to Experience	Pearson Correlation	.131	.041	-.030	.171*	1.000	.134	-.086	-.334**	-.087	.025	.084	-.012	.072
	Sig. (1-tailed)	.086	.334	.377	.036		.080	.184	.000	.187	.398	.206	.450	.245
Emotional Stability	Pearson Correlation	.089	.176*	.026	.446**	.134	1.000	.013	.021	.083	-.045	-.104	-.070	-.105
	Sig. (1-tailed)	.176	.032	.394	.000	.080		.447	.415	.198	.319	.155	.237	.155
Agreeableness	Pearson Correlation	.093	-.056	.104	-.048	-.086	.013	1.000	-.039	.100	-.167*	-.194*	-.121	-.207*
	Sig. (1-tailed)	.165	.278	.140	.307	.184	.447		.342	.153	.040	.028	.106	.022
Conscientious ness	Pearson Correlation	-.054	.005	-.023	-.174*	-.334**	.021	-.039	1.000	.243**	-.003	.026	.070	-.002
	Sig. (1-tailed)	.287	.477	.404	.034	.000	.415	.342		.006	.487	.402	.234	.494
SJT	Pearson Correlation	.058	.135	-.030	.104	-.087	.083	.100	.243**	1.000	.126	.081	.234**	.136

	Sig. (1-tailed)	.275	.083	.378	.144	.187	.198	.153	.006		.099	.218	.008	.098
OPR	Pearson Correlation	.171*	.188*	.087	-.029	.025	-.045	-.167*	-.003	.126	1.000	.684**	.823**	.928**
	Sig. (1-tailed)	.036	.024	.182	.382	.398	.319	.040	.487	.099		.000	.000	.000
Perf Rank	Pearson Correlation	.247**	.288**	.262**	.007	.084	-.104	-.194*	.026	.081	.684**	1.000	.613**	.857**
	Sig. (1-tailed)	.007	.002	.005	.471	.206	.155	.028	.402	.218	.000		.000	.000
BOS	Pearson Correlation	.056	.113	-.012	-.036	-.012	-.070	-.121	.070	.234**	.823**	.613**	1.000	.900**
	Sig. (1-tailed)	.283	.122	.451	.356	.450	.237	.106	.234	.008	.000	.000		.000
MPerf Comp	Pearson Correlation	.153	.179*	.143	-.015	.072	-.105	-.207*	-.002	.136	.928**	.857**	.900**	1.000
	Sig. (1-tailed)	.070	.042	.084	.444	.245	.155	.022	.494	.098	.000	.000	.000	

Note. N = 95-124 due to pairwise deletion of cases

BOS – Behavioural Observation Scale

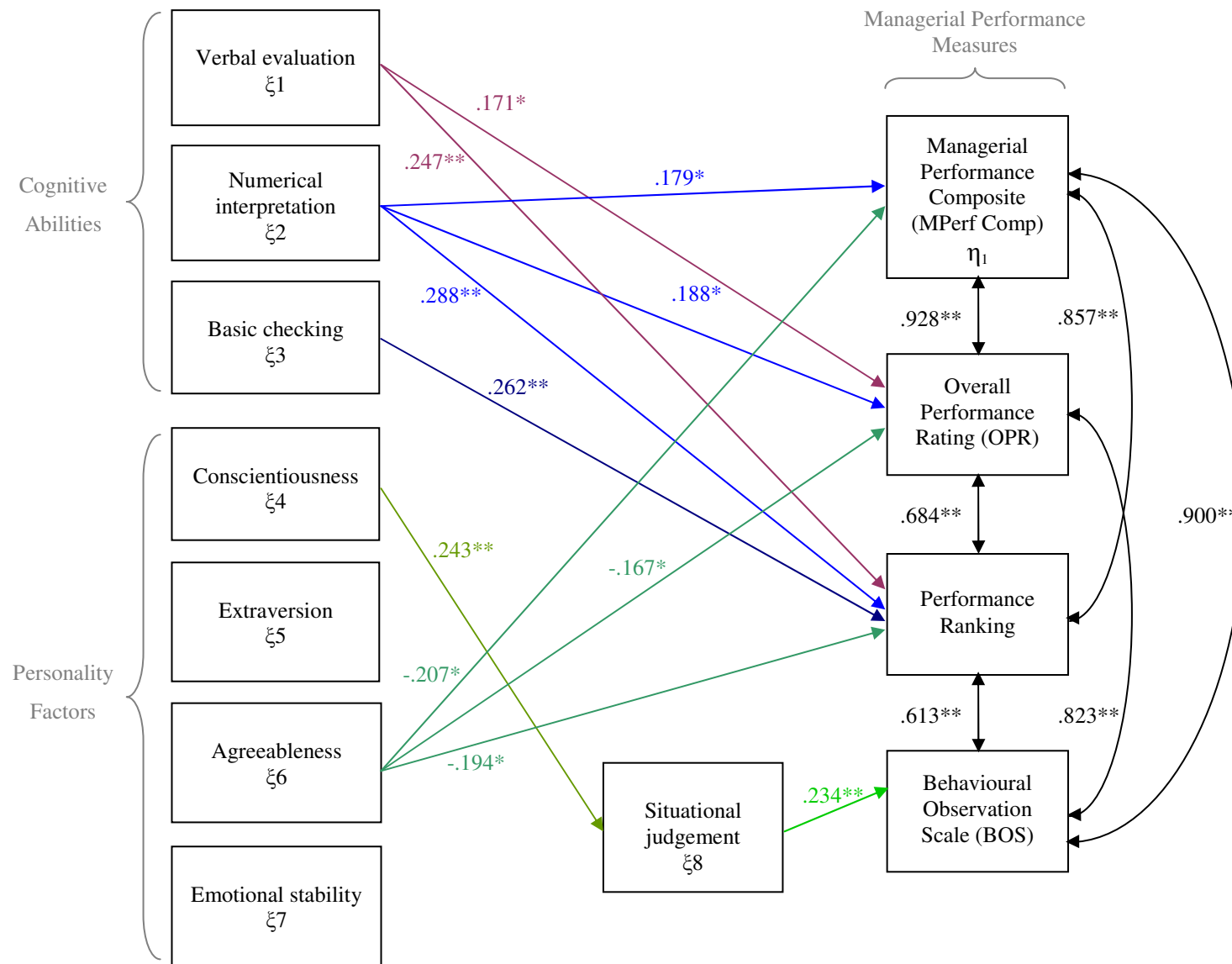
OPR – Overall Performance Rating

Perf Rank – Performance Ranking

MPerf Comp – Managerial Performance Composite

** . Correlation is significant at the 0.01 level (1-tailed)

* . Correlation is significant at the 0.05 level (1-tailed)



Note: **. Correlation is significant at the 0.01 level (1-tailed); *. Correlation is significant at the 0.05 level (1-tailed)

Figure 4.1 Research Model indicating Significant Correlations between Predictors and Criteria

4.5.2 Corrections for Unreliability and Restriction of Range

Due to measures always having a degree of error associated with it, observed correlations between predictors and criteria may be underestimated (Gatewood & Field, 1994). In order to determine what the correlation between two variables would be if the measure of each variable were perfectly reliable, the correlations should be corrected for attenuation. Attenuation can occur either due to unreliability or restriction of range (Nunnally, 1978).

Criterion Unreliability: As noted in section 2.2.2, reliability is a necessary condition for validity. Because the focus in validation is the test, it is appropriate to perform a statistical correction for unreliability in the criterion in order to provide a more accurate assessment of the validity of the test. The following formula by Schmitt and Chan (1998) was used to correct the correlation coefficients for attenuation due to criterion unreliability:

$$r'_{xy} = \frac{r_{xy}}{\sqrt{r_{yy}}}$$

r'_{xy} = correlation between x and y, corrected for attenuation

r_{xy} = original observed correlation between x and y

r_{yy} = reliability of the criterion

Restriction of Range: Range restriction typically occurs when the correlation between two variables (x and y) in a population is estimated, but subjects are selected on x, and data for y are only available for a selected sample (Raju & Brand, 2003). In such a situation, it is appropriate to correct the correlations for the restriction of range. The corrected correlation between x and y for restriction of range can be obtained using the following equation (direct truncation on predictor) by Guion and Highhouse (2006):

$$r_n = \frac{r_o \cdot \frac{s_{xn}}{s_{xo}}}{\sqrt{1 - r_o^2 + r_o^2 \cdot \frac{s_{xn}^2}{s_{xo}^2}}}$$

r_n = new estimate of the coefficient for an unrestricted sample

r_o = old (obtained) coefficient for the available sample

s_{xn} = predictor standard deviation for unrestricted group

s_{xo} = predictor standard deviation for restricted group

The following formula by Ghiselli, Campbell and Zedeck (1981) was applied in order to determine the predictor standard deviation for the unrestricted group:

$$\sigma_c^2 = \sigma_1^2 + \dots + \sigma_k^2 + 2\sigma_1\sigma_2r_{12} + \dots + 2\sigma_{k-1}\sigma_k r_{(k-1)k}$$

Table 4.12 indicates the effect of the corrections for both unreliability and restriction. Two coefficients are presented for the relationships involving each predictor and the BOS measure. The first coefficients are the obtained Pearson product-moment correlation coefficients from the restricted sample bank branch managers. The second coefficients are the correlation coefficients corrected for unreliability and restriction of range.

Table 4.12

Correlations between the Predictors and the Behavioural Observation Scale (BOS) corrected for Unreliability and Restriction of Range

Correlations		
Variable		BOS
Verbal evaluation	Pearson Correlation	(.06)
	Sig. (1-tailed)	.283
Numerical interpretation	Pearson Correlation	(.15)
	Sig. (1-tailed)	.122
Basic checking	Pearson Correlation	(-.02)
	Sig. (1-tailed)	.451
Extraversion	Pearson Correlation	(-.01)
	Sig. (1-tailed)	.356
Openness to Experience	Pearson Correlation	(.00)
	Sig. (1-tailed)	.450
Emotional Stability	Pearson Correlation	(-.01)
	Sig. (1-tailed)	.237
Agreeableness	Pearson Correlation	(.00)
	Sig. (1-tailed)	.106
Conscientiousness	Pearson Correlation	(.01)
	Sig. (1-tailed)	.234
SJT	Pearson Correlation	{.247**}
	Sig. (1-tailed)	.008
OPR	Pearson Correlation	{.868**}
	Sig. (1-tailed)	.000
Perf Rank	Pearson Correlation	{.647**}
	Sig. (1-tailed)	.000
BOS	Pearson Correlation	1.000
	Sig. (1-tailed)	
MPerf Comp	Pearson Correlation	{.949**}
	Sig. (1-tailed)	.000

Note. Values indicated in { } are corrected for unreliability and restriction of range in the criterion.

** Correlation is significant at the 0.01 level (1-tailed)

BOS – Behavioural Observation Scale

OPR – Overall Performance Rating

Perf Rank – Performance Ranking

MPerf Comp – Managerial Performance Composite

4.5.3 Regression Results

Since the relationships between individual variable pairs were discussed in the preceding section, the joint contribution of these variables to explaining variance in the criterion measure will consequently be addressed. In this section, the overall procedure followed in the regression analyses will be discussed first, followed by the results obtained in each phase of the analysis.

Standard multiple regression analysis was used to predict the managerial performance scores from the full set of predictor variables, namely cognitive ability, personality and situational judgment test variables. Standard multiple regression is a statistical technique used to analyse the relationships between a single dependent (criterion variable) and several predictor variables (Tabachnick & Fidell, 1996). In order to examine the extent to which the SJT provides incremental validity in the prediction of managerial performance, hierarchical multiple regression analysis was conducted. The focus in hierarchical regression is on “the change in predictability associated with predictor variables entered later in the analysis over and above that contributed by predictor variables entered earlier in the analysis” (Petrocelli, 2003, p. 11). Since the research initiating question for the present research focused on the added predictive value to be gleaned from a SJT measure, the substantive hypothesis regarding the incremental validity of SJT scores beyond other predictor measures required testing. Therefore, sets of variables were systematically entered into various hierarchical regression models: cognitive ability measures first; personality measures second; and the situational judgement measure in the third step. Partial correlations, which are measures of relative incremental prediction across all predictors in the regression (O’Connell et al., 2007) model, are subsequently reported.

The researchers initially assumed that all the measures of managerial performance would correlate highly, thereby making multiple measures redundant. However, this was not the case (refer to section 4.5.1) and it was found that the predictor-criterion correlation patterns varied substantially, depending on the type of criterion used. In other words, selected predictors would relate to only some of the criterion measures, and not all. As a

result, in order to test the incremental validity hypothesis, it was decided to conduct the regression analyses for the best indicator of managerial performance, i.e., the one that provides the broadest measure of the latent variable, i.e. the Managerial Performance Composite (MPerf Comp). However, for the purposes of confirmation of findings and further illumination of the correlation results obtained earlier, regression analyses were conducted separately for the other three measures as well. However, these results are reported broadly with only the most important resultant statistics.

The regression results are reported firstly for the standard multiple regression, i.e., all predictors enter jointly into a regression model, only to be followed by the hierarchical regression results, where variable sets enter the regression model consecutively by an order determined by the researcher (Cohen & Cohen, 1975).

4.5.3.1 Standard Multiple Regression of Managerial Performance Composite on all Predictors (H₁₀)

Standard multiple regression was performed with managerial performance (as measured by the Managerial Performance Composite) as the dependent variable, and cognitive ability measures (i.e., *verbal evaluation, numerical interpretation, basic checking*), personality measures (i.e., *emotional stability, agreeableness, conscientiousness and extraversion*) and the Situational Judgement Test measure scores, as the independent variables (IV). From this list of IVs, it is clear that not all the Big Five factors were entered into the regression model, but only those hypothesised to relate to managerial performance. The analyses were performed using SPSS REGRESSION. Prior to the main analysis, SPSS EXPLORE was used for the evaluation of assumptions underlying multiple regression analysis (Tabachnick & Fidell, 1996). With the use of a $p < .001$ criterion for Mahalanobis distance, no multivariate outliers among the cases were identified. A summary of the results of the regression analysis is presented in Table 4.13.

