Assessment Centres within the South African Workplace: Relationship between a Situational Judgement Test and Ability and Personality Measures.

Emma Caruth

991236625

A treatise submitted to the Faculty of Humanities, Development and Social Sciences
(University of KwaZulu Natal, Howard College Campus)
in partial fulfilment of the requirements for the degree of

Master of Social Science

in

Industrial Psychology

October 2011

DECLARATION

Submitted in partial fulfilment of the requirements for the degree of Master of Social Science (Industrial Psychology), in the Graduate Programme in

Humanities, University of KwaZulu-Natal, South Africa

I declare that this dissertation is my own unaided work. All citations, references and borrowed ideas have been duly acknowledged. I confirm that an external editor was not used. It is being submitted for the degree of Master of Social Science in the Faculty of Humanities, Development and Social Sciences, University of KwaZulu-Natal, South Africa. None of the present work has been submitted previously for any degree or examination in any other University.

Emma Caruth

Student name and surname

29 October 2011

Date

ACKNOWLEDGEMENTS

Completing one"s research is a destination from a journey never taken alone! I would like to thank the following people for their unique contribution and assistance along the way:

- ❖ My supervisor, Shaida Bobat, for her wisdom and guidance;
- ❖ Tina Joubert from SHL for her statistical advice and professional support;
- ❖ My mentor, Pam Matthias, who provided me with all the impetus required to complete this journey;
- ❖ My most beautiful daughter, Brenna Kean, who shows me every day exactly what life should be about;
- ❖ My father, Geoff Caruth, who provided me with the inspiration to pursue an academic niche;
- ❖ My grandpa, Robin Dean, who always made me feel like "the bee"s knees";
- ❖ My mother, Linda Grobler, for her enduring love and support;
- ❖ My brother and sister-in-law, Andrew and Angela Caruth, for providing relief from my studious diligence;
- ❖ The love of my life, Lee Thompson, who never tires of putting the pieces back together!

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT:

Assessment Centres within the South African Workplace: Relationship between a Situational Judgement Test and Ability and Personality Measures.

This research investigates the relationship of a specific situational judgement test, Scenarios, to that of ability and personality measures through a discriminant validation study. The Scenarios Test was developed in order to assess the construct of Managerial Judgement within the work context. Despite the fact that this instrument has demonstrated reliability and validity evidence, this has been primarily obtained within a European context. Therefore, this study will contribute towards the growing body of evidence supporting the psychometric integrity of the use of Scenarios within a South African context. Moreover, this research aims to determine whether or not a situational judgement test of Managerial Judgement provides information relating to constructs which are not sufficiently measured by ability tests and personality questionnaires. The results indicate that there is significant inter-correlation between the scales of Scenarios, but not between Scenarios and the other instruments used (cognitive ability tests and a trait personality measure). This may support the hypothesis that SJTs could measure a broader underlying construct, such as tacit knowledge, as well as the fact that they may provide unique information relating to candidates" potential job performance which is not provided by tests of ability and personality.

CHAPTER 1: AN INTRODUCTION TO THE RESEARCH TOPIC

1.1 Introduction

This section will explain the background and purpose of this research as well as the key issues which this research will address. Lastly, the value that this research hopes to add to the area of occupational assessment will be discussed.

1.2 Background to and Purpose of the Study

This research attempts to investigate the construct validity of a personnel selection instrument known as Scenarios. Scenarios is a Situational Judgement Test (SJT) developed by Howard and Choi (2004), and distributed by SHL[1]. SHL not only continuously invests in research for maintaining Scenarios" psychometric integrity, but also consults with the product and supervises its use by other suitably trained individuals.

SJTs in general are being used increasingly for various purposes. For example, they are often used as a sift tool for external applicants, as part of selection in an internal promotion process, or as a developmental tool by providing useful feedback on how the candidates" responses compare to the ideal response. However, there have been conflicting views as to whether or not SJTs add more value than cognitive ability tests and/or personality measures. Moreover, the lack of compelling theory surrounding SJTs is perhaps its greatest limitation to its widespread and accepted use (Weekley & Ployhart, 2006). A proper theoretical framework and lack of relevant theory has been purported as the downfall of SJTs. It is for these reasons that this research aims to investigate the construct validity of a specific SJT, Scenarios.

Industrial/Organisational (I/O) Psychologists and Human Resource (HR) professionals use numerous types of assessments to make personnel selection and development decisions in organisations (Conner, 2007). These instruments vary in terms of their content, cost of administration, validity and reliability (amongst other differentiating factors). The major

---

[1] SHL is an international organisation which was founded in 1970 by Professor Peter Saville and Roger Holdsworth. Presently, the organisation operates in over forty different countries, and has psychometric and competency potential instruments available in approximately thirty different languages. To date, SHL is the largest psychometric assessment producer, publisher and distributor globally. Its South African division began in 1994 by Professor Hennie Kriek. There are branches in the cities of Pretoria (where the Head Office is located), Cape Town and Durban. There are also numerous associates throughout South Africa who are able to conduct testing, and provide client support and supervision.

goal of psychometric assessment users is to select assessment instruments and procedures that reliably and cost effectively select individuals from a pool of applicants who will succeed in the position that is being recruited for (Conner, 2007). In other words, they are attempting to maximise predictive validity while minimising costs and adverse impact.

Selection concerns fair discrimination between different candidates based upon the inherent requirements of the job. According to the South African Employment Equity Act (1998), fair discrimination is based on only two grounds; the abilities, characteristics and competencies inherently required by the job, and whether the selection is in accordance with set affirmative action policies and procedures (SHL, 2008b). Should discrimination not be based on these grounds, or the method of assessment is unreliable, not valid, unfairly applied or biased towards certain groups, then the discrimination is categorised as unfair (SHL, 2008b). Unfair discrimination is legally reprehensible.

The Employment Equity Act (1998) provides clear guidance and requirements for legally acceptable labour practices. Chapter two provides a clear indication as to the legally accepted grounds for discrimination. These are that the discrimination must be based on:
1. the inherent requirements of the job, and
2. the affirmative action plans / interventions put in place by the employer.

Chapter two, section eight of the Employment Equity Act (1998) lists specific requirements for the use of any psychological instruments within the workplace. This states that any type of psychological test or similar instrument is only allowed if it is proven to be:
1. scientifically valid and reliable
2. applied fairly, and
3. not biased towards any employee or group.

Only tests which demonstrate sound psychometric properties should be used. The relevant information for judging their psychometric soundness should be available in the test manual. This should include:
- Specification of the skill that the test measures;
- Description of the groups for which the test is appropriate;

- Details of the development process;

- Reliability information;

- Evidence of test validity;

- An outline of the steps taken to avoid ethnic, gender and religious bias;

- A selection of relevant norm groups containing representative proportions of ethnic, religious minorities as well as male and female populations.

Employment Equity, or equal opportunity legislation, has been in operation for a number of years within South Africa. Although many cases have been brought against employers for unfair selection processes, very few of these have been brought against the occupational tests used within the selection process. This reflects the professional and scientific approach taken by most psychometric developers. A psychological assessment development and distributor company should continuously provide evidence for the integrity of the instruments being provided to society. This is due to the nature of the reasons for the use of these instruments. Important decisions regarding recruitment, selection, succession planning, and development are based on the outcome of these assessment tools. Furthermore, providing clients with context specific validation evidence for the use of these tools is invaluable. Clients can therefore trust that decisions based on the outcome of these assessments are well informed, objective and accurate.

The use of psychometric assessments within the work context is also governed by the Health Professions Council of South Africa (HPCSA). The HPCSA provides a regulatory function in terms of the classification and use of psychometric tests. According to the Health Professions Act, Act 56 of 1974, only registered psychologists are permitted to perform psychological acts which, in relation to evaluation, testing and assessment, are defined in Section 37 (2) (a), (b), (c), (d) and (e) as being:

"(a) the evaluation of behaviour or mental processes or personality adjustments or adjustments of individuals or groups of persons, through the interpretation of tests for the determination of intellectual abilities, aptitudes, interests, personality make-up or personality functioning, and the diagnosis of personality and emotional functions and mental functioning deficiencies according to a recognised scientific system for the classification of mental deficiencies;

(b) the use of any method or practice aimed at aiding persons or groups of persons in the adjustment of personality, emotional or behavioural problems or at the promotion of positive personality change, growth and development, and the identification and evaluation of personality dynamics and personality functioning according to scientific methods;

(c) the evaluation of emotional, behavioural and cognitive processes or adjustment of personality of individuals or groups of persons by the usage and interpretation of questionnaires, tests, projections or other techniques or any apparatus, whether of South African origin or imported, for the determination of intellectual abilities, aptitude, personality make-up, personality functioning, psychophysiological functioning or psychopathology;

(d) the exercising of control over prescribed questionnaires or tests or prescribed techniques, apparatus or instruments for the determination of intellectual abilities, aptitude, personality make-up, personality functioning, psychophysiological functioning or psychopathology;

(e) the development of and control over the development of questionnaires, tests, techniques, apparatus or instruments for the determination of intellectual abilities, aptitude, personality make-up, personality functioning, psychophysiological functioning or psychopathology."

Due to the strict governance and regulation of the use of tests within a South African context, it becomes imperative for test developers to research the psychological integrity of their instruments. This research will also help to adapt Scenarios to a South African context by assisting in generating information specific to the South African population. Scenarios has not as yet been adapted for use within a South African occupational context.

1.3 Key Issues to be Addressed by the Study

The key research questions are formulated as follows:

➢ What is the correlation between the SJT of Managerial Judgement and the two cognitive ability tests?

➢ What is the correlation between the SJT of Managerial Judgement and the 32 scales of the personality questionnaire?

Further subsidiary research investigations include the following:

➢ Determine the candidates scores on the assessment instruments used;

➢ Calculate the inter-correlations between the two cognitive ability tests;

- ➢ Calculate the 32 scale inter-correlations of the OPQ32i;
- ➢ Calculate the 10 scales inter-correlations for the Scenarios Test;
- ➢ Determine the reliability of the instruments used for the sample population.

The two hypotheses of this research can thus formulated as follows:

H1: There will be a low correlation between the Situational Judgement Test of Managerial Judgement and the cognitive ability tests.

H2: There will be a low correlation between the Situational Judgement Test of Managerial Judgement and the scales measured by the personality questionnaire.

1.4   Strengths of the Research

Research has shown that well-constructed psychometric tests predict job performance better than most other selection methods (Bartram, 2004; Schmidt & Hunter, 1998). Tests give objective information about a candidate and have been shown in general to lead to better and fairer employment decisions (Schmidt & Hunter, 1998). This has been the finding of numerous research studies, as well as meta-analytic studies of the usefulness of psychometric instruments in predicting job performance (otherwise known as validity generalisation) (Schmidt & Hunter, 1998). Rafilson (1991) states that if validity generalisation evidence is limited, then local criterion-related evidence of validity is required to justify the use of a test. On the other hand, if validity generalisation evidence is extensive, then situation-specific evidence of validity may not be required. Concerning occupational psychometric assessment, due to the fact that there exists extensive research on the topic of the validity of certain instruments in predicting job performance, one can extend these findings to other valid and reliable instruments. Therefore, validation studies of this type provide evidence for the usefulness of instruments in predicting competency potential, and ultimately job performance.

The relationship of a test with other tests contributes to an understanding of the test‟s construct validity. This is generally conducted through an investigation into convergent validity (the degree to which the constructs are similar to other constructs measured by instruments which they theoretically should be similar to) and discriminant validity (the degree to which the constructs diverge from other constructs measured by instruments which they theoretically should be different to). The relationship between different tests

can therefore give useful insight into what tests are (and are not) measuring. SJT construct validity efforts to date have involved little more than examining correlates of SJTs (Ployhart & Erhart, 2001) and have failed to offer much in the way of meaningful insight into convergent and discriminant validity. To this end, this research investigates whether or not a specific SJT, Scenarios, contributes something extra to constructs measured by cognitive ability tests and trait personality questionnaires. Most importantly, one cannot just assume that psychometric assessments have equivalent psychometric properties to others which have demonstrated psychometric integrity. This has to be demonstrated and proved for each psychometric instrument. Thus, one of the strengths of this research is congruent with the usefulness of most validation studies. It will provide objective evidence of the usefulness of the instrument in predicting job performance.

Test utility is an important consideration for both test developers and test users alike. Test utility refers to the conditions under which tests are beneficial within an occupational context, i.e. how small the costs of testing are against the potential gains, with the obvious intention of organisations wishing to increase the utility of tests which they use. Conditions under which test utility is most favourable are when the validity of the test is high, there is a current low success rate with the current techniques being used, and there is a high degree of selectivity (SHL, 2008b). Through conducting validation studies of this nature, evidence towards the test's utility is provided.

1.5   Conclusion

Due to the emphasis placed on the results of tests, and their potential misuse, legislative efforts attempt to guide their use and purpose within the context of selection decisions. These policies need to be taken in to account with any research conducted into the arena of occupational tests.

Furthermore, due to the investment made by many organisations in to the use of psychometric assessment instruments, it is imperative that providers of these tools maintain their psychometric soundness, thereby increasing the utility that these instruments can provide.

CHAPTER 2: LITERATURE REVIEW

2.1   Introduction

A review of the literature includes a consideration of individual difference theory within the context of personnel selection. Managerial performance as a predictor construct will be elucidated. Furthermore, a description of SJTs in general will be provided, outlining their development, scoring methods, and formats. The benefits of using SJTs will be delineated, with a particular reference to the occupational context. A description of what they measure as well as their relationship to other psychometric assessment instruments will be provided, referring to two main schools of thought regarding what value SJTs provide to the assessment process. The specific SJT, Scenarios, employed by this research will be explicated, including the stages of its development and its specific correlations to other forms of assessment. The two cognitive ability tests will be explained, as well the personality questionnaire. Evidence concerning their reliability and validity will be presented.

2.2   Personnel Selection

Personnel selection research has had a rich history within the field of Industrial/Organisational (I/O) Psychology, devoted to investigating the prediction of job performance (Chamorro-Premuzic & Furnham, 2010). Accurate prediction of job performance is said to be critical to the success of organisations. The main aim in personnel selection is to "maximise predictive efficiency by identifying and selecting individuals with the highest job relevant ability" (Fertig, 2009:8). To this end, a central premise of the 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, motivation, knowledge, interests and values. The individual difference model has a number of fundamental assumptions that guide personnel selection (Fertig, 2009). Firstly, adults have a variety of attributes that are relatively stable over time (Fertig, 2009). Secondly, people differ with respect to those attributes (Fertig, 2009). Thirdly, differences among people on those attributes remain relatively constant, even after training or accrued professional experience (Fertig, 2009). Fourthly, different jobs require different attributes (Fertig, 2009). The fundamental principles of personnel selection depend upon maximising

the fit between the attributes of the candidate and the requirements of the job. Lastly, those attributes can be measured by various assessment instruments (Fertig, 2009).

Research has revealed that scientific measurements of individual differences in cognitive ability and personality are valid predictors of job performance. In particular, individual differences in managerial performance are an important determinant of both job and organisational success. In recruitment and selection for managerial positions there is a particular emphasis on appointing the right person to the right job. This is because managerial positions are most often critical to the operation of the business and an organisation's values and culture are often determined by the leaders of the organisation. Selecting a poor performer into the job can have serious cost and staff morale implications for the business. However, it is not always an easy exercise in selecting the right potential job incumbent. This is due to the fact that job performance is multidimensional in nature, and managerial performance in particular involves a broad domain of requisite performance behaviours.

2.3 The Criterion Constructs of Job Performance and Managerial Performance

In order to understand the importance of managerial performance in personnel selection, it is crucial to explicate the theory around job performance in personnel selection theory. A popular conceptualisation of job performance is Campbell's (1990) model which defines three key features of understanding and investigating job performance. First, job performance can only be defined in observable behaviours. Second, job performance only includes those behaviours which are relevant to the organisation's goals. Third, job performance is conceptualised as a multidimensional construct. Campbell (1990) provides a taxonomy of eight performance components:

1. *Job-specific task proficiency:* An individual's capacity to perform the core substantive or technical tasks central to the job.

2. *Non-job-specific task proficiency:* An individual's capacity to perform tasks or execute performance behaviours that are not specific to their particular jobs.

3. *Written and oral communication task proficiency*: An individual's proficiency in writing and speaking, independent of the correctness of the subject matter.

4. *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.

5. *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.

6. *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.

7. *Supervision/leadership:* Proficiency in influencing the performance of subordinates through face-to-face interpersonal interaction and influence.

8. *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.

Campbell (1990) further 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). The eight performance components are a function of the three performance determinants. The performance determinants are affected by individual difference variables, such as ability, personality, motivation and values.