R for Regression was not significantly different from zero, $R = .366$, $F(8, 83) = 1.607$, $p > .05$, with R^2 at .134. The adjusted R^2 value of .051 indicates that 5% of the variability in

general managerial performance is explained by the independent variables. In this model, agreeableness makes the largest contribution ($\beta = -.246$, p (one-tailed) $< .05$), thereby explaining unique variance in the criterion not explained by the other independent variables. No other predictor in the model showed a statistically significant beta coefficient, ps (one-tailed) $> .05$. Given these findings, the null hypothesis H_{010} , stating that managerial job performance can not be reliably predicted from a battery consisting of measures of specific cognitive abilities (i.e., *verbal evaluation*, *numerical interpretation* and *basic checking ability*), personality measures (i.e., *emotional stability*, *agreeableness*, *conscientiousness* and *extraversion*) and the Situational Judgement Test (SJT) measure, can therefore not be rejected. However, this result needs to be interpreted in the light of the relatively high variables-to-cases ratio, which makes the multiple correlation coefficient ($R = .37$) less likely to achieve statistical significance. This interpretation is supported by the substantial drop in the value of R^2 (.134) to Adjusted R^2 (.051). Whereas the former represents the joint contribution of all the variables to explaining variance in the criterion measure, the latter represents an adjustment for possible sampling error due to small sample sizes.

Before the hierarchical multiple regression results are presented, the reader is reminded that the present analysis (standard multiple regression) was also conducted for the constituent measures of managerial performance, for which remarkably different results were obtained. The results clearly show a marked difference in the proportion of variance explained by the joint sets of predictors when other performance measures are used as criterion variables. For example, the multiple correlation coefficients for the Performance Ranking ($R = .442$, $F [8, 85] = 2.576$, $p < .05$, with R^2 at .195), BOS ($R = .308$, $F [8, 96] = 1.254$, $p > .05$, with R^2 at .095) and Overall Performance Rating measures ($R = .318$, $F [8, 97] = 1.368$, $p > .05$, with R^2 at .101) varied substantially. On the one hand, the predictor set predicted the BOS scores the least well, while, on the other hand, it predicted the Performance Ranking scores the best, the latter which was highly statistically significant, $p_{\text{obt}} = .014$. The possible reasons for these discrepancies are discussed in chapter five.

Table 4.13

Standard Multiple Regression of all Managerial Performance Composite on all Predictors

Model Summary ^b												
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson		
					R Square Change	F Change	df1	df2	Sig. F Change			
1	.366 ^a	.134	.051	.87972	.134	1.607	8	83	.135	1.826		
a. Predictors: (Constant), SJT, Basic checking, Emotional stability3, Agreeableness3, Conscientious3, Verbal evaluation, Extraversion3, Numerical interpretation												
b. Dependent Variable: MPerf Comp												
ANOVA ^b												
Model	Sum of Squares		df	Mean Square	F	Sig.						
1	Regression		9.947	8	1.243	1.607 .135 ^a						
	Residual		64.234	83	.774							
	Total		74.181	91								
a. Predictors: (Constant), SJT, Basic checking, Emotional stability3, Agreeableness3, Conscientious3, Verbal evaluation, Extraversion3, Numerical interpretation												
b. Dependent Variable: MPerf Comp												
Coefficients ^a												
Model	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval for B		Correlations		Collinearity Statistics	
	B	Std. Error	Beta				Lower Bound	Upper Bound	Zero-order	Partia l	Part	Tolerance
1 (Constant)	-1.027	1.614			-.636	.526	-4.238	2.184				
Verbal evaluation	.013	.012	.119		1.030	.306	-.012	.037	.153	.112	.105	.781 1.281
Numerical interpretation	.013	.030	.059		.450	.654	-.046	.073	.179	.049	.046	.616 1.625
Basic checking	.014	.014	.118		.988	.326	-.014	.043	.143	.108	.101	.729 1.371
Extraversion3	-.031	.128	-.029		-.243	.809	-.286	.223	-.015	-.027	-.025	.735 1.361
Emotional stability3	-.122	.112	-.127		-1.086	.281	-.345	.101	-.105	-.118	-.111	.764 1.308
Agreeableness3	-.226	.097	-.246		-2.324	.023	-.420	-.033	-.207	-.247	-.237	.930 1.076
Conscientious3	-.045	.105	-.047		-.433	.666	-.253	.163	-.002	-.047	-.044	.882 1.133
SJT	.033	.020	.174		1.593	.115	-.008	.073	.136	.172	.163	.874 1.145
a. Dependent Variable: MPerf Comp												

4.5.3.2 Hierarchical Regression of Managerial Performance Composite on all Predictors (H₁₁)

Hierarchical (Sequential) regression was used to determine if the addition of information regarding predictors of managerial performance improved prediction of this criterion beyond that afforded by the predictors individually. Analysis was performed using SPSS REGRESSION and SPSS EXPLORE was utilised for the evaluation of assumptions. With the use of a $p < .001$ criterion for Mahalanobis distance, no multivariate outliers among the cases were identified.

Table 4.14 displays the hierarchical regression results. R was not found to be significantly different from zero at the end of any of the steps. After step three, with all independent variables in the equation, $R = .366$, $F(8, 83)$ and $R^2 = .134$, $F(8, 83) = 1.607$, $p > .05$. The adjusted R^2 value of .051 indicates that 5% of the variability in managerial performance is predicted by measures of verbal evaluation, numerical interpretation, basic checking, personality and the Situational Judgement Test.

After step one, with verbal evaluation, numerical interpretation and basic checking in the equation, $R^2 = .042$, $F_{inc}(3, 88) = 1.292$, $p > .05$. After step two, with conscientiousness, emotional stability, agreeableness and extraversion in the equation, $R^2 = .108$, $F_{inc}(4, 84) = 1.539$, $p > .05$. The addition of the personality variables to the equation resulted in an increase in R^2 and the adjusted R^2 ; however these increases were not significant. After step three, with the Situational Judgement Test in the equation, $R^2 = .366$, $F_{inc}(1, 83) = 2.538$, $p > .05$ showing that the addition of the Situational Judgement Test to the equation did not reliably improve R^2 .

Considering the regression results when using the composite measure of managerial performance, two aspects deserve highlighting. Firstly, it appears that the pattern of results suggests that the proportion of variability in managerial performance obtained when all predictors are entered jointly into the model is relatively small, at 5%. The model for which this result was obtained was also not statistically significant. In this

model, agreeableness was the only significant contributor ($\beta = -2.324, p < .05$). It is noteworthy that, in this case, the beta coefficient was negative. Secondly, when considering the incremental validity hypothesis (and still using the composite measure of performance as a criterion) the results showed that the addition of variable sets did not improve the predictiveness of the model at successive steps. While R , R^2 and the adjusted R^2 increased with the addition of the personality variables and the SJT, the changes were not statistically significant. Therefore, the personality variables and the SJT did not add incremental validity to the measures of cognitive ability. These results confirm the findings from the standard multiple regression analyses, in other words, managerial job performance (when measured by a broad composite measure) can not be reliably predicted from a battery consisting of measures of specific basic cognitive abilities, specific personality constructs and an SJT. Furthermore, the Situational Judgement Test scores did not provide incremental validity in the prediction of managerial performance relative to the predictability provided jointly by cognitive ability and personality measures. Based on the findings of the regression analyses and the above discussion, H_{010} and H_{011} cannot be rejected.

Prior to reaching preliminary conclusions based on the aforementioned results, the hierarchical multiple regression results for the specific measures of managerial performance from which the composite measure was computed, must be considered (see Table 4.13). Here, it is evident that incremental validity (also, statistically significant, with $R^2 = .118, F_{\text{inc}}(3, 90) = 4.013, p < .05$) is achieved after adding personality measures to a battery of cognitive ability measures, when Performance Ranking is used as a measure of managerial performance. Of greater importance, though, is the result that is achieved when considering the BOS measure of managerial performance. In this case, the addition of neither the measures of basic specific cognitive abilities, nor the specific personality measures, improve the ability of the model to predict BOS scores at a greater-than-chance ($p < .05$) level. However, when the SJT measure is added to this model, incremental validity is achieved ($R^2_{\text{change}} = .051, F_{\text{inc}}(1, 96) = 5.441, p < .05$). Interpreted differently, the SJT's inclusion in the regression model improves the ability of measures of basic cognitive abilities and personality to predict managerial performance by 5%. The

possible reasons for the finding of incremental validity (yielded by the SJT) that is limited only to cases where the BOS operationalisation of managerial performance is used, are discussed in chapter five.

Table 4.14

Hierarchical Regression of Managerial Performance on all Predictors

Model Summary ^d										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.205 ^a	.042	.010	.89855	.042	1.292	3	88	.282	
2	.328 ^b	.108	.033	.88773	.065	1.539	4	84	.198	
3	.366 ^c	.134	.051	.87972	.026	2.538	1	83	.115	1.826

a. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation
b. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation, Conscientious3, Emotional stability3, Agreeableness3, Extraversion3
c. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation, Conscientious3, Emotional stability3, Agreeableness3, Extraversion3, SJT
d. Dependent Variable: MPerf Comp

ANOVA ^d						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.130	3	1.043	1.292	.282 ^a
	Residual	71.050	88	.807		
	Total	74.181	91			
2	Regression	7.983	7	1.140	1.447	.198 ^b
	Residual	66.198	84	.788		
	Total	74.181	91			
3	Regression	9.947	8	1.243	1.607	.135 ^c
	Residual	64.234	83	.774		
	Total	74.181	91			

a. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation
b. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation, Conscientious3, Emotional stability3, Agreeableness3, Extraversion3
c. Predictors: (Constant), Basic checking, Verbal evaluation, Numerical interpretation, Conscientious3, Emotional stability3, Agreeableness3, Extraversion3, SJT
d. Dependent Variable: MPerf Comp

Table 4.14 Continued

		Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations		Collinearity Statistics		
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-1.057	.802		-1.319	.191	-2.650	.536					
	Verbal Evaluation	.009	.012	.086	.736	.464	-.015	.034	.153	.078	.077	.800	1.250
	Numerical Interpretation	.025	.029	.109	.853	.396	-.033	.083	.179	.091	.089	.670	1.492
	Basic Checking	.008	.014	.068	.568	.572	-.020	.036	.143	.060	.059	.768	1.303
2	(Constant)	.509	1.307		.389	.698	-2.089	3.107					
	Verbal Evaluation	.012	.012	.118	1.013	.314	-.012	.037	.153	.110	.104	.781	1.280
	Numerical Interpretation	.021	.030	.094	.722	.472	-.038	.080	.179	.079	.074	.634	1.579
	Basic Checking	.011	.014	.092	.772	.442	-.017	.039	.143	.084	.080	.743	1.346
	Extraversion3	.000	.128	.000	-.006	.995	-.255	.253	-.015	.000	.000	.752	1.330
	EmotStability3	-.127	.113	-.132	-1.118	.267	-.352	.099	-.105	-.121	-.115	.765	1.307
	Agreeableness3	-.203	.097	-.221	-2.090	.040	-.396	-.010	-.207	-.222	-.215	.951	1.052
	Conscientious3	.000	.102	.000	.003	.998	-.202	.202	-.002	.000	.000	.954	1.049
3	(Constant)	-1.027	1.614		-.636	.526	-4.238	2.184					
	Verbal Evaluation	.013	.012	.119	1.030	.306	-.012	.037	.153	.112	.105	.781	1.281
	Numerical Interpretation	.013	.030	.059	.450	.654	-.046	.073	.179	.049	.046	.616	1.625
	Basic Checking	.014	.014	.118	.988	.326	-.014	.043	.143	.108	.101	.729	1.371
	Extraversion3	-.031	.128	-.029	-.243	.809	-.286	.223	-.015	-.027	-.025	.735	1.361
	EmotStability3	-.122	.112	-.127	-1.086	.281	-.345	.101	-.105	-.118	-.111	.764	1.308
	Agreeableness3	-.226	.097	-.246	-2.324	.023	-.420	-.033	-.207	-.247	-.237	.930	1.076
	Conscientious3	-.045	.105	-.047	-.433	.666	-.253	.163	-.002	-.047	-.044	.882	1.133
	SJT	.033	.020	.174	1.593	.115	-.008	.073	.136	.172	.163	.874	1.145

a. Dependent Variable: MPerf Comp

Table 4.15

Summarised Hierarchical Regression results of the BOS, Perf Rank and OPR Performance Measures

Model Summary											
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Predictors	
					R Square Change	F Change	df1	df2	Sig. F Change		
	1	.343 ^a	.118	.089	.261088	.118	4.013	3	90	.010	BC, VE, NI
Perf Rank	2	.433 ^b	.187	.121	.256409	.069	1.829	4	86	.131	BC, VE, NI, Consc, ES, A, E
	3	.442 ^c	.195	.119	.256645	.008	.841	1	85	.362	BC, VE, NI, Consc, ES, A, E, SJT
	1	.136 ^a	.019	-.011	.64623	.019	.638	3	101	.592	BC, VE, NI
BOS	2	.208 ^b	.043	-.026	.65107	.025	.627	4	97	.645	BC, VE, NI, Consc, ES, A, E
	3	.308 ^c	.095	.019	.63666	.051	5.441	1	96	.022	BC, VE, NI, Consc, ES, A, E, SJT
	1	.212 ^a	.045	.017	1.92999	.045	1.601	3	102	.194	BC, VE, NI
OPR	2	.286 ^b	.082	.016	1.93055	.037	.985	4	98	.419	BC, VE, NI, Consc, ES, A, E
	3	.318 ^c	.101	.027	1.91975	.020	2.105	1	97	.150	BC, VE, NI, Consc, ES, A, E, SJT

Note. Perf Rank: $N = 98$; BOS: $N = 108$; OPR: $N = 111$

BC: Basic checking; VE: Verbal evaluation; NI: Numerical interpretation; Consc: Conscientiousness; ES: Emotional stability; A: Agreeableness;

E: Extraversion; SJT: Situational Judgement Test

4.5.4 Do Situational Judgment Test (SJT) Scores Mediate the Influence of Personality Measures on Managerial Performance? (H₁₂)

The twelfth, and last, null hypothesis (H₀₁₂) stated that SJT scores will not mediate the effects of conscientiousness, agreeableness, and emotional stability on managerial performance. Although there are various methods to test for mediation, that include structural equation modeling (SEM; Cole & Maxwell, 2003) and bootstrapping (MacKinnon, Lockwood & Williams, 2004; Shrout & Bolger, 2002), the regression-based tests for testing mediating effects is still the most prevalent (Fritz & MacKinnon, 2007). In order to test for mediation effects, adequate sample size is required to detect the mediated effect (Fritz & MacKinnon, 2007). In their article, Fritz and MacKinnon (2007) present necessary sample sizes for six of the most common and the most recommended tests of mediation for various combinations of parameters. Using their tables (see appendix C), however, it was apparent that the present sample does not contain enough participants to achieve adequate power to test for mediation. Therefore, the formal test of H₀₁₂ was abandoned.