As can be seen above, the performance domain of most jobs is complex, and may even be more so managerial jobs (Bartram, Robertson & Callinen, 2002; Kurz & Bartram, 2002). Fleishman, Mumford, Zaccaro, Levin, Korotkin and Hein (1991) examined leader behaviour classifications to develop a taxonomy which describes behaviour requirements for effective leadership. Four broad super-ordinate dimensions of behaviour were identified. Firstly, information search and structuring which involves the leader‟s acquisition of information, organising that information, and providing guidance or feedback to subordinates based upon that information (Fleishman et al, 1991). Secondly, information use and problem solving, which involves applying information in the pursuit of solving problems through identifying needs and requirements of the group, communicating that information, and coordinating unit efforts (Fleishman et al, 1991). Thirdly, managing personnel resources, including activities such as obtaining, allocating, developing and motivating (Fleishman et al, 1991). Fourthly, managing material resources,

including obtaining, allocating, utilising and monitoring the use of those resources (Fleishman et al, 1991).

Tett, Guterman, Bleier & Murphy (2000) developed a comprehensive managerial performance taxonomy that combines twelve models of managerial performance developed by other theorists. Tett et al (2000:43) identified 53 dimensions of managerial performance grouped into nine areas:

1. *Traditional Functions*

    i.   Problem awareness: perceives situations that may require action to promote organisational success.

    ii. Decision making: uses good judgement in resolving problems.

    iii. Directing: clearly specifies to subordinates what needs to be done.

    iv. Decision delegation: assigns true decision-making authority to qualified subordinates.

    v.   Short-term planning: prepares the steps needed to complete tasks before action is taken.

    vi. Strategic planning: develops long-term plans to keep the organisation aligned with future demands.

    vii. Coordinating: organises the activities of subordinates and the allocation of resources.

    viii. Goal setting: identifies organisational work unit objectives and the methods for achieving them.

    ix. Monitoring: compares current work unit progress to predetermined standards, objectives, and deadlines.

    x.   Motivating by authority: influences subordinates directly using rewards and/or punishments.

    xi. Motivating by persuasion: persuades others to achieve excellence for its own sake.

    xii. Team building: identifies and integrates distinct subordinate roles in a spirit of collaboration.

    xiii. Productivity: accomplishes goals set by self or others.

2. *Task Orientation*
    i.   Initiative: takes preliminary steps to do what needs to be done without direction.
    ii. Task focus: stays on task despite complexity and/or ambiguity.
    iii. Urgency: responds quickly to pressing organisational demands.
    iv. Decisiveness: does not hesitate in making tough decisions.

3. *Person Orientation*
    i.   Compassion: shows genuine concern for the welfare of others.
    ii.   Cooperation: seeks to accomplish work goals through collaboration with others.
    iii. Sociability: initiates and energetically maintains friendly interactions with others inside and outside of work.
    iv.   Politeness: demonstrates proper manners when dealing with others.
    v.   Political Astuteness: takes advantage of political relationships and the distribution of power in pursuing goals.
    vi.   Assertiveness: states views confidently, directly, and forcefully.
    vii. Seeking Input: actively pursues others" contributions to work related discussion.
    viii. Customer Focus: seeks to maintain or enhance customer satisfaction.

4. *Dependability*
    i.   Orderliness: maintains a high degree of organisation in his or her physical work environment.
    ii. Rule Orientation: realises the importance of organisational rules and policies, and willingly follows them.
    iii. Personal Responsibility: accepts responsibility for own actions, decisions, and directions to subordinates.
    iv. Trustworthiness: maintains confidentiality in dealing with sensitive information about the company, its customers, and/or its workers.
    v.   Timeliness: shows appreciation for and abides by routine job-related time limits.
    vi. Professionalism: demonstrates the standards of his or her career or occupational group.

vii. Loyalty: shares the company's goals and values.

5. *Open mindedness*

    i. Tolerance: values judgements different from his or her own.

    ii. Adaptability: readily adapts to new situations and immediate work demands.

    iii. Creative Thinking: fosters creative thinking within the organisation or work unit.

    iv. Cultural Appreciation: appreciates diversity in cultural experiences and/or beliefs.

6. *Emotional Control*

    i. Resilience: maintains a positive attitude in response to failure.

    ii. Stress Management: deals effectively with feelings of job-related stress and their causes.

7. *Communication*

    i. Listening skills: actively attends to what others are saying.

    ii. Oral Communication: expresses thoughts verbally in a clear, pleasant, and straightforward manner.

    iii. Public Presentation: is effective and comfortable in presenting material to groups of people.

    iv. Written Communication: expresses self clearly and succinctly in writing.

8. *Developing Self and Others*

    i. Developmental Goal Setting: collaborates with individual subordinates to establish work objectives for their career advancement.

    ii. Performance Assessment: evaluates individual co-workers' performance with respect to their personal developmental objectives.

    iii. Developmental Feedback: gives regular, specific, and timely feedback to subordinates in relation to personal goals.

    iv. Job Enrichment: gives employees learning opportunities to expand job-related expertise.

    v. Self Development: seeks out and engages in self-improvement opportunities.

9. *Occupational Acumen and Concerns*
   i. Technical Proficiency: knows what it takes to do the job.
   ii. Organisational Awareness: knows how the organisation works as a whole and in terms of individual work units.
   iii. Quality Concern: works to meet or exceed existing organisational quotas or standards.
   iv. Financial Concern: understands the importance of generating and saving money for the organisation.
   v. Safety Concern: emphasises accident prevention at the workplace.

SHL has a range of competency frameworks designed to describe various occupations and used to model managerial performance (Bartram et al, 2002; Kurz & Bartram, 2002). The eight factor competency framework was developed, which divides 112 specific competencies into eight categories (Bartram & Brown, 2005:5):

1. Leading and Deciding: takes control and exercises leadership; initiates action, gives direction and takes responsibility.

2. Supporting and Cooperating: 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 that complement those of the organisation.

3. Interacting and Presenting: communicates and networks effectively; successfully persuades and influences others; relates to others in a confident and relaxed manner.

4. Analysing and Interpreting: shows evidence of clear analytical thinking; gets to the heart of complex problems and issues; applies own expertise effectively; quickly learns new technology; communicates well in writing.

5. Creating and Conceptualising: open 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.

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.

7. Adapting and Coping: adapts and responds well to change; manages pressure effectively and copes with setbacks.

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.

The above taxonomies reveal overlaps across a number of areas. Whilst some are more comprehensive than others, what is important to note is the fact that not all competencies are as important as one another to all managerial professions. Various industries and organisations demand a different set of competencies relevant to their respective context. Therefore, a routine practice within personnel selection is the undertaking of a formal job analysis process in order to determine the relevance of various competencies to the specific job requirements. All taxonomies of managerial performance are based on the assumption that the competencies operate as predictors of managerial performance which can be measured and assessed by a variety of methods. Researchers and practitioners are investigating how to maximise the prediction of managerial performance by developing new predictor constructs and measures (Fertig, 2009). Of significance, SJTs have become an increasingly popular selection tool, and evidence to date suggests that they are valid predictors of requisite performance, in particular, for managerial positions. An important consideration concerns how SJT performance relates to managerial performance in comparison to measures of cognitive ability and personality. Although this research does not seek to investigate the incremental validity of the SJT in predicting managerial performance over and above that of cognitive ability and personality measures, it does seek to examine the relationship between the different instruments. As the central premise of this research maintains that variations in job performance is a result of individual differences, the individual difference measures used in this study will be further explained, viz. situational judgement, cognitive ability and personality.

2.4   Description of Situational Judgement Tests

SJTs are a type of psychological test which present the test taker with scenarios focused on the tasks and the behaviours inherent within the target job. These scenarios represent realistic but hypothetical vignettes, and describe a work related situation in which the test taker is required to identify an appropriate course of action. The item responses are presented in multiple choice format, where several possible options are provided. The task of the test taker is to choose the best response which reflects their intended actions and decisions based upon their analysis of the situation.

A review of situational judgement literature reveals that there are two basic stances regarding what SJTs are and what they measure (Conner, 2007). One stance, led by Sternberg and colleagues (Sternberg & Hedlund, 2002; Sternberg, Wagner, Williams & Horvath, 1995; Wagner & Sternberg, 1986) contends that SJTs measure a construct in and of itself which is independent of any other individual difference variables. This construct has been labelled as "tacit knowledge" (Sternberg & Wagner, 1986). On the other hand many scholars, such as McDaniel and colleagues (Chan & Schmitt, 2002; Clevenger, Pereira, Wiechmann, Schmitt & Harvey, 2001; McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001; McDaniel & Nguyen, 2001, Smith & McDaniel, 1998), argue SJTs are an assessment method similar to selection interviews and assessment centres which measure various job related constructs. Therefore, they do not specifically measure a single construct, but rather a range of individual difference variables those of which may be critical for job success.

According to the former camp, tacit knowledge is a subcomponent of the more general term of practical intelligence. Tacit knowledge is defined as knowledge that reflects the ability to learn from experience, and consists of the accumulation of habits and culture which cannot easily be taught or shared. With tacit knowledge, people are not often aware of the knowledge that they possess or how it can be valuable to others. Contrary to explicit knowledge, which can be easily communicated and understood, tacit knowledge cannot be easily codified or articulated (Hedlund, Antonakis & Sternberg, 2002). Similarly, others have argued that because SJT items are multi-dimensional it is likely that situational, or tacit, knowledge mediates the relationship between knowledge, skills, ability and job

performance (Chan & Schmitt, 1997; Motowildo, Hanson & Crafts, 1997). Unfortunately, there is limited empirical evidence to substantiate these claims (Trippe, 2002)

.

According to the opponents of SJTs measuring a form of tacit knowledge, McDaniel et al (2001) found a relatively strong relationship between SJTs and cognitive ability. They also found that this relationship was dependent on how the test had been developed. For example, if a job analysis had been used to develop an SJT, the relationship with cognitive ability was higher compared to when a job analysis had not been conducted. One reason for this finding relates to the establishment of content validity which a formal job analysis provides. In this way an SJT can be designed in accordance with the content requirements of the job. Interestingly, these SJTs were also found to be better predictors of job performance. This finding would therefore suggest that the more robust SJTs are more likely to be contaminated by cognitive ability. Thus, rather than testing aspects of job knowledge, they are actually testing logical and deductive reasoning. In other words, SJTs are a method for measuring a variety of constructs (Bergman et al, 2006). Many selection tests are construct centred as they are designed to measure a specific construct (e.g. cognitive ability, integrity, emotional intelligence). They are therefore labelled as such (e.g. cognitive ability test, integrity test, emotional intelligence questionnaire). In contrast, predictors such as interviews, work samples and SJTs are developed with a job-centred focus in that they are designed to reflect aspects of the work itself as opposed to a specific construct (Christian et al, 2010). These selection instruments are therefore described in method-based, as opposed to construct-based, terms. Due to SJTs being constructed by collecting critical incidents required during job performance, they generally tap into a variety of predictor constructs simultaneously (Christian et al, 2010). This conceptualisation of SJTs is as hollow measurement methods rather than indicators of a unique construct or combination of constructs. If this is accurate then one should be able to measure any number of constructs using SJT items as the modality (Trippe, 2002). As with the argument that SJTs measure tacit knowledge, there is limited empirical evidence supporting the conceptualisation of SJTs as a hollow measurement method (Trippe, 2002).

Christian, Edwards and Bradley (2010) contend that the difficulty with defining what SJTs measure is due to the limited attention test developers and researchers give to the constructs measured, and instead tend to report results based on overall, or composite, SJT scores. This is supported by Bledow and Frese (2009) where they contend that although

Page 16 of 81

SJTs have been found to be valid predictors of performance, they have rarely been used to measure particular constructs. Moreover, the development of SJTs is largely atheoretical (Bledow & Frese, 2009; Motowildo et al, 2001; Motowildo, Dunnette & Carter, 1990), contributing towards its nebulous quality. The majority of SJTs are not designed with an emphasis on construct validity because they are developed for applied purposes and are thus concerned with criterion related validity (Trippe, 2002). It is argued that despite the rising interest in and use of SJTs, the fundamental properties of these instruments remain poorly understood, leading to limitations in the utility of SJTs. Furthermore, the variability in the SJT literature can also be attributed to the fact that so many people are using different SJTs measuring different constructs. This is reflected in McDaniel et al"s (2001) meta-analytic study where an enormous variability in research results could be underpinned by the fact that the SJTs used measured a wide variety of content despite being the same modality of measurement. The fact that the modality is the same does not mean that a latent construct is measured which has reliable correlations with one another, or with constructs such as cognitive ability and personality. Therefore, it becomes very difficult to cross-compare SJT construct validity results. It is assumed that those who do not specify constructs or dimensions believe SJTs measure a unidimensional judgement construct. The problem here is that a measure of communication and leadership skills is going to differ dramatically from one that is constructed to measure motivation or simply "judgement". It is not necessary, or possible, for two SJTs to correlate with one another if they are measuring different constructs. It is unclear why researchers attempt to make universal statements about SJTs when each instrument is designed very differently (Trippe, 2002).

In Christian et al"s (2010) meta-analysis of the criterion related validities of SJTs it was found that a substantial number (33%) of the SJTs did not report the constructs measured, did not provide enough information to determine the constructs measured, or provided only a composite score which collapsed across multiple constructs. Several possible explanations for why researchers neglect to specify the constructs measured by SJTs are proposed by Christian et al (2010). Firstly, SJTs are commonly considered methods, and purport to measure a single construct such as situational judgement, tacit knowledge or practical intelligence; second, it may be difficult to create SJTs that can be scored at the construct level; lastly, I/O psychology has only recently begun developing and implementing a construct-orientated paradigm for selection research (Christian et al,

2010). This last point is supported by Schmitt and Chan (cited in Christian et al, 2010) where they state that I/O psychology is more apt to discuss the validity of methods rather than the validity of measurement constructs.

What emerges as important is a consideration of how the SJT is developed. In other words, an SJT can be developed with particular consideration of measuring a construct. To this end, theory and literature surrounding the construct under investigation will be utilised in the design of the items of the test. In contrast, an SJT may be developed to reflect the content of the requirements of a particular job, i.e. through the use of critical incidents and various other work sample examples. In the former example, validation of the instrument would centre around divergent and convergent validation studies with other assessment instruments. In the latter example, validation would require concurrent or predictive validation studies with the criterion (i.e. job performance).

Bess (2001) and Strahan, Fogarty and Machin (2005) outline further difficulties in defining constructs measured by SJTs in the context of the workplace. They assert that job performance is a multi-dimensional construct, including both task performance and contextual performance (Bess, 2001, Strahan et al, 2005). Task performance refers to the duties contained within a job description, whereas contextual performance refers to knowledge of facts, principles and procedures which relate to situations that require helping and cooperating with others. Contextual activities differ from task activities in four ways. First, contextual activities do not support the technical core of work, but rather the social and psychological environment surrounding the technical core (Bess, 2001). Second, contextual activities are much the same in all jobs and organisations (Bess, 2001). Third, variations in contextual performance are influenced by volition and predisposition, whereas variations in task performance are mediated by proficiency (Bess, 2001). Fourth, contextual activities are not role-prescribed and thus not formally recognised as part of the job (Bess, 2001).

Borman and Motowildo (1997) present five categories that operationalise contextual performance. These include volunteering to carry out tasks and activities that are not formally within the job description, persisting with extra enthusiasm or effort, helping and cooperating with others, following organisational rules and procedures even when

personally inconvenient, and endorsing, supporting and defending organisational objectives (Borman & Motowildo, 1997).

It is further claimed that both task and contextual performance have separate antecedents (Bess, 2001). Specifically, the differential predictors of personality and cognitive ability tend to better predict contextual and task performance respectively (Bess, 2001; Strahan et al, 2005). Bess (2001) suggests that situational judgement is a multi-dimensional evaluation methodology that assesses task and/or contextual job knowledge, and therefore, in any given SJT there may be items that better tap contextual knowledge while other items may better tap task knowledge. This may explain the inconsistencies found in the SJT literature. Bess (2001) developed an SJT specifically for the study where items were constructed based upon tapping into either task or contextual knowledge. Bess" (2001) study showed incremental validity for personality but that cognitive ability did not contribute to the significance of the model. It is claimed that the task-contextual knowledge distinction may be a key to solving the mystery of the constructs underlying situational judgement (Bess, 2001).

A study conducted by Bergman et al (2006) investigated an SJT (Leadership Skills Assessment) in predicting leadership and overall job performance above and beyond the effects of personality and cognitive ability. The Leadership Skills Assessment (LSA) is a 21-item computerised multi-media assessment of leadership skills. The Wonderlic Personnel Test (WPT) was used to measure cognitive ability. The WPT contains 50 items involving word comparison, disarranging sentences, sentence parallelism, number comparison, number series, analysis of geometric figures, and word problems requiring mathematical or logical solution. Respondents are given 12 minutes to complete as many items as they can. Substantial evidence exists for the reliability and construct validity of WPT scores. Test-retest reliabilities for the WPT range from 0.82 to 0.94. Personality was measured by the Sixteen Personality Factor Questionnaire (16PF). The 16 PF contains 185 items mapped onto 16 primary factors and/or five global factors. The five global factors and their concomitant reliabilities are: extraversion (0.90), Anxiety (0.90), Self-Control (0.88), Independence (0.81) and Toughmindedness (0.84). Scale correlations between the LSA and the WPT and 16PF did not exceed 0.32, suggesting "that the LSA is not overly redundant with cognitive ability or personality" (Bergman et al, 2006:229).