In conclusion, we computed both the power and expected cross-validity of our standard multiple regression models using G*Power version 1.3.0. (Faul, Erdfelder, Lang & Buchner, 2007). Firstly, the achieved power was calculated post hoc, for a given error probability ($\alpha = .05$), sample size ($N = 98 - 111$), effect size f^2 (small to medium) and number of predictors (8). All power estimates ranged between .67 (for predicting the composite criterion) and .87 (for predicting the Performance Ranking criterion), which were deemed acceptable for the present study. However, the relatively lower level power (.67) for the composite criterion regression analysis may have led to a lowered ability of the analysis to detect possible effects, which cautions the reader to interpret the lack of significant multiple correlation coefficients in this specific case within the light of inadequate statistical power.

Table 4.16 provides the cross-validity estimates for the various standard regression models. It is obvious from this table that the regression coefficients can be expected to show more stability when investigating similar regression models in new samples in the case of the Performance Ranking criterion. It is also evident that the relatively low

multiple correlation coefficient, combined with the number of predictors:sample size ratio in the case of the BOS criterion leads to lower cross-validity estimates.

Table 4.16

Cross validity estimates for Standard Multiple Regression Models

Estimate	Criterion (DV)			
	MPerf Comp	BOS	Perf Rank	OPR
Population multiple correlation ^a	.24	.15	.35	.17
Population cross-validity ^b	.163	.082	.284	.107

a. Calculated using Wherry's formula (Schmitt & Chan, 1998)

b. Cattin (1980) formula estimates (Schmitt & Chan, 1998)

4.6 CONCLUSION: CHAPTER FOUR

The findings from the inter-correlations and regression results revealed that managerial performance, as measured by the Managerial Performance Composite, could not be predicted at better than chance levels by using a battery of predictors. However, certain predictors were valid predictors of managerial performance, amongst others, measures of numerical interpretation ability (i.e., cognitive factors), as well as measures of agreeableness (i.e., personality). However, when decomposing the composite managerial performance measure into distinct measures that differ in terms of their format of administration and item content, and regressing these measures against the predictor set, strikingly different results were obtained. For example, the Performance Ranking measure correlated significantly with all three specific cognitive ability measures, as well as negatively with agreeableness. On the other hand, the BOS measure correlated significantly with the SJT, whereas the Overall Performance Rating correlated significantly with verbal evaluation and numerical interpretation, and negatively with agreeableness.

Similar inconsistent findings were obtained from the regression results. The standard multiple regression model (on the Managerial Performance Composite) shows that, in combination, the predictors account for 5% of the variance in managerial

performance, of which agreeableness makes the largest significant contribution. The hierarchical regression model using the Managerial Performance Composite, Performance Ranking, as well as the overall performance criterion, indicated that the SJT did not exhibit meaningful or statistically significant incremental prediction over cognitive ability and personality on the Managerial Performance Composite criterion. However, the hierarchical regression model using the BOS, demonstrated that the SJT provided incremental validity over cognitive ability and personality. Therefore, even though not all the hypotheses in the current study were supported, further discussion of these results is necessary in order to contextualise the findings in the body of literature, and to finally draw conclusions, which follows in chapter five.

CHAPTER FIVE

DISCUSSION OF RESULTS

5.1 INTRODUCTION

Based on the research findings presented in chapter four, this final chapter discusses the ways in which the present results relate to previous research (presented in chapter two). Subsequently, we interpret the results of our data analyses to form general conclusions that can be drawn from the findings. Thereafter, the limitations of the study, in terms of (a) the generally recognised limitations of survey research, and (b) the specific potential problems of the present study, are provided. The chapter concludes with recommendations for future research.

5.2 GENERAL CONCLUSIONS

The objective of the current study was to determine whether a Situational Judgement Test (SJT) would significantly add to the prediction of managerial performance beyond other measures used for managerial selection, such as measures of cognitive ability and personality. It further aimed to investigate the relationship between the SJT and such predictor measures to shed light on the construct validity of the SJT, i.e., which latent person characteristics the SJT is measuring. For these purposes, the researchers tested the hypotheses within a sample of branch managers in a South African bank.

Three criterion measures, i.e., a Behavioural Observation Scale (BOS), Performance Ranking (Perf Rank) measure, and an Overall Performance Rating (OPR), were used to gather data on the performance of the participants. An unweighted linear composite (MPerf Comp) of the standardised scores of these individual measures was created and ultimately used to decide whether to reject or confirm the hypotheses that were formulated. We intended to overcome the weaknesses of each individual measure — these generally suffer from methodological weaknesses — by integrating the three distinct measures into a composite variable. It was decided to include the specific performance measures in conjunction with the composite measure in order to explore

the effect of using different performance measures on predictive validity of the predictor set, especially considering Bartram's (2007) research indicating that validity can be increased by using forced-choice criterion measures over normative measures. The results of our statistical analysis and the subsequent findings relating to the hypothesised relationships will therefore be discussed next. We use the composite criterion measure's regression results in order to explain the main findings, but will continuously refer to the results for the specific managerial performance measures. Firstly, the inter-correlation results are discussed, followed by the presentation of the multiple regression results. Finally, an holistic integration of results is presented.

5.2.1 Inter-correlations

In our first set of analyses, we used Pearson's product-moment correlation coefficients, r , to test the plausibility of various hypothesised relationships between the study variables that were based on the research model proposed in the culmination of the literature review. These results will consequently be integrated with the existing body of knowledge, per hypothesised relationship.

5.2.1.1 Relationship between Cognitive Ability and Managerial Performance

It is well accepted in the literature that general cognitive ability creates variance in job performance in general, as well as in managerial performance in particular (see literature review). In Baydoun and Neuman's (1992) review of management selection techniques, they suggest that, of the various methods, cognitive ability testing provides the strongest criterion-based validity when used for managerial selection. Furthermore, in their seminal article, Schmidt and Hunter (1998) found that up to 33% of managerial job performance is accounted for by estimates of managers' cognitive ability.

In the present study, we did not use the traditional set of measures that tap fluid intelligence or general cognitive ability (g). The cognitive measures' scores used in the present research were extracted from a selection database that yielded the scores that were used to screen an incumbent sample of bank branch managers. Mostly by nature of default, these instruments tapped very basic cognitive abilities that appear to

represent lower-order cognitive abilities, as compared to higher-order cognitive factors. Even though the bulk of prior research into cognitive ability predictors of managerial job performance used general cognitive ability measures, we expected that these specific cognitive abilities would also predict managerial performance, but would probably be smaller in magnitude than those observed when higher-order cognitive abilities are measured. These specific abilities were identified as being critical to job success in the job analysis and competency modelling process that was followed in this study. Managers are required to evaluate verbal (written) information, to interpret financial reports and to execute a control function by overseeing the correctness of client records, legal and fiduciary documentation and others. It was, therefore, hypothesised that the three specific cognitive abilities – verbal evaluation, numerical interpretation and basic checking – impact directly on managerial performance. The results, however, revealed inconsistent correlations for the prediction of the different criterion measures due to the difference in the nature of the criterion measures.

Cognitive factors: Even though verbal evaluation ability significantly correlated with the Performance Ranking criterion measure and the Overall Performance Rating criterion measures, it did not relate to the Managerial Performance Composite. The results also lent support for our second hypothesis, which was that numerical interpretation ability would predict the Managerial Performance Composite. Lastly, basic checking ability did not appear to play a role in the Managerial Performance Composite. This ability (basic checking) was however significantly related to the Performance Ranking criterion measure.

To summarise, all three specific cognitive abilities, verbal evaluation, numerical interpretation, and basic checking had significant correlations with the Performance Ranking criterion measure (refer to Table 5.1). Two of the three specific cognitive abilities namely, verbal evaluation and numerical interpretation, significantly correlated with the Overall Performance Rating measure. None of the specific cognitive abilities correlated significantly with the BOS, and consequently only numerical interpretation was found to be significantly correlated with the Managerial Performance Composite.

Table 5.1

Correlations between Cognitive Ability Predictors and Criteria

	Correlations			
	OPR	Perf Rank	BOS	MPerfCom p
Verbal evaluation	.171*	.247**	.056	.153
Numerical interpretation	.188*	.288**	.113	.179*
Basic checking	.087	.262**	-.012	.143

** . Correlation is significant at the 0.01 level (1-tailed)

* . Correlation is significant at the 0.05 level (1-tailed)

As one would expect, cognitive abilities do appear to play a role in managerial performance, similar to prior research (e.g., Baydoun & Neuman, 1992; Schmidt & Hunter, 1998). However, two aspects are worthy of mention. Firstly, it appears that specific cognitive abilities that resemble lower-level cognitive functions (i.e., basic clerical actions) are *less* predictive of higher complexity jobs such as the class of position studied in the present research due to the general overall nature of the performance measure. In the present study, perhaps basic cognitive abilities were required as a minimum in order to qualify for the present position, but as tenure in the position increases, higher level cognitive skills become more salient as the demands of the position rise in terms of its complexity. This would typically be observed in a bivariate heteroscedastic distribution, i.e., where the variance of basic cognitive abilities is greater at lower levels of job performance, but more proficient bank managers could be expected to vary less in terms of basic cognitive abilities and more on higher level cognitive functioning. Stated more directly, basic cognitive abilities would play a role in performance only up to a certain point (e.g., passing the screening test for entry into a junior branch manager position), but higher levels of these abilities would lead to little increases in job performance beyond these initial stages. This interplay between basic and higher level cognitive abilities deserve greater research attention, since this topic has typically received little research attention. The present research, however, does shed some light on the role of basic cognitive abilities in higher level job performance. A corollary of this issue is the possibility that the present sample was restricted in terms of the variance of lower level cognitive abilities, which would have resulted in the deflation of observed

correlation coefficients (Schmitt & Chan, 1998), similar to those found in terms of the cognitive ability-performance bivariate relationships reported in the present study.

The second aspect that requires mention in terms of the results that relate to the cognitive ability measures, is the finding that the basic cognitive abilities related to only *certain* criterion measures. This is in line with what Bartram (2007) found in his study that validities are increased with the use of forced-choice criterion measures. Although the BOS was developed to represent a high-fidelity measure of branch performance, one interpretation of the low variance could have been that the supervisors did not discriminate sufficiently between the different statements and rated individuals highly on all the BOS statements. Another interpretation could be that the distribution of performance in the population of branch managers is such that one would not expect variance to be large, since sub-minimum levels of performance would have been weeded out by self-selection, performance management, or even turnover of poor performers (Murphy & Davidshofer, 1998). Thus, range restriction in the criterion measure scores represents a realistic depiction of (relative) restricted levels of job performance. Another possible explanation could be gleaned from taking the rater perspective. It could be speculated that supervisor raters could have been focussing more on general person characteristics of managers when assigning ratees into rank-order positions. These characteristics may not have been purely job relevant behaviours (i.e., in contrast to the BOS measure items), while they were assigning Performance Rankings. The target characteristics raters had in mind could have been influenced by their perceptions of the proficiency of target bank managers at lower level abilities, similar to the ones measures by the cognitive measures, such as proficiency at clerical tasks, accuracy of checking of reports and documents, and basic verbal skills. The BOS, on the other hand, had the intention (by design) of forcing the raters to evaluate the managers in the light of specific normative performance dimensions. These performance dimensions had little to do with the type of tasks in which basic cognitive abilities (e.g., checking) could have been expected to play a role. Therefore, the lack of job-relatedness of these basic cognitive abilities simply manifested in low predictor-criterion relationships, partly due to a lack of conceptual overlap with the core task dimensions associated with the target job.

Another possible explanation for the low correlations on the BOS could be found in restriction of range. The sample was highly homogenous, as job incumbents rather than applicants were used in this study. Little variance was found in the BOS scores, as raters generally scored the managers high on performance (see descriptive statistics). Poor management is normally easily noticed and those managers are typically dismissed, thereby reducing the variance and consequently, observed correlations. This is not an uncommon finding in validation research of this kind, and these effects have been found to be substantial (Murphy & Davidshofer, 2005).

Finally, as mentioned previously, the bulk of prior research into cognitive ability predictors of managerial performance used *general* cognitive ability measures. In contrast, the current study investigated *specific* cognitive abilities as predictors and since such measures represent minimum-entry requirements for the position at stake, we did not expect the obtained correlations to be large. We, therefore, recommend the use of measures of higher level cognitive abilities in future selection research relating to similar positions, although the present findings regarding the predictive validity of basic cognitive abilities should be extended to research that links the dynamic interplay between basic and higher-order cognitive abilities (e.g., using the hierarchical model of intelligence of Carrol, 1993) to career progression through jobs that vary in complexity. Another recommendation would be to investigate the relationship between specific abilities and detailed multiple criteria instead one overall job performance score as found in the literature.

5.2.1.2 Relationship between Personality and Managerial Performance

Apart from cognitive ability, personality factors were also expected to predict managerial performance as they influence a person's work-related behaviour, namely their job performance. Recent meta-analytic studies of personality and job performance have confirmed that the Big Five can serve as a useful tool in personnel selection, although some of the factors have proven to be more useful than others (e.g., Barrick & Mount, 1991; Barrick et al., 2001; Borman et al., 2001; Tett et al., 1991). It was hypothesised that two of the Big Five factors, conscientiousness and extraversion, would influence managerial job performance directly.

This is in line with what Bartram (2007) found in his study that validities are increased with the use of forced-choice criterion measures. Although the BOS was developed to represent a high-fidelity measure of branch performance, the supervisors did not discriminate sufficiently between the different statements and rated one individual high on all the statements. The restriction of range found on the BOS criterion measure seems to be evidence of this rater error. This is a limitation of the study and should be taken into consideration when interpreting the results.

Conscientiousness and Extraversion: Conscientiousness was expected to be positively related to managerial performance, because the former assesses personal characteristics, such as responsibility, carefulness, the ability to plan, persistence, the capacity to work hard and the ability to follow instructions and procedures, which are important attributes for accomplishing managerial tasks. Some researchers have even proposed that conscientiousness might be the ‘g’ of personality and predict performance in most occupational areas (Robertson, Baron, Gibbons, MacIver & Nyfield, 2000). In addition, extraversion was expected to be positively related to managerial performance due to the frequent interactions of managers with others, as well as their responsibility for influencing others in their leadership and supervisory roles. The quality of extraversion relates to personal characteristics, such as being active, talkative and assertive. The low correlations (see Table 5.2) between conscientiousness and the Managerial Performance Composite, and extraversion and the Managerial Performance Composite, are therefore not easily explained. Both the literature (e.g. Barrick & Mount, 1991; Salgado, 1997) and the job analysis indicated that these two Big Five factors are related to managerial performance.

One possible explanation for the contradicting findings is that, even though none of the criterion measures focussed the raters’ attention on personality traits per se, the Overall Performance Rating and Performance Ranking measures could have focussed the raters’ attention on general person characteristics of managers, which may not have been job relevant. This would have resulted in lower observed correlations between job performance measures and personality traits. While the BOS measures forced the raters to evaluate the managers in the light of specific performance dimensions, it could have been expected to be even less related to personality traits. This question cannot be conclusively answered from the present data. However, the

simpler explanation is probably more plausible. It is possible that personality factors simply play less of a role in driving performance in as far as it is viewed from the perspective of regional managers. Regional managers typically assess branch manager performance in terms of objective criteria (e.g., sales, compliance, customer feedback) and not based on personal interaction. Therefore, personality factors would not necessarily impact on regional managers' ratings of branch manager job performance. This notion could be extended to another perspective, for future research. Consistent with this idea, subordinates would be expected to base their assessments of branch managers more on personality characteristics than would regional managers. In their day-to-day interactions with branch managers, they could also be expected to come into contact with success relevant behaviours by branch managers driven by personality factors such as extraversion. Therefore, we expect subordinate rating sources to shed another interesting angle on the predictive validity of certain personality traits.

Again, another possible explanation for the insignificant correlations on the different criterion measures could be restriction of range. During the selection process in the host organisation, the aim is to select individuals with a particular type of personality profile. As a result, this organisation should show a consistent personality profile, which implies that many other personality profiles are missing. In addition, it is possible that most individuals with a particular personality profile leave or get dismissed, while those with a somewhat different profile tend to get promoted. The corporate culture can therefore, over time, easily lead to significant homogeneity of employees. This finding has been confirmed by recent research that found support for this 'homogeneity hypothesis' in terms of personality profiles in organisations (Schneider, Smith, Taylor & Fleenor, 1998).

Table 5.2

Correlations between Personality Measures and Criteria

Correlations				
	OPR	Perf Rank	BOS	MPerf Comp
Conscientiousness	-.003	.026	.070	-.002
Extraversion	-.029	.007	-.036	-.015
Agreeableness	-.167*	-.194*	-.121	-.207*
Emotional Stability	-.045	-.104	-.070	-.105
Openness to Experience	.025	.084	-.012	.072

*. Correlation is significant at the 0.05 level (1-tailed)

Agreeableness: Various meta-analyses, on testing the relationship between agreeableness and managerial performance, found very low, but positive correlations (e.g., Barrick & Mount, 1991; McDaniel & Nguyen, 2001; Tett & Christiansen, 2007). A significant finding, though contrary to the extant literature, relates to the significant negative correlation that was found between agreeableness and managerial performance, regardless of criterion measure used. In the present research, agreeableness had a significant but negative correlation with the Managerial Performance Composite. Based on the specific nature of the industry within which this study was conducted, a rival hypothesis for this finding could be offered. Agreeableness, as a personality trait, represents an individual's desire to be liked and accepted by others. Agreeable individuals therefore tend to be cooperative so as to maintain good relationships. However, managers are required to lead and make decisions based on what is the best in a certain situation, rather than to be influenced by their need for acceptance. This is of high relevance in the banking industry: Managers are required to adhere to very strict policies and procedures such as the credit policy. For example, a branch manager with high levels of agreeableness will find it extremely hard to decline a loan to a client who is in desperate need of money. Therefore, a person lower in agreeableness would perform better in this position. In this sense, the present research highlights the importance of considering the specific nature of the situation when interpreting the results of predictor-criterion relationships. Without a solid understanding of the nature of the organisational context, as well as that of the target job, the prospect of interpreting observed validity coefficients becomes increasingly unattainable. Moreover, the choice of selection

instruments should also strongly consider the specific nature of the target position. For instance, a sole reliance on the meta-analytic results to choose measures for branch manager selection could have, in this case, turned out to be a poor decision. It appears that the managerial performance, despite the generic performance models discussed in the literature review, also lends itself to industry specific contexts that demand a judicious choice of selection measures that also take into account the specific context of the industry.

5.2.1.3 Mediating Affect of Individual Differences on Performance on the Situational Judgement Test

Even though statistical power considerations precluded mediation analysis in the present research, tentative indications about these effects could be gathered from the present results. Consequently, we discuss the relationships between personality measures, SJT performance, and managerial performance. Available research directed towards understanding the constructs measured by SJTs (e.g., McDaniel et al., 2007), or why they are related to performance, is inconclusive. Therefore, one of the correlary objectives of the present study was to understand the correlates of the SJT by measuring the most relevant and likely individual difference correlates of SJT responses. It was hypothesised that conscientiousness, agreeableness and emotional stability would be significantly related to performance on the Video-based Simulation (VBS) SJT, which was the specific SJT used in the current study.

Personality: It is frequently argued that personality influences the judgements that people make about appropriate and inappropriate courses of action, and that SJTs, at least in part, capture such personality constructs (Clevenger et al., 2001). Therefore, we expected that such personality constructs would indirectly impact on managerial performance. Such an expectation is consistent with Weekley and Ployhart's (2005) findings that SJTs partially served to mediate the effects of personality on performance. In addition, McDaniel and Nguyen's (2001) meta-analytical exploration of the relationship between SJTs and the Big Five personality dimensions found that SJTs correlated with agreeableness, conscientiousness, and emotional stability. However, our results clearly show that the SJTs did not correlate with extraversion and openness to experience.

Results indicate a positive relationship between conscientiousness and performance on the VBS (.243, $p < .01$), therefore lending support to hypothesis six (see Table 5.3). This finding is consistent with that of Mullins and Schmitt and Smith and McDaniel (cited in Clevenger et al., 2001) who reported that the SJTs used in their studies, were most strongly correlated with conscientiousness ($r = .26$) and ($r = .32$) respectively.

The correlation between agreeableness and performance on the VBS was low and insignificant, therefore leading to a rejection of hypothesis seven (see Table 5.3). This finding contradicts that of McDaniel and Nguyen's (2001) meta-analysis that found that agreeableness had a significant correlation with SJTs (mean $r = .25$, $k = 12$, $N = 12855$). However, McDaniel and Nguyen also found that when excluding one of the studies from their meta-analysis that reported substantially large correlations between agreeableness and SJTs, the mean correlation was pulled down to (mean $r = .13$, $k = 11$, $N = 8483$).

The correlation between emotional stability and performance on the VBS was low and insignificant, therefore leading to a rejection of hypothesis eight (see Table 5.3). This finding contradicts with that of McDaniel and Nguyen's (2001) meta-analysis that found that emotional stability had the highest correlation with SJTs (mean $r = .31$, $k = 11$, $N = 7482$).

Table 5.3

Correlations between Individual Differences and the VBS

Correlations								
	Verbal evaluation	Numerical interpretation	Basic checking	Extraversion	Openness to Experience	Emotional Stability	Agreeableness	Conscientiousness
VBS	.058	.135	-.030	.104	-.087	.083	.100	.243**

** . Correlation is significant at the 0.01 level (1-tailed)

To summarise, although conscientiousness does not directly impact on managerial performance, the VBS has been found to mediate the effect of conscientiousness on

managerial performance. Thus, conscientiousness influences the judgements that people make about appropriate and inappropriate courses of action when being assessed by the VBS. The effect of agreeableness, on the other hand, was found not to be mediated by the VBS. Furthermore, the fact that emotional stability was found to have no significant relationship with performance of the VBS is consistent with our findings that this personality factor has no correlation with managerial performance.

Cognitive ability: No significant relationship was expected between the specific cognitive abilities and performance on the SJT. It is evident from the literature that, depending on the response instruction, general cognitive ability which represents higher order cognitive functioning, such as abstract thinking, deductive reasoning and general cognitive complexity, correlates highly with SJTs (McDaniel & Nguyen, 2001). However, the cognitive ability tests used in the current study measured lower level basic specific abilities (*verbal evaluation, numerical interpretation and basic checking*) and not general intelligence. The nature of the VBS SJT-measure was designed to measure knowledge of appropriate interpersonal behaviour and problem solving in daily managerial scenarios, so that it does not provide the opportunity to measure verbal, numerical or checking abilities. The expectation that no significant relationship exists between the specific cognitive abilities and performance on the VBS was supported by the results as indicated in Table 5.3.

5.2.1.4 Relationship between Performance on the Situational Judgement Test and Managerial Performance

Evidence to date indicates that SJTs are valid predictors of performance, especially for managerial positions in which interpersonal interactions are important (Motowidlo et al., 1990). SJTs can assess job-related skills that remain untapped by other measures, ranging from skills relating to conflict management, interpersonal communication, problem solving, negotiation, and teamwork facilitation (Chan & Schmitt, 1997; Lievens & Sackett, 2007; McDaniel & Whetzel, 2005; O'Connell et al., 2007; Weekley & Jones, 1999). Performance on the SJT used in this study, the Video-based Simulation (VBS), was therefore expected to directly affect managerial performance.

In addition, the correlation between performance on the VBS and managerial job performance was expected to be significant, due to the high fidelity of the VBS when compared to the nature of tasks required in the target job. Since the VBS is presented in video format, it enables the respondents to see the scenario unfold, including the behaviour of the individuals involved, as if they were making real-life observations. Presentation formats with high fidelity are more reflective of actual job experiences, and are, therefore, likely to be more predictive of on-the-job behaviour. For example, an SJT presented in video format was found to have higher predictive validity than the same SJT presented in written format (Lievens & Sackett, 2006).

Although the VBS did not emerge as a significant predictor of the Managerial Performance Composite ($r = .136, p > .05$), the Performance Ranking measure or the Overall Performance Rating, the SJT did however, correlate moderately with the BOS measure ($r = .234, p < .01$) (see Table 5.3). The significant correlation between the SJT and the BOS measure is in line with earlier research on the validity of SJTs in predicting managerial performance (Motowidlo et al., 1990). A possible explanation could be that the high correlation between the VBS and the BOS scores reflect the congruence between the performance constructs that underlie both measures, by default of design. Both measures, the SJT and the BOS, are strongly grounded in, and developed from the performance dimensions needed for success in the position. The VBS SJT is a high fidelity 'work sample' that was developed by identifying a general task list and typical critical incidents that were likely to be encountered by first-line managers, based on a detailed job and task analysis, as well as by means of quantitative and qualitative integration. Similarly, the BOS measure was developed from a job analysis and competency modelling process to measure the entire construct domain of managerial performance. The VBS items focused on specific performance dimensions of managers namely: *maintenance of discipline; conflict management; negotiation skills; time management; decision making; prioritisation; participative management; motivating of staff; and sensitivity*. The BOS focused on similar competencies required by managers, including: *deciding and initiating action, leading and supervising, persuading and influencing, relating and networking, analysing, achieving personal work goals and objectives, delivering results and meeting customer expectations, following instructions and procedures* (refer to appendix B for behavioural indicators of these competencies). There is a strong conceptual

correspondence between the items of the VBS and the BOS measure, which could explain the strong correlation that was found between them. Hence, both measures are focused on the performance dimensions of the position and, consequently, are operationalised to yield construct relevant and construct inefficient measures of managerial job performance.

Table 5.4

Correlations between Performance on the VBS and Criteria

Correlations				
	OPR	Perf Rank	BOS	MPerf Comp
Video-based Simulation	.126	.081	.234**	.136

** . Correlation is significant at the 0.01 level (1-tailed)

5.2.2 Regression Results

Multivariate analysis techniques were used to investigate a number of the latter hypotheses. Standard multiple regression analysis was used to predict the managerial performance scores from the full set of predictor variables, namely cognitive ability, personality and situational judgment test variables. Following this procedure, hierarchical multiple regression analysis was conducted to examine the extent to which the SJT provided incremental validity in the prediction of managerial performance.

It was hypothesised that managerial job performance can be reliably predicted from a battery consisting of measures of specific cognitive abilities (*verbal evaluation, numerical interpretation* and *basic checking*), personality (*conscientiousness* and *extraversion*) and an SJT. The standard multiple regression model on the Managerial Performance Composite showed that, in combination, the predictors accounted for 5% of the variance in managerial performance, of which agreeableness made the largest significant contribution. Agreeableness therefore explained unique variance in the criterion not explained by the other independent variables.

A key question concerning the usefulness of SJTs concerns their incremental prediction ability over that of other selection tests. This question has relevance for both practical and theoretical reasons. Practically, because tests that yield increased predictive accuracy lead to higher selection utility. Theoretically, since tests that show incremental validity potentially map out areas of the construct domain not tapped by existing measures.

As a measurement method of a multidimensional nature, SJTs can assess various constructs relating to job performance. They can, therefore, be expected to measure other aspects of performance, which cannot be measured by specific tests, such as cognitive ability and personality, therefore adding incrementally to such tests (Weekley & Ployhart, 2005). To examine whether the Video-based Simulation (VBS) SJT-measure provided incremental validity in the prediction of managerial performance relative to the predictability provided jointly by cognitive ability and personality (conscientiousness and extraversion), an hierarchical multiple regression analysis was conducted.