In general, research attempting to examine the construct validity of SJTs has been largely inconsistent. Constructs that are measured across all SJTs have thus far proven elusive in that no consistent relationships have emerged with other variables such as personality and cognitive ability (Chan & Schmitt, 1997). General cognitive ability has been among the most enduring constructs examined, but results have been mixed, with some researchers finding significant correlations (e.g. r=0.42 and 0.46) (Weekly & Jones, 1999; McDaniel et al, 2001) and others finding non-significant relationships (e.g. r= -0.04 to 0.09) (Mullins & Schmitt, 1998; Motowildo et al, 1990; Ployhart, 1999). McDaniel et al's (2001) meta-analysis suggested a corrected mean correlation of 0.46 with cognitive ability, but with enough non-trivial variability (i.e. credibility values ranging from 0.17 to 0.75) to cast doubt on any conclusions based on this data.

There has been a paucity of research in the personality domain. Some researchers have found relationships between SJTs and Five Factor Model (FFM) personality variables such as Conscientiousness (r=0.26 to 0.32), Extraversion (r=0.19 to 0.20), Emotional Stability (r=0.16 to 0.19) and Agreeableness (r=0.22 to 0.24) (Bess, 2001; Mullins & Schmitt, 1998; Smith & McDaniel, 1998).

An SJT designed to measure team role knowledge was developed and validated by Mumford, Van Iddekinge, Morgeson and Campion (2008). The estimated internal consistency reliability of the Team Role Test (TRT) was given as 0.83. As part of their investigation into criterion-related validity of the TRT, the instrument was validated against measures cognitive ability and personality. The cognitive ability test used was the Wonderlic Personnel Test (WPT). Internal consistency reliability of the WPT in the study was estimated at 0.89. Personality was measured using the 50-item International Personality Item Pool (IPIP) instrument (measuring the Big Five personality factors). Each IPIP scale comprises 10 items. For each item respondents indicate the extent to which the statement is a *very inaccurate* (1) to a *very accurate* (5) description of who they are. The study obtained coefficient alphas for the five personality scales ranging from 0.73 for openness to experience to 0.91 for extraversion. Significantly, the TRT demonstrated low correlations with cognitive ability and personality. The results of the study led the researchers to believe that SJTs are best thought of as a method, rather than a specific construct, which can be used to assess a wide range of predictor constructs (Mumford et al,

2008). It was also concluded that SJTs capture declarative and procedural knowledge relevant to the target construct/s (Mumford et al, 2008).

In Fertig's (2009) study the incremental validity of SJTs over personality and cognitive ability tests as a predictor of managerial performance was explored. The results of the study indicated that the SJT did not exhibit meaningful or statistically significant incremental prediction over cognitive ability and personality to predict the composite managerial performance measure.

Results such as these may lead organisations to question the benefits of using SJTs over and above the use of cognitive ability tests and personality measures. However, Weekley and Jones (1999) found that SJTs account for all the predictive validity of cognitive ability tests but the reverse was not true. One explanation for this finding is that SJTs do not measure one unique construct, as ability tests have been designed to measure, but are multidimensional tools that encompass several job relevant skills and abilities, a sentiment shared by the opponents of tacit knowledge. To this end, SJTs tap into other aspects of job knowledge other than that just required by cognitive ability alone. This further supports the finding by Weekley and Jones (1999) of a strong relationship between SJTs and job experience, such as experience in the industry, longest tenure with one employer and exposure to multiple employers. Stemler and Sternberg (2006) support this research and add that SJTs appear to capture some unique variance that is not captured by personality or by general ability.

2.3.1 The Format and Scoring Strategies of SJTs

SJTs can appear in various types of formats. These include paper-and-pencil, video based, as well as online versions. Concerning the equivalence of forms, research findings have been mixed. Chan and Schmitt (1997) found that video based SJTs were significantly more face valid and revealed significantly lower levels of adverse impact than their paper-and-pencil counterparts. Further, Konradt, Hertel and Joder (2003) developed and concurrently validated a web-based SJT which also yielded strong face validity and acceptance.

Unlike cognitive ability tests, SJT items do not have objectively correct answers. All of the responses are possible, however, some may be better than others as opposed to the right and wrong course of action (Bergman, Drasgow, Donovan, Henning & Juraska, 2006).

There are a number of ways to determine the best course of action. Three main categories of scoring methods are described in the literature: (1) empirical, (2) theoretical, and (3) expert-based (Bergman et al, 2006). Empirical approaches score items according to their relationships with a criterion measure (McDaniel & Nguyen, 2001). The process includes choosing a criterion, developing decision rules, weighting items and cross-validating the responses (Bergman et al, 2006). Theoretical scoring requires that items and options be constructed to reflect theory, or that theory is used to identify the best and worst options (Bergman et al, 2006). Lastly, expert-based scoring requires a group of Subject Matter Experts (SMEs) to determine the relative effectiveness of the responses, and assign ordinal weights to each response option. Generally, the scoring of these tests can be accomplished by providing one point for each correct response, or the various choices for each scenario can be weighted. For the latter, the weighted responses scoring approach can provide more points for choosing the responses determined as the best choices, and less points (or a reduction in points) for choosing the worst choice (Joiner, 2008).

2.5   Benefits of Using Situational Judgement Tests

The popularity and the use of SJTs has increased over the past 10 to 15 years since its introduction as a low-fidelity simulation by Motowildo, Dunette and Carter (1990) for the selection of entry level managers. A meta-analytic study conducted by McDaniel, Morgeson, Finnegan, Champion and Braverman (2001) revealed the criterion related validity of SJTs to be comparable with some of the best predictors of job performance. Moreover, the following advantages have been cited regarding their increasing use and popularity: firstly, they show less adverse impact against minorities when compared to that demonstrated by cognitive ability tests (Creighton & Scott, 2006). In other words, differences in mean scores between racial subgroups are typically smaller than those reported for cognitive ability tests (Motowildo & Tippens, 1993). This finding was supported by Nguyen and McDaniel (2003) where it was found that racial differences were a function of the cognitive saturation of the test, i.e. when respondents to the SJT were asked to indicate their knowledge, the SJT was more cognitively saturated and displayed greater sub-group differences than when respondents were asked to provide their behavioural tendencies. Secondly, they can be used to directly measure job relevant behaviours as the simulation exercises presented by the vignettes can be designed to reflect the context and content of the job. Therefore, they are more acceptable and engaging to candidates when compared to cognitive ability tests as they are based on real incidents and

so are appeared to have more face validity (Motowildo, Dunette & Carter, 1990). Thirdly, they can be administered in bulk, either via paper-and-pencil or online. It can be relatively easy and inexpensive to develop, administer and score SJTs (Clevenger, Pereira, Wiechmann, Schmitt & Harvey, 2001; Gessner & Klimoski, 2006). Fourthly, it is unlikely that practice will enhance candidate performance as the answers cannot be arrived at logically, i.e. a response to a situation may be appropriate in one organisation but inappropriate in another (www.humanassets.co.uk/pagecontent.asp). Lastly, they can tap a variety of constructs, ranging from problem solving and decision making, to interpersonal skills (www.humanassets.co.uk/pagecontent.asp). Albeit, it is the last point which has given rise to a gap in our knowledge concerning these instruments, i.e., what exactly do these tests measure, and do scores on these tests correlate with those found on personality and ability instruments? Considering the fact that organisations invest much resources in administering additional instruments over and above an SJT, does the SJT provide extra and appropriate information that is not simply obtained from cognitive tests and personality questionnaires? The answer to this question provides significant direction for the future of assessment processes in the context of selection decisions.

2.6 Description of the Specific Situational Judgment Test Employed by this Research: Scenarios

Scenarios is a low-fidelity simulation that is specifically designed to measure Managerial Judgement. In reference to its low-fidelity nature, the bandwidth fidelity of an instrument concerns the level of spread of various jobs that an instrument can encompass versus the accuracy and job specificity it can cover (Ones & Viswesvaran, 1996). In the context of Scenarios, this instrument provides broad and generic situations as opposed to job or industry specific ones. However, these situations encompass details which would be relevant to a broad range of decisions which would need to be made in most managerial settings. This allows the instrument to be used in a wide variety of jobs across a large number of industries. However, it must be considered that as the fidelity of a simulation decreases, so too can the element of situational realism.

Managerial Judgement (and other judgements in general) occurs in response to situations where unique decisions have to be made when no clear precedent or procedure has been established for action (Howard & Choi, 2004). Managerial Judgement can be assessed in a variety of ways, such as through the use of in-tray exercises and a situational interview.

Scenarios employs a situational method whereby individuals are presented hypothetical situations which are designed to rely on Managerial Judgement in order to solve them. In essence, Scenarios is described as a test of Managerial Judgement.