The hierarchical regression model using the Managerial Performance Composite showed that the measures of cognitive ability did not significantly predict managerial performance and that neither the personality nor VBS measures added incremental validity when added to the model. In short, the results indicated that the VBS did not exhibit meaningful or statistically significant incremental prediction over cognitive ability and personality on the Managerial Performance Composite criterion. The hierarchical regression model using the Performance Ranking criterion measure showed that measures of cognitive ability and personality significantly predicted managerial performance, but that the SJT did not. In addition, when adding the SJT to the model, it did not add incremental validity. However, the third hierarchical regression model using the BOS measure of performance showed that, although a battery of cognitive ability, personality and SJT measures did not significantly predict managerial performance, the addition of the SJT to the model added incremental validity.

To summarise these findings, the hierarchical regression model using the BOS demonstrated that the VBS provided incremental validity over cognitive ability and

personality. This finding is consistent with research indicating that SJTs can validly predict job performance incrementally over cognitive ability and personality tests (Chan & Schmitt, 2002; Clevenger et al., 2001; McDaniel et al., 2001; McDaniel et al., 2007; O'Connell et al., 2007; Weekley & Ployhart, 2005). A possible explanation why the VBS added incremental validity only when using the BOS measure, could be that a strong conceptual correspondence exists between the items of the VBS and the BOS measure, as both measures are focused on the performance dimensions of the position and are operationalised according to this framework (as discussed in section 5.2.1.4). In addition, it could suggest that the SJT captures something unique in performance and that it is not merely a substitute for cognitive ability or personality, but rather relates to performance for reasons different from such predictor constructs. Whether this is 'judgement' or some other unmeasured variable such as general knowledge needs to be clarified in future research.

In conclusion, although the multiple regression models did not reach significance, as generally small multiple correlations were observed, even small correlations can have a large practical significance in situations where high base rates and low selection ratios are evident (Murphy & Davidshofer, 2005). Therefore, future use of such SJT measures in similar settings is encouraged.

5.3 LIMITATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The findings in this study must be interpreted in light of several potential limitations, and future research should be designed to address these limitations. First, the SJT used in this study provided incremental validity relative to cognitive ability and personality only when using the Behavioural Observation Scale (BOS) as criterion measure. When using the Managerial Performance Composite measure, the SJT did not provide incremental validity. In this regard, we argue that the Managerial Performance Composite measure provides the best overall operationalisation of the managerial performance construct as each of its constituent individual measures traditionally suffer from methodological weaknesses (Guion, 1998) and by integrating the separate measures into a composite score, we attempted to overcome such weaknesses. This method of integrating different performance measures have, to our knowledge, not

been used in similar studies. However, since inconsistent results were obtained from the different criterion measures, it may be premature to make direct practical implications for personnel selection based on our findings. Further research is needed to determine how the criterion should be measured.

In addition, this study investigated the incremental validity of the SJT over *specific* cognitive abilities. However, to be able to confidently make specific practical recommendations, future research is needed to determine whether the SJT will be able to provide incremental validity over *general* cognitive ability as well and, in greater magnitude than over measures of specific cognitive abilities. To our knowledge, no studies have investigated the interplay between specific *and* general cognitive abilities in as far as they relate to SJT performance. We expect that the SJT will have less incremental validity over general cognitive ability which could mean that no additional predictive validity might be obtained. In this regard, Schmidt and Hunter (1998, p. 262) state that “from the point of view of practical value; the most important property of a personnel assessment method is predictive validity”. It would therefore seem insensible to use a more expensive selection method if no additional predictive validity is gained, when the same predictive validity could be obtained from inexpensive general cognitive ability measures. However, it is important to note that although predictive validity is the most important property of a personnel assessment method, other factors such as fairness (Pulakos & Schmitt, 1996), test-taker reactions (Chan & Schmitt, 2005) should be considered.

We also note that the participants in this concurrent validation study were incumbents. This raises two concerns. First, because incumbents were selected in part according to their cognitive ability and personality profile, restriction in range for these variables might explain why the SJT was not more strongly correlated with them. Consequently, there is likely more restriction of range in the predictors than would be found in applicant samples. The second concern is that criterion measures would also be restricted by the effects of the use of incumbents. In addition, the criterion is likely effected by range restriction because poor performers may have left the organisation. It is therefore recommended that our work be replicated in a predictive validity study using applicant samples.

From inspection of the descriptive statistics of some of the criterion measures (e.g., the BOS measure) it is evident that it is negatively skewed with a smaller standard deviation than those of the other measures. However, this is a common finding where subjective ratings of performance are used to measure criterion variables (Guion, 1998). In this regard, the HR Executive of the participating research organisation expressed the view that, in his opinion, performance levels of the participants varied more in the sample than those observed in the BOS measure scores. Furthermore, it can be speculated that the lower levels of variance in the BOS ratings could have resulted from various factors, such as rater errors (e.g., halo). The fact that this restriction of range on the criterion might not be a true reflection of the performance of the branch managers is another limitation of the study and should be taken into consideration when interpreting the results. On the other hand, however, the ratio of variance to mean scores in the other measures of performance (e.g., ranking) is relatively large when compared to other published norms. Therefore, we are of the opinion that the use of a composite measure used in this study had the effect of overcoming the limitations of each of the specific measures.

Moreover, the correlations between cognitive ability and the SJT, as well as personality and the SJT, reported in this study are lower than those reported in recent meta-analytic studies (e.g., McDaniel et al., 2001). This is likely to be one of the important reasons why the SJT in this particular study provided substantial incremental validity when using the BOS measure and, in this sense, our results are consistent with McDaniel et al.'s (2006) suggestion that the incremental validity of the SJT will vary according to its correlates. More research on the construct validity of the SJT used in the present research would be useful in understanding what it actually measures.

In addition, traditionally, estimating the reliability for SJTs of a multidimensional nature is problematic because the scale and item heterogeneity makes Cronbach's alpha an inappropriate reliability index (Whetzel & McDaniel, 2009). Test-retest reliability has been suggested as better measures for assessing reliability (Hough & Dilchert, 2009; McDaniel & Whetzel, 2005). However, due to the limitations of the current research, measuring the test-retest reliability was not possible in this case. We

agree with McDaniel and Whetzel (2005) that more research is needed on the appropriate methods by means of which to assess the reliability of SJTs.

Finally, since the results of a validation study should be interpreted with caution as validity coefficients can fluctuate from one sample to the next (Kaplan & Saccuzzo, 2001), it is furthermore recommended that the findings of this study should be cross-validated in a future study with other samples before extensive conclusions can be drawn. In this regard, we calculated estimates of expected cross-validity by formula, since these estimates have been shown to be reasonable proxies for empirical cross-validation (Schmitt & Chan, 1998).

5.4 CONCLUSIONS AND IMPLICATIONS

The main objective of this study was to investigate whether a Situational Judgement Test significantly adds to the prediction of managerial performance over other measures typically used for managerial selection in contemporary selection programs, such as measures of cognitive ability and personality.

The results of the present study contribute to the existing literature by demonstrating that different results are obtained when using different types of criterion measures to operationalise the criterion construct. Our study found that when one type of criterion measure (the BOS) was used, the SJT provided significant predictive and incremental validity. However, when the Overall Performance Rating or Performance Ranking measures were used, the SJT provided no significant predictive and incremental validity. This is a gap in the existing body of knowledge that we addressed.

This study further aimed to make a unique contribution by addressing an important gap in the current body of knowledge identified by Weekley and Ployhart (2005) who urged that:

... more research into the relative contribution of cognitive and non-cognitive constructs as determinants of SJT performance is warranted. Although the evidence indicates that cognitive ability is related to nearly all SJTs (McDaniel et al., 2001), it is likely that noncognitive correlates will vary as a function of the item content of the SJT itself (as dictated by

changes in the job). Future research should continue to explore the correlates of SJTs across different jobs and organizations. A particularly important point is that if SJTs are measurement methods, research should not attempt to identify the correlates assessed by SJTs in general, but rather focus on the correlates of SJTs in particular classes of jobs. What is clearly needed is some theoretical basis for understanding how and why personality traits might be differentially related to various SJTs. (p. 100)

The present research study investigated the relationship of SJT scores to an array of managerial performance measures, in conjunction with other predictor measures traditionally used for managerial selection.

In terms of construct validity evidence related to constructs underlying the competencies that drive SJT performance, we found no significant relationship between specific cognitive abilities and performance on the SJT. Further more, conscientiousness was found to be the only personality factor to be related to the SJT. The reason for this finding is not clear, and should be investigated in future research.

Our results suggest that SJTs captures something unique and that they are not merely substitutes for cognitive ability or personality. SJT appear to relate to performance for reasons different from such predictor constructs. Whether this is 'judgement' or some other unmeasured variable such as general knowledge needs to be clarified in future research.

LIST OF REFERENCES

- Babbie, E. R. & Mouton, J. (2001). *The practice of social research*. Cape Town: Oxford.
- Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, 66, 1-6.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.
- Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, 78, 715-722.
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90, 1185-1203.
- Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, 15, 263-272.
- Bartram, D., & Brown, A. (2004). Online testing: Mode of administration and the stability of OPQ 32i scores. *International Journal of Selection and Assessment*, 12, 278-284.
- Bartram, D & Brown, A. (2005). *Five Factor model (Big Five) OPQ32 report*. *OPQ32 Technical Manual Supplement*. Thames Ditton: SHL Group plc.

- Bartram, D., Robertson, L. T., & Callinen, M. (2002). Introduction: A framework for examining organizational effectiveness. In I. T. Robertson, M. Callinen & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 1-11). Chichester: Wiley
- Baydoun, R. B., & Neuman, G. A. (1992). The future of the General Aptitude Test Battery (GATB) for use in public and private testing. *Journal of Business and Psychology*, 7(1), 81-91.
- Beaujouan, Y. (2000). The convergence of the factorial structure of the OPQ (Occupational Personality Questionnaire) towards the Five Factor Model, in seven countries. *European Review of Applied Psychology*, 50, 349-357.
- Becker, T. E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, 13, 225-232.
- Bergman, M. E., Drasgow, F., Donovan, M. A., & Henning, J. B. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14, 223-235.
- Bertua, C., Anderson, N., & Salgado, J. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organisational Psychology*, 78, 387.
- Bledow, R., & Frese, M. (2009). A Situational Judgement Test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 229-258.
- Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, 6, 1-21.
- Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisory experience in supervisor performance. *Journal of Applied Psychology*, 78, 443-449.

- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment* 9, 52-69.
- Borman, W. C., White, L. A., & Dorsey, S. J. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology*, 80, 167-177.
- Boudreau, J. W. (1984). Decision theory contributions to HRM research and practice. *Industrial Relations*, 23, 198-217.
- Bryman, A. & Cramer, D. (2001). *Quantitative data analysis with SPSS Release 10 for Windows*. London: Routledge.
- Campbell, J. R. (1990). Modelling the performance prediction problem in industrial and organisational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organisational psychology* (2nd ed.), (Vol. 1), (pp.687-732). Palo Alto, CA: Consulting Psychologists Press.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). *Managerial behaviour, performance, and effectiveness*. New York: McGraw-Hill.
- Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258-299). Hillsdale, NJ: Erlbaum.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt, W. C. Borman & Associates (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.

- Carroll, J. B. (2005). *The three-stratum theory of cognitive abilities*. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 69-76). (2nd ed.). New York: Guilford Press.
- Cartwright, S., & Cooper, C. L. (2008). (Eds.) *The Oxford handbook of personnel psychology*. Oxford handbooks. Oxford: Oxford University Press.
- Cascio, W.F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall International.
- Chan, D. (2006). Interactive effects of situational judgment effectiveness and proactive personality on work perceptions and work outcomes. *Journal of Applied Psychology, 91*(2), 475-481.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143-159.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15*, 233-254.
- Chan, D., & Schmitt, N. (2005). An agenda for future research on applicants' reactions to selection procedures: A construct-orientated approach. *International Journal of Selection and Assessment, 12*, 9-23.
- Christensen, L. B. (1985). *Experimental methodology* (3rd ed.). Boston: Allyn & Bacon.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt-Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*, 410-417.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioural sciences*. Hillsdale, N.J.: Lawrence Erlbaum.

- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558–577.
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*(6), 653-665.
- Creighton, P. & Scott, N. (2006). An introduction to situational judgement inventories. *Selection and Development Review, 22*, 3-6.
- Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment, 14*, 142-155.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.
- Ferguson, E., Payne, T., & Anderson, N. (1994). Occupational personality assessment: Theory, structure and psychometrics of the OPQ FMX5–Student. *Personality and Individual Differences, 17*, 217–225.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: SAGE Publications.
- Fletcher, C. (1991). Candidates' reactions to assessment centres and their outcomes: A longitudinal study. *Journal of Occupational Psychology, 64*, 117-127.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233-239.
- Furnham, A. (2008). *Personality and intelligence at work: Exploring and explaining individual differences at work*. Howe: Routledge.

- Gatewood, R. D., & Feild, H. S. (1994). *Human resource selection* (3rd ed.). Chicago: Dryden.
- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the Behavioral Sciences*. San Francisco: W.H. Freeman & Co.
- Goffin, R. D., Rothstein, M. G., & Johnston, N. G. (1996). Personality testing and the assessment center: Incremental validity for managerial selection. *Journal of Applied Psychology, 81*, 746-756.
- Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*, 13-23.
- Gravetter, F. J. & Forzano, L. B. (2003). *Research methods for the behavioral sciences*. Belmont, CA: Wadsworth.
- Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). New York: Wiley.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Guion, R. M. and Highhouse, S. (2006). *Essentials of personnel assessment and selection*. London: Lawrence Erlbaum Associates.
- Hanson, M. A., & Borman, W. C. (1995). Construct validation of a measure of supervisory job knowledge. In *The 10th Annual Conference of the Society for Industrial and Organizational Psychology*, Orlando, FL.
- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469-477.