Individuals are obviously not born with Managerial Judgement. Its nature is such that very little, if any, can be learnt at school or even during higher education (Howard & Choi, 2004). Most of the skill develops either on the job or through direct training within the corporate environment (Howard & Choi, 2004). According to most theoretical approaches, including that proposed by SHL, Managerial Judgement is dependent upon three variables. These include applying one‟s reasoning ability, formal management training experiences, and tacit skill concerning management. Thus, an individual‟s reasoning skills play some role in shaping Managerial Judgement (Howard & Choi, 2004). However, one should not exaggerate the role which cognitive ability plays. For example, even a cognitively astute manager can make wrong decisions in corporate situations (Howard & Choi, 2004). One also needs to consider the influence of training and management programmes designed to increase correct managerial decision making in a variety of contexts. However, these two ingredients (cognitive ability and training and development) cannot lead to a full development of the skill. The third contributor has been described as an everyday awareness and understanding of how organisations operate (Howard & Choi, 2004). This alludes to the importance of pragmatism in decision making. It is this awareness which has been described as tacit knowledge, practical knowledge, or the "tricks of the trade" (Howard & Choi, 2004:5).

Some crucial characteristics concerning tacit knowledge include the following: firstly, it is difficult to verbalise, i.e. explaining what one has learnt is often difficult to elucidate (Howard & Choi, 2004). Secondly, some aspects of life other than direct experiences within managerial roles may assist indirectly with the development of tacit skill (Howard & Choi, 2004). Evidence for this is shown when mature entrants to organisations show quite well developed Managerial Judgement without having been in managerial roles (Howard & Choi, 2004). Thirdly, some tacit skill can be taught directly to trainees such as analysing an expert‟s skill in Managerial Judgement in different situations, and breaking this down into component parts (Howard & Choi, 2004). Lastly, individuals can develop tacit skill through acquiring knowledge informally on a day-to-day basis (Howard & Choi, 2004). This process can be accelerated through mentorship and coaching programmes, and

deliberately reflecting on events (such as critical incidents and setbacks) (Howard & Choi, 2004).

The Scenarios Test presents a succession of managerial scenarios and associated responses (Howard & Choi, 2004). Individuals are asked to evaluate each of the possible responses for their effectiveness in addressing the presented situation. These evaluations are then scored against a set of ideal answers to provide an assessment of Managerial Judgement (Howard & Choi, 2004). When individuals are asked to respond to hypothetical situations there is always a concern that their response does not reflect how they would act in a real life incident. However, given the realism of the Scenarios vignettes, it is reasonable to assume that the answers given reflect what people feel would be the correct action to take, and therefore how they would behave in a real life setting (Howard & Choi, 2004). A further concern has been that individuals may "fake good" when answering. However, this criticism does not make sense as knowledge of what the right thing to do is exactly what the test attempts to measure (Howard & Choi, 2004). Some have also mentioned that using tests such as Scenarios for managerial recruitment and selection may lead to managerial cloning within the organisation, stifling diversity and innovation (Howard & Choi, 2004). However, Managerial Judgement is only one of the critical skills required by managers. There are a myriad of other attributes which contribute to successful management, such as persuasiveness, interpersonal skills, commercial awareness, among others.

2.5.1   The Stages of Development of Scenarios

The development of the Scenarios Test was guided by the following criteria: to measure Managerial Judgement relevant to working in a medium to large sized commercial organisation, to be based on sound psychometric principles, to be face valid with managers, to be usable by HR and training managers and other professionals, to provide information which can aid in recruitment, development and training interventions, and to be based on real life managerial practice (Howard & Choi, 2004). At present no South African norms or research are currently available for this instrument.

Detailed information relating to the development of Scenarios can be found in the Technical Manual (Howard & Choi, 2004). The steps to its development took place in nine distinct phases and will be briefly explained below:

Step 1 - Initial Pilot Work

A version of Wagner and Sternberg's (1985) Tacit Knowledge for Business Management was utilised in presenting situations and associated responses to 53 early career graduates. It was found that their scores were significantly correlated with four performance appraisal ratings given by their direct supervisors.

Step 2 – The Managerial Judgement Model

Management competency models were examined for competency definitions which incorporated the use of Managerial Judgement, e.g. "aspects of prioritising, decision making, managing staff, concern for impact, networking, organisational politics, and so on" (Howard & Choi, 2004:13). Furthermore, theories were explored surrounding how Managerial Judgement was considered and measured in the past. Four major and ten minor themes were identified as covering the area of Managerial Judgement. These themes directed the initial item pool generation.

Step 3 – The Beginnings of the Scenarios Test

Possible responses to the initial item pool of 34 situations were devised based on the developer's experience of working in and with large organisations. These 34 situations were then sent to directors and experienced managers (in pairs and 45 in total). Two types of responses were requested: effective and ineffective ones. These replies were then analysed qualitatively for content, with some items being discarded for providing poor discrimination between good and poor responses. This process resulted in 19 scenarios with a total of 251 response items.

The themes being measured at this stage were identified as:
- Managing People
- Managing Self / Priorities
- Managing Managers
- Influencing
- Managing Career and Reputation
- Networking / Managing Contacts

Step 4 – Preparing the Scenarios Test for Trialling

The item pool of 19 situations and 251 items was too large to be completed in its entirety by one individual, and so was split into four trial versions, A, B, C and D. At this stage it was established that a negative set of rating options (-1 to -3) alongside the positive options (+1 to +3) was needed to reflect that should someone execute an inappropriate task, matters can be made worse when compared to undertaking no action at all. Therefore, Managerial Judgement is not only concerned with appropriate action, but also knowing what one should definitely not do. In order to correct answers to the test items, 123 experts in the business/corporate management field were consulted. Consensus-based answers were established. Items that did not achieve this clear modal response were dropped. These responses provide the initial "scoring key" for the item pool.

Step 5: Trialling the Scenarios Test (Part 1)

At this stage groups of individuals who differed in their amount of scope of experience of working in large organisations completed one of the four versions described above. There were three subject groups (N=125); students (N=58), junior managers (N=27) and senior managers (N=40). Items were retained at this stage if they showed the potential to differentiate between the expert group and the other three groups. Further, detailed performance ratings were available for the senior and junior manager groups, making it possible to correlate the items with this performance data.

At this stage the following themes were being measured:
- Managing Objectives (24 items)
- People Management (36 items)
- Managing and Influencing Superiors (20 items)
- Reputation Management (28 items)

Step 6 – Trialling the Scenarios Test (Part 2)

In order to establish the final correct answers for the Scenarios Test items, the opinions of further experts were obtained (N=174). These were primarily chosen by examining modal answers. Furthermore, the assembled set of correct answers were then checked statistically through reliability checks. In total, 84 items and 16 situations were retained from the 108 items in the final trial version following the final scoring key development.

Step 7 – Trialling the Scenarios Test (Part 3)

Further data was collected for each of the three subject groups described earlier. The scoring key and scale items were used in order to construct both norm tables and to estimate reliability information. A total of 212 subjects" responses were analysed to establish internal reliability information and to provide data on scale construction. Three Managerial Judgement sub-scales were constructed from the items in the final Managerial Judgement scale using internal reliability checks. Items were assigned on a purely rational basis dependent on item content. However, cross-checks were run to ensure that an item did not load more highly on one of the other sub-scales. The three sub-scales were established as:

- Managing Objectives (30 items)
- People Management (28 items)
- Reputation Management (26 items)

In order to maintain a balance of correct responses throughout the test, 16 further relatively neutral items were added to the final 84 items. These new items were effectively dummy or research (non-scored) items.

Step 8 – Development of the Styles Scales

This stage concerned the format of test feedback, i.e. how could individuals" results be best presented to them in order to provide meaningful information for development purposes? Information provided by the three sub-scale scores was felt to be too broad to give individuals meaningful feedback for developmental purposes. Bearing this in mind, six style scales were developed in order to provide more detailed information. These scales are intended for feedback only, and should not be used to support selection decisions. These six style scales provide specific information to the candidate in terms of the behavioural trends or styles favoured in addressing situations.

The final test contains the following scales and items:

The Managerial Judgement Scale (84 items)

- Managing Objectives (30 items)
    - Big Picture Style (16 items)

- Delegative Style (14 items)
- People Management (28 items)
    - Team Style (13 items)
    - One-to-One Style (15 items)
- Reputation Management (26 items)
    - Personal Recognition Style (9 items)
    - Company Protocol Style (17 items)

Additional Research Items (16 items)

Total Items: 100

Step 9 – Standardisation and Validation

In order to provide norms for the Scenarios Test, the samples already collected during the previous steps were added. The samples included senior managers/professionals, high-flier managers, junior managers and professionals, team leader applicants, undergraduates and A-level students.