- Hogan, J. & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88, 100-112.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, 11, 209-244.
- Hough, L. M. & Dilchert, S. (2009). Personality: Its measurement and validity for employee selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection*. Mahwah, NJ: Erlbaum.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learnt. *International Journal of Selection and Assessment*, 9, 152-194.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Hunter, J. E., & Schmidt, L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, & Law*, 2(3-4), 447-472.
- Huysamen, G. K. (1980). *Psychological test theory*. Cape Town: Boschendal.
- Ivancevich, J. M., & Matteson, M. T. (1993). *Organizational behaviour and management*. Homewood, IL: Irwin.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport: Praeger.
- Jensen, A. R. & Nyborg, H. (2003). *The scientific study of general intelligence*. Oxford: Elsevier.
- Kaplan, R. M. & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications and issues* (5th ed.). Sydney: Wadsworth.

- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology*. (Vol. 1, pp. 233–265). New York: Oxford University Press.
- Kerlinger, F. N. (1992). *Foundations of behavioral research* (3rd ed.). Orlando, FL: Harcourt Brace.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Routledge.
- Kriek, H. J. (1994). *Handleiding vir die video-gabaseerde-simulasie (VBS) vraelys*. Unpublished manuscript.
- Kurz, R. & Bartram, D. (2002). Competency and individual performance: Modelling the world of work. In I. T. Robertson, M. Callinen & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology*. (pp. 227-257). Chichester: Wiley.
- La Grange, L. & Roodt, G. (2001). Personality and cognitive ability as predictors of the job performance of insurance sales people. *Journal of Industrial Psychology*, 27(3), 35-43.
- Landy, F. J. (1993). The concept of validity and the validation process. *Journal of Business and Psychology*, 7(4), 369-371.
- Landy, F. J. & Conte, J. M. (2007). *Work in the 21st century: An introduction to industrial and organizational psychology* (2nd ed.). Malden, MA: Blackwell.
- LeBreton, J.M. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815-852.
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91, 1181-1188.

- Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology, 92*, 1043-1055.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a videobased situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*, 426-441.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-103.
- Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247-256.
- Matthews, G., & Stanton, N. (1994). Item and scale factor analyses of the Occupational Personality Questionnaire. *Personality and Individual Differences, 16*, 733-743.
- McDaniel, M. A., Hartman, N. S., & Grubb, W. L. (2003). *Situational judgment tests, Knowledge, Behavioral Tendency, and Validity: A Meta-Analysis*. Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63-91.

- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.
- McDaniel, M. A. & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence, 33*, 515-525.
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N., & Grubb, W.L. (2006). Situational judgment tests: Validity and an integrative model. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application*. (pp. 183-204). San Francisco: Jossey-Bass.
- McKenzie, P., Mitchell, P., & Oliver, P. (1995). *Competence and accountability in education*. Aldershot: Arena.
- Morgeson, F., Campion, M., Dipboye, R., Hollenbeck, J., Murphy, K., & Schmitt, N. (2007a). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*, 683-729.
- Morgeson, F., Campion, M., Dipboye, R., Hollenbeck, J., Murphy, K., & Schmitt, N. (2007b). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60*, 1029-1049.
- Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology, 66*, 337-344.
- Motowidlo, S., Dunnette, M. D., & Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640-647.

- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology, 91*, 749-761.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480.
- Murphy, K. R. (2002). Can conflicting perspectives on the role of g in personnel selection be resolved? *Human Performance, 15*(1/2), 173-186.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: principles and applications* (6th ed.). London: Prentice-Hall.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment, 15*, 19-29.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995-1027.
- Ones, D. S., Dilchert, S., Viswesvaran, C., & Salgado, J.F. (2009). Cognitive abilities. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection*. Mahwah, NJ: Erlbaum.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*(4), 679-703.
- Ottz, J. L. (2002). The role of cognitive ability tests in employment selection. *Human Performance, 15*, 161-171.

- Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development, 36*, 9-22.
- Ployhart, R. E., & Harold, C. M. (2004). The applicant attribution-reaction theory (AART): An integrative approach of applicant attributional processing. *International Journal of Selection and Assessment, 12*, 84-98.
- Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment, 11*, 1-16.
- Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*, 241-258.
- Raju, N. S., & Brand, P. A. (2003). Determining the Significance of Correlations Corrected for Unreliability and Range Restriction. *Applied Psychological Measurement, 27*(1), 52-71.
- Ree, M. J., Carretta, T. R., & Steindl, J. R. (2001). Cognitive ability. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1) (pp. 219-232). London: Sage Publications.
- Reeve, C. L. & Hakel, M. D. (2002). Asking the right questions about g. *Human Performance, 15*(1/2), 47-74.
- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology, 85*, 880-887.

- Robertson, I. T., Baron, H., Gibbons, P., MacIver, R., & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology*, 73, 171-180.
- Robertson, I., & Kinder, A. (1993). Personality and job performance: An examination of the criterion related validity of the Occupational Personality Questionnaire. *Journal of Occupational Psychology*, 66, 225-244.
- Rosen, N. A. (1961). How supervise? *Personnel Psychology*, 14, 87-99.
- Rothmann, S. Meiring, D. Van der Walt, H.S. & Barrick, M. (2002). Predicting job performance using personality measures in South Africa. Paper presented at the 17th Annual Conference of the Society for Industrial and Organisational Psychology, April 12-14, Toronto, Canada.
- Salgado, J. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Salgado, J., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance*, 15(1), 75-96.
- Salgado, J., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalisation of GMA and cognitive abilities. *Personnel Psychology*, 56, 573-605.
- Saville, P., Sik, G., Nyfield, G., Hackston, J., & MacIver, R. (1996). A demonstration of the validity of the Occupational Personality Questionnaire (OPQ) in the measurement of job competencies across time and in separate organizations. *Applied Psychology: An International Review*, 45, 243-262.
- Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks: Sage.

- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Schneider, B. (1984). Productivity on the job: Psychological perspective. *Gifts & Decorative Accessories*, 85, 206-211.
- Schneider, B., Smith, D. B., Taylor, S., & Fleenor, J. (1998). Personality and organizations: A test of the homogeneity hypothesis. *Journal of Applied Psychology*, 83, 462-470.
- Shavelson, R. J. (1981). *Statistical reasoning for the behavioral sciences*. Boston: Allyn & Bacon.
- Saville & Holdsworth Ltd (SHL) (1999). Validity study. (Study number V009). Occupational Personality Questionnaire Concept Model. Retrieved from <http://www.research.shl.co.za>
- SHL. (2000a). Reliability study. (Study number R007). Occupational Personality Questionnaire 32 (Ipsative). Retrieved from <http://www.research.shl.co.za>
- SHL. (2000b). Reliability study. (Study number R036). Occupational Personality Questionnaire 32 (Ipsative). Retrieved from <http://www.research.shl.co.za>
- SHL. (2003a). Reliability study. (Study number R045). Critical Reasoning Test Battery. Retrieved from <http://www.research.shl.co.za>
- SHL. (2003b). Reliability study. (Study number R049) Critical Reasoning Test Battery. Retrieved from <http://www.research.shl.co.za>
- SHL. (2003c). Reliability study. (Study number R046). Customer Contact Aptitude Series. Retrieved from <http://www.research.shl.co.za>

- SHL. (2003d). Reliability study. (Study number R048). Customer Contact Aptitude Series. Retrieved from <http://www.research.shl.co.za>
- SHL. (2003e). Reliability study. (Study number R047). Personnel Test Battery. Retrieved from <http://www.research.shl.co.za>
- SHL. (2003f). Reliability study. (Study number R050). Personnel Test Battery. Retrieved from <http://www.research.shl.co.za>
- SHL. (2004). Validity study. (Study number V032). Personnel Test Battery. Retrieved from <http://www.research.shl.co.za>
- SHL. (2006). *Occupational personality questionnaire: Technical Manual*. Esher, UK: Saville & Holdsworth Ltd.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445.
- Society for Industrial and Organisational Psychology of South Africa (SIOPSA). (2005). *Guidelines for the validation and use of assessment procedures for the workplace*. Johannesburg: SIOPSA.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology, 15*, 201-293.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins.
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967-993.

- Tett, R. P., Guterman, H. A., Bleier, A. & Murphy, P. J. (2000). Development and content validation of a 'hyperdimensional' taxonomy of managerial competence. *Human Performance, 13*, 205-251.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.
- Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1994). Meta-analysis of bi-directional relations in personality-job performance research. *Human Performance, 12*, 1-29.
- Theron, C. C. (2007). In search of selection fairness and minimum adverse impact. *PAI Symposium and Workshop on Fair Assessment in Organizations*, University of Stellenbosch, Stellenbosch, 19 October 2007. Retrieved from http://www.pai.org.za/docs/workshops/Theron_PAI_Adverse_impact.pdf
- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology, 61*, 871-925.
- Viswesvaran, C. (1993). *Modeling job performance: Is there a general factor?* (Doctoral dissertation, University of Iowa, 1993).
- Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*, 216-226.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-574.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108-131.

- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology, 50*, 25-49.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology, 52*, 679-700.
- Weekley, J. A. & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance, 18*, 81-104.
- Weekley, J. A., & Ployhart, R. E. (2006). An introduction to Situational Judgement Testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application*. (pp. 1-12). San Francisco: Jossey-Bass.
- Weekley, J. A., Ployhart, R. E., & Harold, C. M. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance, 17*(4), 433-461.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research, *Human Resource Management Review, 19*(3), 188-202.
- Wilk, S. L., Desmarais, L. B., & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology, 80*, 79-85.
- Young, B. S., Arthur, Jr. W., & Finch, J. (2000). Predictors of managerial performance: More than cognitive ability. *Journal of Business and Psychology, 15*, 53-72.

Appendix A

WPS Competency Profile for Branch Managers



wps

> **Competency Profile - SHL UCF™ 20**

Job title: **Branch Manager**

Date: 24/10/2007

INTRODUCTION

The competency profile report identifies the most job relevant competencies based upon an analysis of the tasks, activities and work context that comprise this job. The competency profile report is based upon an analysis of the content and context of this job. It does not take account of the current relationships among employees or the prevailing culture of the organisation.

The SHL competency model used for this profile is called **The SHL Universal Competency Framework™** or **SHL UCF™20**. The SHL UCF™20 is a competency model of 20 *generic competencies* found to contribute to superior performance in all roles and positions in the organisation.

SECTIONS OF THE REPORT

This report consists of the following sections:

1. JOB INFORMATION

This section provides fundamental information about a job. Beyond a person's ability to perform tasks, which make up the job, other issues may impact a person's motivation or interest in a job, such as work location, salary and reporting relationships. These details are specified directly by the WPS user.

2. COMPETENCY PROFILE

The competency profile is a graphical representation of the relative importance level of each competency organised by groups of competencies. The relative importance level of each competency for the job in questions is shown in the bar graph according to the scale definitions. Definitions of the competencies and importance levels follow the competency profile.

3. BACKGROUND REQUIREMENTS

Background requirements refer to the education, training and work experience necessary for a person to be a successful performer. Specific knowledge requirements and other requirements (e.g. willingness to relocate) may be specified by the WPS user to provide a more complete profile.

I

> 01	WPS Version 2	Competency Profile Report (UCF) - Branch Manager	© SHL Group Limited, 2007
------	---------------	---	---------------------------

1. JOB INFORMATION

Job Title:	Branch Manager
Organisation:	Confidential
Department:	Operations
Location:	Western Cape
Salary Range:	R
Upward Reporting Relationships:	Regional Manger
Downward Reporting Relationships:	Consultant; ATM Assistants; CRO; Hit Squads; Admin

MAIN PURPOSE OF THE JOB

To manage a profitable branch through co-ordinating and efficient operational functioning, client service excellence and co-ordination of marketing activities.

JOB OBJECTIVES

To achieve and exceed profit targets within agreed time frames and consistently maintain operational effectiveness standards of the branch within set policies and guidelines.
 To increase market share and achieve marketing objectives through timeous implementation of campaigns, own initiatives, collection, analysis and application of competitor information.
 To sustain Capitec standards of client service, housekeeping, governance and employee standards at all times and under all conditions.
 To increase productivity and maintain effective employee relations through providing induction, coaching, motivating, performance feedback, scheduling and planning of people management in the branch.
 To reduce Capitec's exposure to risk through the effective application of the risk policy.

2. COMPETENCY PROFILE

The competency profile outlines the behaviours that are key to the successful performance of the role as determined during the job analysis session.

	Importance Level			
	Not Relevant	Less Relevant	Desirable	Essential
1. Leading and Deciding				
1.1	Deciding and Initiating Action			
1.2	Leading and Supervising			
2. Supporting and Co-operating				
2.1	Working with People			
2.2	Adhering to Principles and Values			
3. Interacting and Presenting				
3.1	Relating and Networking			
3.2	Persuading and Influencing			
3.3	Presenting and Communicating Information			
4. Analysing and Interpreting				
4.1	Writing and Reporting			
4.2	Applying Expertise and Technology			
4.3	Analysing			
5. Creating and Conceptualising				
5.1	Learning and Researching			
5.2	Creating and Innovating			
5.3	Formulating Strategies and Concepts			
6. Organising and Executing				
6.1	Planning and Organising			
6.2	Delivering Results and Meeting Customer Expectations			
6.3	Following Instructions and Procedures			
7. Adapting and Coping				
7.1	Adapting and Responding to Change			
7.2	Coping with Pressures and Setbacks			
8. Enterprising and Performing				
8.1	Achieving Personal Work Goals and Objectives			
8.2	Entrepreneurial and Commercial Thinking			

The competency profile is presented using the following framework:

Cluster Name	Not Relevant	Less Relevant	Desirable	Essential
Competency 1				
Competency 2				

Diagram labels and connections:

- Competency Factor**: Points to the 'Cluster Name' header.
- Competency Name**: Points to the 'Competency 1' and 'Competency 2' rows.
- Level of Importance**: Points to the 'Not Relevant', 'Less Relevant', 'Desirable', and 'Essential' columns.

Definitions of importance level:

- Not Relevant**.....Basic level of competency expected in all jobs, *not unique* to this job or directly linked to job objectives.
- Less Relevant**.....Slightly more important for this job - relatively more important for meeting at least *some* job objectives.
- Desirable**.....More important for this job - relatively more important for meeting *most* job objectives.
- Essential**.....Much more important for this job - essential for meeting *nearly all* job objectives.