Validation studies were also conducted on the above groups to provide initial data on the construct and criterion-related validity of the Scenarios Test. These are presented in detail in the Scenarios technical manual (Howard & Choi, 2004).

The Scenarios Test which is available today consists of 16 managerial-type work situations, each of which is followed by a number of responses or possible courses of action. Candidates are expected to complete the assessment instrument within 50 minutes, however it is not strictly timed. Normal testing conditions do apply and supervision of its completion is recommended.

2.5.2   Description of the Scales of Scenarios

The overall scale of Managerial Judgement (a total scale which includes the three sub-scales of Managing Objectives, People Management and Reputation Management) measures how effective one is at weighing up a variety of managerial situations and deciding on the best ways of handling them. Low score on the Managing Objectives sub-scale indicates that one is less effective at distributing their energies at work and delegating, with resultant less emphasis on the wider organisational objectives (SHL,

2006). Conversely, high scorers make the best possible use of their energies at work. The People Management sub-scale provides information relating to the handling of staff and team issues. Lastly, low scorers on the Reputation Management sub-scale tend to place their own and the team"s needs ahead of the organisation (SHL, 2006).

Low scorers on the first of the six style scales, Big Picture, are generally more inclined than their peers to focus on immediate team objectives. This is in contrast to high scorers who tend to advocate wide consultation with other teams on objectives and decisions, as well as align their efforts with the wider organisational objectives (SHL, 2006). The Delegative style scale indicates the degree to which individuals may become preoccupied with detail and take thinks on personally rather than delegate when appropriate (SHL, 2006). High scorers on the One-to-One style scale tend to approach others who may have motivation and performance issues. They offer these individuals support and coaching whilst emphasising organisational objectives. However, low scorers tend to be more inclined to reprimand, ignore or replace an individual with performance issues (SHL, 2006). High scorers on the Team style scale consult, clarify and tackle team motivation and performance problems, and communicate change well. People who score lowly may give the team the impression of being unable to assist with performance problems (SHL, 2006). The Personal Recognition style scale describes individuals who score highly as taking a balanced approach in attempting to manage their own image within the organisation. Further, they may overestimate the importance of letting results speak for themselves rather than occasionally highlighting their own achievements, whereas low scorers may overestimate their own personal contributions (SHL, 2006). Lastly, high scores on the Company Protocol style scale indicates that one is more likely to focus on reaching decisions and objectives through established roles and procedures. Moreover, high scores would indicate that one is more willing to accommodate the needs of other managers. On the other hand, low scores indicate that one is more likely to focus on outcomes as opposed to rules and procedures. Further, they may prioritise their own needs ahead of other managers" needs (SHL, 2006).

2.5.3   Validity and Reliability Evidence for Scenarios

Extensive validity and reliability evidence is available in the test"s technical manual. Tests of ability should have reliability figures above 0.70. Since individuals" performance on the Scenarios Test improves with time in many cases, e.g. as they acquire managerial

experience, the test-retest method is not the most appropriate way of assessing reliability. No alternate form is available so this method is also not appropriate. The reliability of the Scenarios Test was assessed via the internal consistency method. Cronbach‟s Alpha for the Managerial Judgement Scale was 0.85 (Howard & Choi, 2004). Results thus confirm the reliability of the Scenarios Test.

In terms of establishing validity, a number of different studies were undertaken to prove that Managerial Judgement was indeed the skill measured by the Scenarios Test. These studies examined construct validity in the following areas (included is a short summary of the outcome) (Howard & Choi, 2004):

- ➢ The relationship between the sub-scales of the Scenarios Test: Two studies indicated that all three sub-scales show close relationships with one another. This is to be expected as the three sub-scales all measure an aspect of Managerial Judgement.

- ➢ The Scenarios Test and cognitive ability: Two studies confirmed small overlaps with cognitive ability tests. Reasons for this include that cognitive ability is one component of Managerial Judgement, and that the format of the Scenarios Test partly relies on the ability to understand written information (Howard & Choi, 2004). Factor analytic studies strongly suggest that the Scenarios Test measures a construct separate from cognitive ability (Howard & Choi, 2004).

- ➢ The Scenarios Test and personality: The personality tests included Costa and McCrae‟s NEO Five Factor Inventory (1985) and the 16PF5. Factor analysis demonstrated the independence of the Scenarios Test from the major spheres of personality measured by the personality questionnaires.

- ➢ The Scenarios Test and biographical variables: Performance on the Scenarios Test is relatively independent of age and years‟ experience, but shows strong improvement with job level, and to a lesser degree, education level.

- ➢ The Scenarios Test and equal opportunities: Studies confirmed that Managerial Judgement as a skill is slightly dependent on cognitive ability and learning from experience, but not on gender or race.

In terms of the criterion related validity of Scenarios, 11 studies were conducted and are cited in the technical manual (Howard & Choi, 2004). The conclusion from the tests of validity is that Scenarios measures Managerial Judgement, a skill which is partly

dependent on learning from on-the-job experience and relevant training. It is independent from both cognitive ability and personality and hence the Scenarios Test is a useful addition to any battery of tests. This outcome of this particular research may help to assist in the accumulation of evidence towards the test"s construct validity in a South African context.

2.7 The Cognitive Ability Tests: Verbal Evaluation (VC1.1) and Interpreting Data (NC2.1)

Within the context of the workplace, tests measuring IQ do not have predictive validity in terms of job performance. For example, an Engineer and a Psychologist may have the same IQ, but this does not necessarily mean that the Psychologist would be a good Engineer and vice versa. The best process would be to conduct a job analysis in order to determine what abilities (as well as competencies and personality characteristics) are required at what level of complexity. This helps to establish content validity of the instruments selected to the requirements of the job. To this end, Guilford"s differential model of abilities has informed the conceptual framework for the selection of relevant cognitive ability tests. Guilford"s model proposes that the measurement of differential cognitive abilities relevant to the job requirements has far more utility and predictive validity than the measurement of a single general mental ability.

Verbal Evaluation (VC1.1) measures the ability to understand written passages and the logical evaluation of arguments. The test is highly relevant to the world of work where reading, interpreting and evaluation of written material are required.

The test consists of fifteen passages, each of which is followed by four statements. The candidate is required to read each passage and then to evaluate the statements which follow in the light of the passage which preceded them. The statements can be true, false or cannot be determined without further information. In selecting the correct response candidates must be able to comprehend the passage, select pertinent information from it, evaluate the relationship between the passage and the statement, recognise the assumptions on which the statement is based, and finally evaluate the logic of the statement in the context of the passage. The test is strictly timed, providing candidates with 30 minutes to complete 60 questions. This test combines both the power and speed characteristics of ability tests.

The passages refer to different occupational areas and are based on the type of information encountered at the work levels for which the test was intended, namely supervisors, administrators or junior to middle managers. The passages include extracts from text likely to feature in policy statements, overviews, guidelines or brochures. In an educational context, high scorers on the test are likely to have the ability to follow courses at school or tertiary institutions where verbal material is prevalent and where different views or arguments need to be assessed.

The internal consistency score for VC1.1 is recorded as 0.90 (SHL, 2008a). This was obtained from a South African sample of 11,235 applicants to, and incumbents of, various positions in various industry sectors, including mining and manufacturing, financial services, retail, telecommunication professional services, travel and tourism and consulting.

The Interpreting Data (NC2.1) test measures the ability to make correct decisions or inferences from numerical data. Straightforward statistical information and other numerical data are presented. The test is appropriate for any job involving analysis or decision-making based on numerical facts.

The test consists of eight tables of information with five questions relating to each table. The tables refer to degree results, allocation of sales revenue, exchange rates, workshop absence, population statistics, grocer's sales, an activities survey, and factory output. These have been deliberately chosen to illustrate different work contexts and different kinds of relationships between numbers. The task in each question is to select the appropriate table of information, recognise the relevant figures, identify the method for calculating the answer, perform the calculation and then select the correct answer from the five available options. "Cannot say" is also sometimes presented as a possible option and applies to additional information or data being required in order for the candidate to make a conclusion. The test consists of 40 items and candidates have 30 minutes in which to complete it.

Occupations in which this skill is a key element include that of accountant, auditor, cashier, financial clerk, valuer insurance broker, general manager, banker, local

government officer, office manager, personnel manager, solicitor, company secretary, market researcher, statistical clerk, sales representative, and branch and store manager. In an educational context, school or college subjects which involve the ability to make decisions on the basis of numerical or statistical data include Economics, Mathematics Statistics, Chemistry, Physics and the Social Sciences.

The internal consistency score for NC2.1 is recorded as 0.86 (SHL, 2008a). The sample was the same as that used to determine the reliability for VC1.1, i.e. a South African sample of 11,235 applicants to, and incumbents of, various positions in various industry sectors, including mining and manufacturing, financial services, retail, telecommunication professional services, travel and tourism and consulting.