COMPETENCY DEFINITIONS

The definitions of the competencies are provided below:

FACTOR 1: LEADING AND DECIDING		
Takes control and exercises leadership. Initiates action, gives direction and takes responsibility.		
1.1	Deciding and Initiating Action	Takes responsibility for actions, projects and people; takes initiative and works under own direction; initiates and generates activity and introduces changes into work processes; makes quick, clear decisions which may include tough choices or considered risks.
1.2	Leading and Supervising	Provides others with a clear direction; motivates and empowers others; recruits staff of a high calibre; provides staff with development opportunities and coaching; sets appropriate standards of behaviour.

FACTOR 2: SUPPORTING AND CO-OPERATING		
Supports others and shows respect and positive regard for them in social situations. Puts people first, working effectively with individuals and teams, clients and staff. Behaves consistently with clear personal values which complement those of the organisation.		
2.1	Working with People	Shows respect for the views and contributions of other team members; shows empathy; listens, supports and cares for others; consults others and shares information and expertise with them; builds team spirit and reconciles conflict; adapts to the team and fits in well.
2.2	Adhering to Principles and Values	Upholds ethics and values; demonstrates integrity; promotes and defends equal opportunities, builds diverse teams; encourages organisational and individual responsibility towards the community and the environment.

FACTOR 3: INTERACTING AND PRESENTING		
Communicates and networks effectively. Successfully persuades and influences others. Relates to others in a confident and relaxed manner.		
3.1	Relating and Networking	Easily establishes good relationships with customers and staff; relates well to people at all levels; builds wide and effective networks of contacts; uses humour appropriately to bring warmth to relationships with others.
3.2	Persuading and Influencing	Gains clear agreement and commitment from others by persuading, convincing and negotiating; makes effective use of political processes to influence and persuade others; promotes ideas on behalf of oneself or others; makes a strong personal impact on others; takes care to manage one's impression on others.
3.3	Presenting and Communicating Information	Speaks fluently; expresses opinions, information and key points of an argument clearly; makes presentations and undertakes public speaking with skill and confidence; responds quickly to the needs of an audience and to their reactions and feedback; projects credibility.



FACTOR 4: ANALYSING AND INTERPRETING		
Shows evidence of clear analytical thinking. Gets to the heart of complex problems and issues. Applies own expertise effectively. Quickly takes on new technology. Communicates well in writing.		
4.1	Writing and Reporting	Writes convincingly; writes clearly, succinctly and correctly; avoids the unnecessary use of jargon or complicated language; writes in a well-structured and logical way; structures information to meet the needs and understanding of the intended audience.
4.2	Applying Expertise and Technology	Applies specialist and detailed technical expertise; uses technology to achieve work objectives; develops job knowledge and expertise (theoretical and practical) through continual professional development; demonstrates an understanding of different organisational departments and functions.
4.3	Analysing	Analyses numerical data and all other sources of information, to break them into component parts, patterns and relationships; probes for further information or greater understanding of a problem; makes rational judgements from the available information and analysis; demonstrates an understanding of how one issue may be a part of a much larger system.

FACTOR 5: CREATING AND CONCEPTUALISING		
Works well in situations requiring openness to new ideas and experiences. Seeks out learning opportunities. Handles situations and problems with innovation and creativity. Thinks broadly and strategically. Supports and drives organisational change.		
5.1	Learning and Researching	Rapidly learns new tasks and commits information to memory quickly; demonstrates an immediate understanding of newly presented information; gathers comprehensive information to support decision making; encourages an organisational learning approach (i.e. learns from successes and failures and seeks staff and customer feedback).
5.2	Creating and Innovating	Produces new ideas, approaches, or insights; creates innovative products or designs; produces a range of solutions to problems.
5.3	Formulating Strategies and Concepts	Works strategically to realise organisational goals; sets and develops strategies; identifies, develops positive and compelling visions of the organisation's future potential; takes account of a wide range of issues across, and related to, the organisation.

FACTOR 6: ORGANISING AND EXECUTING		
Plans ahead and works in a systematic and organised way. Follows directions and procedures. Focuses on customer satisfaction and delivers a quality service or product to the agreed standards.		
6.1	Planning and Organising	Sets clearly defined objectives; plans activities and projects well in advance and takes account of possible changing circumstances; identifies and organises resources needed to accomplish tasks; manages time effectively; monitors performance against deadlines and milestones.
6.2	Delivering Results and Meeting Customer Expectations	Focuses on customer needs and satisfaction; sets high standards for quality and quantity; monitors and maintains quality and productivity; works in a systematic, methodical and orderly way; consistently achieves project goals.
6.3	Following Instructions and Procedures	Appropriately follows instructions from others without unnecessarily challenging authority; follows procedures and policies; keeps to schedules; arrives punctually for work and meetings; demonstrates commitment to the organisation; complies with legal obligations and safety requirements of the role.

FACTOR 7: ADAPTING AND COPING		
Adapts and responds well to change. Manages pressure effectively and copes well with setbacks.		
7.1	Adapting and Responding to Change	Adapts to changing circumstances; tolerates ambiguity; accepts new ideas and change initiatives; adapts interpersonal style to suit different people or situations; shows an interest in new experiences.
7.2	Coping with Pressures and Setbacks	Maintains a positive outlook at work; works productively in a pressurised environment; keeps emotions under control during difficult situations; handles criticism well and learns from it; balances the demands of a work life and a personal life.

FACTOR 8: ENTERPRISING AND PERFORMING		
Focuses on results and achieving personal work objectives. Works best when work is related closely to results and the impact of personal efforts is obvious. Shows an understanding of business, commerce and finance. Seeks opportunities for self-development and career advancement.		
8.1	Achieving Personal Work Goals and Objectives	Accepts and tackles demanding goals with enthusiasm; works hard and puts in longer hours when it is necessary; seeks progression to roles of increased responsibility and influence; identifies own development needs and makes use of developmental or training opportunities.
8.2	Entrepreneurial and Commercial Thinking	Keeps up to date with competitor information and market trends; identifies business opportunities for the organisation; maintains awareness of developments in the organisational structure and politics; demonstrates financial awareness; controls costs and thinks in terms of profit, loss and added value.

3. BACKGROUND REQUIREMENTS

FORMAL QUALIFICATIONS

Formal qualification or education involves the acquisition of knowledge and skills through learning where subject matter is imparted systematically. Formal qualifications are obtained by studying at formal institutions e.g. universities, technikons, colleges, etc.

Qualifications required	Essential	Desirable
Grade 12	✓	
Business Economics		
		✓
Mathematics and Economics		

JOB RELATED WORK EXPERIENCE

Experience is obtained through opportunities for exposure and practice at work. It includes all working experience that has some bearing on the job and is not restricted to the current organisation. Supervised on-the-job training, internships and Learnerships are incorporated within this category.

Work experience required	Time span	Essential	Desirable
Managing people in a client service front office, retail environment.	4 to 6 years	✓	
		✓	

JOB RELATED KNOWLEDGE

Job-related knowledge is typically gained through formal or informal training programs (these exclude programs through which Formal Qualifications are attained). It includes knowledge of facts, data and information and understanding the rationale behind models, theories and principles.



Job related knowledge	Time span	Essential	Desirable
Basic management principles	2 years	✓	
Labour legislation, Disciplinary Procedures	12 months	✓	
Credit Act	12 months	✓	
Areas policy , IT Policy	12 months	✓	
Capitec Operations Manual	12 months	✓	
Branch Income Statement	3 months	✓	
Client Service Model	1 month	✓	
Product knowledge	3 months	✓	



JOB RELATED SKILLS

Job-related skills are typically gained through formal or informal training programs (these exclude programs through which Formal Qualifications are attained). Skills refer to how to do things. They are demonstrated in the application of techniques and procedures.

Job related skills	Time span	Essential	Desirable
Basic MS Office suite	12 months	✓	
Client Relationship skills	12 months	✓	
Organisation specific operational, technical and people problem solving skills	2 years	✓	
Presentation skills	2 years		✓
Marketing skills	2 years	✓	
Conflict resolution skills	2 years	✓	
People management and leadership	2 years	✓	

REPORT DOCUMENTATION**DESCRIPTION**

The Project Details Section contains the most important details of this WPS Project, including the list of JAQs and any applicable project caveats. This information is important for documentation purposes.

PROJECT DETAILS

Project Name	Branch Manager
Project Id. Number	1
Project Description	Branch Manager
Job Title	Branch Manager
Questionnaire Type	101
Date Created	2007/10/24 03:06:25 PM
Date Modified	2007/10/24 03:06:25 PM
Task Sections Reranked	Reranked (default)
WPS Version	2.3.0.0
Job Analyst	Johan Struwig

ANALYST CONTEXT VARIABLES

Type of Organisation	Banking, finance, insurance, business services and leasing
Size of Organisation	2000 to 2999
Ease of Finding Qualified Staff	Moderate supply of qualified people
Regional Availability	Moderate regional variations in availability
Site Locations	No demographic constraints

JAQs IN THIS PROJECT

JAQ ID	Respondent Name	Respondent Job Title	Modified
1	F Liebenberg; N Ndlovu; F Hendricks; N Adriaans; R	Unnamed Job	2007/10/24

Competency Criticality Scores
CF1=56.918, CF2=61.174, CF3=49.031, CF4=42.293, CF5=71.460, CF6=62.136, CF7=38.421, CF8=37.781, CF9=24.165, CF10=38.338, CF11=33.107, CF12=39.296, CF13=39.641, CF14=50.552, CF15=84.262, CF16=52.818, CF17=47.121, CF18=54.058, CF19=75.619, CF20=44.511 CFES=56.500, CFDE=50.000, CFLR=42.293

REPORT COMMENTARY

This report was generated using the Work Profiling System Version 2 module of the SHL® Human Resource Management System. The report is computer-generated from the results of one or more job analysis questionnaires answered by subject matter experts and substantially reflect the answers provided by them. Due regard of this must be taken in the interpretation of this data.

This report has been generated electronically - the user of the software can make amendments and additions to the text herein - SHL cannot accept any liability for the consequences of the use of this report and cannot guarantee that the contents are the unchanged output of the computer system.

SHL Group Limited and its associated companies cannot guarantee that the contents of this report are the unchanged output of the computer system. We can accept no liability for the consequences of the use of this report and this excludes liability of every kind (including negligence) for its contents. This report is confidential and should not be published in any way. SHL Group Limited cannot accept any liability if it is.

The SHL Universal Competency Framework and SHL Online Competency Profiler are protected by copyright (c) 2007, SHL Group Limited and are trademarks of SHL Group Limited. All rights reserved

© SHL Group Limited, 2007. All Rights Reserved

Internet> <http://www.shl.com>

Email> info@shl.com

SHL is a trademark of SHL Group Limited. Registered in the UK and in other countries.

> 010	WPS Version 2	Competency Profile Report (UCF) - Branch Manager	© SHL Group Limited, 2007
-------	---------------	---	---------------------------

Appendix B

Questionnaires

VBS QUESTIONNAIRE BIOGRAPHICAL INFORMATION

For office use

3

Date	
Name and Surname	
ID Number	
Branch	
For how long have you been the BM of your branch?	
For how long have you been a BM in this company?	
Were you a manager prior to working for this company? If yes, for how long?	
Age	
Sex	
Marital Status	
Race	

Regional Manager	
Region	
Ops Manager	

Informed Consent

I understand herewith that these ratings are confidential, only for research purposes and that my identity will be treated with confidentiality.

Signature

Branch Manager Behavioural Questionnaire (BOS)

Instructions

1. Complete the background information on yourself as the rater as well as the Branch Manager that is being assessed. Please note that this information is necessary in statistical analysis.
2. Rate each Branch Manager in your region on a separate questionnaire.
3. Circle responses with the pencil provided.
4. Respond to all 16 items.
5. Rate each statement independently.
6. Avoid using one overall impression.
7. Avoid rating all Branch Managers either high or low.
8. Avoid the middle response (3) as much as possible. Try to use the full five-point scale.
9. In order to get the most value out of the questionnaire please respond honestly and frankly. Confidentiality is guaranteed.

Branch Manager Behavioural Questionnaire Statements

Statement 1:	Makes effective decisions
Behavioural Indicators	Makes well-informed, objective decisions; Uses principles, values and sound business sense to make decisions; Takes time to fully understand the implications before making a decision; Takes responsibility for decisions.
Statement 2:	Influences others to reach objectives
Behavioural Indicators	Effectively gets team to reach targets and objectives; Involves others in a process or decision to ensure their support; Argues effectively to win influence for desired course of action; Presents compelling case to team, superiors, and key decision-makers.
Statement 3:	Initiates action
Behavioural Indicators	Takes action that is consistent with available facts, constraints, and probable consequences; Starts projects without being prompted.

Statement 4:	Demonstrates leadership skills
Behavioural Indicators	Positively influences the attitudes and actions of others; Personally models appropriate values, behaviours, and work practices; Views him/herself as part of the team, not above it; Adapts leadership styles to various situations; Provides team members with a clear sense of direction; People tend to follow him/her Coaches staff in the performance of their tasks.
Statement 5:	Supervises performance and ensures corrective action is taken
Behavioural Indicators	Organizes resources and steers others toward successful task accomplishment; Ensures that team members have the necessary information to operate effectively
Statement 6:	Develops and maintains positive relationships with others
Behavioural Indicators	Gets along well with team and other colleagues; Shows awareness of the personal circumstances of the team; Sets the example for dealing with and relating to clients; Shows appreciation for healthy working relationships between team members.
Statement 7:	Building and maintaining networks of contacts to further the organization's goals
Behavioural Indicators	Builds and maintains wide and effective networks of interest, e.g. employers and competition, to manage branch more effectively; Keeps support areas like Marketing and HR informed of priorities, needs, and issues, in pursuit of responsive service.
Statement 8:	Persuades others to gain agreement and commitment
Behavioural Indicators	Identifies and presents information that will have an impact on others; Presents several different arguments in support of a decision; Communicates viewpoints, arguments and objectives in a manner that leads to commitment; Convinces others of his/her opinion.
Statement 9:	Evaluates and interprets information/reports in order to understand issues
Behavioural Indicators	Makes rational judgements from the available information and analysis; Demonstrates an understanding of how one issue may be a part of a much larger system.
Statement 10:	Achieves personal work goals
Behavioural Indicators	Sets and achieves realistic personal goals and work plans that are consistent with the business needs and strategies; Accepts and tackles demanding goals with enthusiasm; Works hard and puts in longer hours when it is necessary.

Statement 11:	Reaches objectives and targets
Behavioural Indicators	Consistently reaches branch targets Demonstrates a clear understanding of the relationship between team objectives and broad organizational goals and objectives; Supports employee efforts to achieve job and organizational goals (e.g., by providing resources, removing obstacles, acting as a buffer).
Statement 12:	Meeting clients' needs in a manner that provides satisfaction for the client
Behavioural Indicators	Understands clients' requirements/expectations and applies this knowledge to effectively deliver valued solutions; Quickly and effectively solves client problems; Is accessible and provides prompt, attentive service; Looks for external trends that are likely to shape the wants and needs of clients in the near future.
Statement 13:	Follows instructions from superiors in order to reach company goals and objectives
Behavioural Indicators	Appropriately follows instructions from superiors without unnecessarily challenging authority; Keeps to schedules; Effectively performs the job in a timely and efficient manner; Demonstrates commitment to the organization.
Statement 14:	Follows policies and procedures
Behavioural Indicators	Understands and appropriately applies procedures, requirements, and policies; Completes all reports and documents according to procedures and standards; Monitors the team's work to ensure it aligns with formal policies and procedures.
Statement 15:	Analyzes information/data in order to understand issues
Behavioural Indicators	Implements a structured process of collecting and analyzing information; Analyses numerical data and other sources of information to determine trends and make decisions; Conduct comparisons between alternative scenarios on dashboard with multiple variables (e.g., product demand, price, competition, margins).
Statement 16:	Delivers results and reaches targets
Behavioural Indicators	Consistently achieves goals and targets; Addresses problems and issues promptly; Monitors and maintains quality and productivity.

For office use

Branch Manager Behavioural Questionnaire Rating Sheet

1

Date	
Regional Manager Name and Surname	
Region	
Ops Manager	

Branch Manager	
Branch	

Informed Consent

I understand herewith that these ratings are confidential, only for research purposes and that my identity will be treated with confidentiality.

Signature

Overall Performance Rating

1) Have you had sufficient opportunity to observe this individual's job-related behaviour?

YES

NO

If '**NO**' please provide a reason, eg this BM only started working in a branch in my region 2 months ago

2) Rate this individual's overall job performance as a Branch Manager, from 1 (top performer) to 10 (worst performer).

1 2 3 4 5 6 7 8 9 10

Branch Manager	
Branch	

Statement 1	Almost never					Almost always
	1	2	3	4		5
Statement 2	Almost never					Almost always
	1	2	3	4		5
Statement 3	Almost never					Almost always
	1	2	3	4		5
Statement 4	Almost never					Almost always
	1	2	3	4		5
Statement 5	Almost never					Almost always
	1	2	3	4		5
Statement 6	Almost never					Almost always
	1	2	3	4		5
Statement 7	Almost never					Almost always
	1	2	3	4		5
Statement 8	Almost never					Almost always
	1	2	3	4		5
Statement 9	Almost never					Almost always
	1	2	3	4		5
Statement 10	Almost never					Almost always
	1	2	3	4		5
Statement 11	Almost never					Almost always
	1	2	3	4		5
Statement 12	Almost never					Almost always
	1	2	3	4		5
Statement 13	Almost never					Almost always
	1	2	3	4		5
Statement 14	Almost never					Almost always
	1	2	3	4		5
Statement 15	Almost never					Almost always
	1	2	3	4		5
Statement 16	Almost never					Almost always
	1	2	3	4		5

Please ensure that you have completed all the items.

 YES
 NO

Branch Manager Ranking Sheet

For office use

2

Date	
Regional Manager Name and Surname	
Region	
Since when are you a RM in your Region?	
Since when are you a RM in this company?	
Ops Manager	

Instructions: Rank the Branch Managers in your region from 1 – top performer to 10 – worst performer (The top performer you would typically promote first).

Rank	Branch Manager
1 (Top Performer)	
2	
3	
4	
5	
6	
7	
8	
9	
10 (Poor Performer)	

Appendix C

**Fritz & MacKinnon's (2007) necessary sample sizes for six of the most common
and the most recommended tests of mediation for various combinations of
parameters**

TABLE 3
Empirical Estimates of Sample Sizes Needed for .8 Power

Test	Condition															
	SS	SH	SM	SL	HS	HH	HM	HL	MS	MH	MM	ML	LS	LH	LM	LL
BK ($\tau' = 0$)	20,886	6,323	3,039	1,561	6,070	1,830	883	445	2,682	820	397	204	1,184	364	175	92
BK ($\tau' = .14$)	562	445	427	414	444	224	179	153	425	178	118	88	411	147	84	53
BK ($\tau' = .39$)	531	403	402	403	405	158	124	119	405	125	75	59	405	122	60	38
BK ($\tau' = .59$)	530	404	402	403	406	158	124	120	405	125	74	58	404	122	59	36
Joint significance	530	402	403	403	407	159	124	120	405	125	74	58	405	122	59	36
Sobel	667	450	422	412	450	196	144	127	421	145	90	66	410	129	67	42
PROCLIN	539	402	401	402	402	161	125	120	404	124	74	57	404	121	58	35
Percentile bootstrap	558	412	406	398	414	162	126	122	404	124	78	59	401	123	59	36
Bias corrected bootstrap	462	377	400	385	368	148	115	118	391	116	71	53	396	115	54	34

Note. All sample sizes have been rounded up to the next whole number. In the condition labels, the first letter refers to the size of the α path, and the second letter refers to the size of the β path; S = 0.14, H = 0.26, M = 0.39, and L = 0.59 (e.g., condition SM is the condition with $\alpha = 0.14$ and $\beta = 0.39$). All results, except for those for Baron and Kenny's (1986) test (BK), have been collapsed across τ' conditions.

Appendix D

Ethics Form

ETHICS COMMITTEE APPLICATION FORM
UNIVERSITY OF STELLENBOSCH
SUBCOMMITTEE A

2006

Application to the University of Stellenbosch SUBCOMMITTEE A
for clearance of new/revised research projects

This application must be typed or written in capitals

Name: Prof/Dr/Mr/Ms:	MS SIGLIND FERTIG
Position/Professional Status:	STUDENT
Affiliation: Research Programme/Institution:	STELLENBOSCH UNIVERSITY

Telephone and extension no.	0722125369
Email address:	sigifertig@capitecbank.co.za

Title of research project: (*Do not use abbreviations*)

The incremental validity of a situational judgement test relative to personality and cognitive ability tests to predict managerial performance.

Where will the research be carried out?

The research will be carried out within Capitec Bank at the various regional conferences in South Africa. All the branch managers in each region gather once a year where they are informed of the strategic plan for the year ahead and to discuss various other operational issues and challenges. The data for this research project will be gathered at each of these conferences, therefore including all branch managers.

All the following sections must be completed (Please tick all relevant boxes where applicable)

1. FUNDING OF THE RESEARCH: How will the research be funded?

The researcher (S. Fertig) is employed by Capitec Bank, which is funding the research. No ethical risk is foreseen given the fact that the researcher's employer will fund the research.

2. PURPOSE OF THE RESEARCH:

I/O psychologists are continuously searching for ways to predict job performance more accurately, especially when using different measuring instruments together, to ensure that the utility of such instruments are considered.

Furthermore, South African legislation require employers to be able to proof that measuring instruments used for selection purposes are valid and reliable.

The purpose of the present study is to examine whether situational judgement tests (SJT) add incrementally to the prediction of managerial performance over other measures used for managerial selection, such as measures of cognitive ability and personality?

3. AIMS AND OBJECTIVES OF THE RESEARCH: *(Please list objectives)*

This research attempts to further the understanding of the incremental validity of SJTs over personality and cognitive ability tests as a predictor of managerial performance.

4. SUMMARY OF THE RESEARCH *(give a brief outline of the research plan – not more than 200 words)*

This study will be conducted in a South African retail bank. A correlative ex post facto design will be used to reach the objective of this study. The specific research design is the concurrent validity design. The sample consists of branch managers who are geographically spread throughout the country. As the researcher seeks to obtain as many respondents as possible, a non-probability convenience sampling method was chosen for this study. Existing selection data generated by the company and SHL will be analysed for this study. The data for the criterion measures were collected by the bank's HR staff and a supervising industrial psychologist during the annual regional Branch Manager Conferences. Once a year, all branch managers gather at their regional Branch Manager Conference in a classroom setting to discuss various matters. Arrangements were made with the Head of Operations to allow some time at each region's Conference for all branch managers to complete the SJT and for their supervisors to complete a behavioural questionnaire (BOS) on each of the branch managers reporting to them. The SJT data were thus collected

5. NATURE AND REQUIREMENTS OF THE RESEARCH

5.1 How should the research be characterised (*Please tick ALL appropriate boxes*)

5.1.1 Personal and social information collected directly from participants/subjects	
5.1.2 Participants/subjects to undergo physical examination	
5.1.3 Participants/subjects to undergo psychometric testing	X
5.1.4 Identifiable information to be collected about people from available records	X
5.1.5 Anonymous information to be collected from available records	
5.1.6 Literature, documents or archival material to be collected on individuals/groups	

5.2 Participant/Subject Information Sheet attached? (*for written and verbal consent*)

YES	X
NO	

5.3 Informed Consent form attached? (*for written consent*)

YES	X
NO	

5.3.1 If informed consent is not necessary, please state why:

Since this project was initiated by Capitec Bank's HR department, their own informed consent form was completed. Refer to attached document for more information.

NB: If a questionnaire, interview schedule or observation schedule/framework for ethnographic study will be used in the research, it must be attached. The application cannot be considered if these documents are not included.

5.4 Will you be using any of the above mentioned measurement instruments in the research?

YES	X
NO	

6 PARTICIPANTS/SUBJECTS IN THE STUDY

6.1 If humans are being studied, state where they are selected:

The population of Capitec branch managers, who are geographically spread throughout the country, is selected.

6.2 Please mark the appropriate boxes:

Participants/subjects will:	YES	NO
be asked to volunteer	X	
be selected		

6.2.1 State how the participants/subjects will be selected, and/or who will be asked to volunteer:

6.3 Are the participants/subjects subordinate to the person doing the recruiting?

YES	
NO	X

6.3.1 If yes, justify the selection of subordinate subjects:

6.4 Will control participants/subjects be used?

YES	
NO	X

6.4.1 If yes, explain how they will be selected:

6.5 What records, if any, will be used, and how will they be selected?

Each branch manager completed the cognitive ability tests as well as the personality questionnaire as part of their recruitment process. This data will be retrieved from the Recruitment department's database. The letter from Capitec Bank in which formal permission was granted to use this data for this research is attached.

6.6 What is the age range of the participants/subjects in the study?

21-45

6.6.1 Was assent for guardians/consent for participants/subjects obtained?

YES	
NO	X

If YES, please attach the appropriate forms.

6.6.2 If NO, please state why: N/A

6.7 Will participation or non-participation disadvantage the participants/subjects in any way?

YES	
NO	X

6.7.1 If yes, explain in what way:

No, it will not disadvantage participants or non-participants since the performance appraisal data will not be used for any decision-making.

6.8 Will the research benefit the participants/subjects in any direct way?

YES	
NO	X

6.8.1 If yes, please explain in what way:

7. PROCEDURES

7.1 Mark research procedure(s) that will be used:

Literature	X
Documentary	
Personal records	X
Interviews	
Survey	X
Participant observation	
Other (please specify)	X
Video situational judgment test.	

7.2 How will the data be stored?

The captured data is stored on a personal computer that is located in a secure office on the premises of a company (SHL) that conducts psychometric testing.

7.3 If an interview form/schedule; questionnaire or observation schedule/framework will be used, is it attached?

7.4 Risks of the procedure(s): Participants/subjects will/may suffer:

No risk	X
Discomfort	
Pain	
Possible complications	
Persecution	
Stigmatisation	
Negative labeling	
Other (please specify)	

7.4.1 If you have checked any of the above except "no risk", please provide details:

8. RESEARCH PERIOD**(a) When will the research commence:**

The research makes use of archival data was collected in 2007-2008 by SHL and Capitec Bank.

(b) Over what approximate time period will the research be conducted:

One year.

9. GENERAL**9.1 Has permission of relevant authority/ies been obtained?**

YES	X
NO	

9.1.1 If yes, state name/s of authority/ies:

Leon Venter Chief Human Resources, Capitec Bank. Refer to formal letter in attached document.

9.2 Confidentiality: How will confidentiality be maintained to ensure that participants/subjects/patients/controls are not identifiable to persons not involved in the research:

No individual scores will be made available to the company.

9.3 Results: To whom will results be made available, and how will the findings be reported to the research participants?

Capitec Bank management will be informed of the results of the research. No individual information will be disclosed. Ms Tina Joubert of SHL, who will assist with the data analysis, will be aware of the results of the research. The results will not be reported to research participants since the measures are continuously used in personnel selection.

In addition, the results will be reported in the thesis and might be published in a journal article in an accredited journal.

9.4 There will be financial costs to:

participant/subject	
institution	X
Other (please specify)	

9.4.1 Explain any box marked YES:

Plane tickets, car rentals, accommodation for the in-house researchers to travel to the regional conferences to conduct the research. This research falls within ambit of HR normal duties.

9.5 Research proposal/protocol attached:

YES	X
NO	

9.6 Any other information which may be of value to the Committee should be provided here:

The agreement between SHL and the researcher as well as Stellenbosch University is attached to this document.


Date: 12 May 09**Applicant's signature****Ms Siglind Fertig****Who will supervise the project?****Name:****Mr Francois de Kock****Programme/Institution/Department:****Department of Industrial Psychology\
Stellenbosch University****Date:****Signature: _____****Director/Head/Research Coordinator of Department/Institute in which study is conducted:****Name:****Date:****Signature: _____**