Six validation studies (conducted in South Africa) have been documented outside of the information contained within the VC1.1 and NC2.1 technical manual. In summation, significant correlations have been found between the VC1.1 and NC2.1 to competencies assessed in interviews, presentations and role-plays. Moreover, a validation study also confirmed that the VC1.1 and NC2.1 were significantly correlated to the criterion College marks attained by candidates selected to attend a training course. Lastly, convergent validity was established by high correlations between VC1.1 and NC2.1 to the Cognitive Process Profile test (CPP). Of particular importance, the VC1.1 and NC2.1 have been validated for use within the South African context.

2.8   The Personality Questionnaire: Occupational Personality Questionnaire (OPQ32i)

SHL has developed a range of personality questionnaires which are designed to assess personality traits which affect an individual‟s performance at work. The personality questionnaire is based upon a trait theoretical approach. These theorists generally assume that traits are relatively stable over time, they differ among individuals, and that they influence behaviour. Personality also functions as a critical determinant of job success as people often search for environments where they can express their interests, values and attitudes. Specifically, the OPQ32 assesses 32 personality attributes within the following areas:

- Relationships with people

- Thinking and problem solving style

- Emotions, drives and motivation

- Team working styles

- Leadership and subordinate styles

- Selling and influencing styles

The OPQ is one of the world"s most widely used personality questionnaires within the context of work. There are two versions of the instrument; a normative and an ipsative. The normative questionnaire (OPQ32n) is used in low stakes applications, such as development, and has a multiple choice format to which individuals indicate the extent to which they agree or disagree with a series of statements. This version measures 32 attributes of personality and allows for the measurement of a Social Desirability scale, as opposed to a Consistency scale characteristic of the OPQ32i. The Social Desirability scale provides an indication of how open, honest and frank the candidate has been by presenting items such as "I never tell a lie". Should the candidate respond that this is "Always True", an element of socially desirable responding can be deduced. Individuals indicate the extent to which they agree or disagree with the 230 statements. The questionnaire takes approximately 35 minutes to complete.

The ipsative version (OPQ32i) is usually used in high stakes situations as it contains a Consistency scale which measures how consistent candidates have responded across the questionnaire, i.e. the more the answers to the questions change as the candidate moves through the questionnaire, the lower the Consistency score. Individuals choose between options indicating which in a block of four statements is most and least like them. Candidates normally take between 45 minutes to one hour to complete the questionnaire which contains 104 blocks of 4 statements each. This research utilises the OPQ32i.

Reliabilities of personality scales should fall within the range of 0.60 to 0.80. Extensive validity and reliability evidence is available in the technical manual (Bartram, Brown, Fleck, Inceoglu & Ward, 2006). Data from USA, South Africa, Japan, UK and other European countries suggest that while there are a few scales with internal consistency close to 0.7, the overall median consistency of the scales in the instrument can be considered to be in the range of 0.75 to 0.80. A large data set for South Africa shows a

median reliability of 0.81. The latest OPQ32i reliability study in South Africa was conducted in 2008 with a sample size of 13,523. Overall, the statistics show good internal consistency across the scales. Twenty-one of the scales obtained coefficients of 0.70 and higher and eleven scales ranged between 0.60 and 0.69 (Reliability Study for the OPQ32i, R062, 2008). All ethnic groups in this study obtained acceptable reliabilities indicating that the OPQ32i can be used with confidence in South Africa.

In terms of construct validity of the OPQ32i, comparisons with other instruments showed high correlations between scales with similar content and low correlations between scales that are different (Bartram et al, 2006). In addition, the OPQ32i was compared with other instruments measuring personality, such as the 16PF, Hogan Personality Inventory (HPI) and Myers-Briggs Type Indicator (MBTI). Criterion related validity has also been established through 7 studies, covering a total of 36 samples, with a total sample size of 5,416. In summary, the studies clearly reveal that personality scales of the OPQ32i are important predictors of workplace competencies (Bartram et al, 2006).

Studies concerned with establishing the relationship between Scenarios, cognitive ability and personality were conducted shortly after its development. Links with cognitive tests have shown to be moderate at best. Given the nature of the purported relationships between Scenarios and various cognitive ability tests, it is crucial to check whether or not Scenarios is operating as a cognitive ability test. This has formed the major impetus for this research. This research will not provide a definitive answer to the construct validity of SJTs, however, it will contribute evidence towards the construct validity of Scenarios through a discriminant validity investigation, i.e. a study of how Scenarios (as a particular SJT) is different to cognitive ability and personality.

2.9   Conclusion

SJTs are a type psychological test which presents the test taker with realistic yet hypothetical situations in an attempt to measure the candidate‟s judgement and action. Although this section has explained a number of benefits regarding their use, it is necessary to elucidate whether or not they provide additional information over and above that provided by personality instruments and ability tests.

This section has also outlined the specific SJT used in this research, namely Scenarios, and has provided a step by step account of the research into its development. Lastly, research which has been developed regarding Scenarios" correlation to other forms of assessment has also been delineated. In order to demonstrate the psychometric integrity of the cognitive ability tests and personality questionnaire, a description of their format, and reliability and validity evidence have been provided.

CHAPTER 3: RESEARCH DESIGN

3.1   Introduction

This section outlines the aims of the research. The sample group is described, as well as the method of data collection. The specific tools utilised in this project are provided, which includes both the ability tests, and the personality questionnaire. The research methodology is provided, including both the strengths and the limitations of this form of research. Lastly, ethical considerations are acknowledged.

3.2   Research Aim

The aim of this research is to investigate whether or not a relationship exists between individuals" scores on Scenarios and their personality and cognitive ability test results. Furthermore, the aim is to generate research specific to the South African context as Scenarios has yet to be adapted to use within a South African occupational context.

3.3   Sample

The data was collected from a group of candidates employed by a leading international tobacco company. The division from which the data was collected was based in the Gauteng area. The sample consists of 93 candidates; 71 male and 22 female (please refer to Table 1). The mean age is 31.27 years, with a minimum of 22 years and a maximum of 44 years (please refer to Table 2). However, 23 respondents did not complete this information. In terms of the number of years within the job position, only 43 respondents provided this information (please refer to Table 3). The average was 2.7 years, with a minimum of 0 (indicating less than 12 months) and a maximum of 12. 52.86% of the sample is African, 8.57% Indian, and 38.57% White. Again, 23 candidates did not provide their ethnic category (please refer to Table 4). Concerning level of education, 70 respondents provided this information, where 15.71% had a matric, 10% a post-matric certificate, 51.43% a degree and 22.86% a post-graduate qualification (please refer to Table 5).

Although the sample size was 93 respondents for the Scenarios Test and personality questionnaire, not all in the sample completed the same ability instruments. Only 79 of the 93 completed the ability instruments from the Critical Reasoning Test Battery (viz. VC1.1

and NC2.1). The remaining respondents (N=14) completed different ability tests from the Management and Graduate Item Bank (viz. VMG3 and NMG3).

Table 1: Gender Composition

Gender	Count	Cumulative	Percent	Cumulative %	% of all	Cumulative %
Male	71	71	76.34	76.34	76.34	76.34
Female	22	93	23.66	100.00	23.66	100.00
Missing	0	93	0.00		0.00	100.00

Table 2: Age Composition

	Valid N	Mean	Minimum	Maximum	Standard Deviation
Age (Years)	70	31.27	22	44	4.43

Table 3: Number of Years within the Position

	Valid N	Mean	Minimum	Maximum	Standard Deviation
Years in Position	43	2.70	0	12	2.04

Table 4: Ethnicity

Ethnicity	Count	Cumulative	Percent	Cumulative %	% of all	Cumulative %
African	37	37	52.86	52.86	39.78	39.78
Indian	6	43	8.57	61.43	6.45	46.24
White	27	70	38.57	100.00	29.03	75.27
Missing	23	93	32.86		24.73	100.00

Table 5: Education Level

Education Level	Count	Cumulative	Percent	Cumulative %	% of all	Cumulative %
Matric	11	11	15.71	15.71	11.83	11.83
Certificate	7	18	10.00	25.71	7.53	19.35
Degree	36	54	51.43	77.14	38.71	58.06
Post Graduate	16	70	22.86	100.00	17.20	75.27
Missing	23	93	32.86		24.73	100.00

3.4 Data Collection

The data was collected from respondents in a group format through pre-arranged assessment sessions where the four assessments (VC1.1, NC2.1, Scenarios Test and OPQ32i) were administered to the sample groups.

SHL employs trained test administrators to conduct occupational test administration sessions for clients. The data obtained from the instruments used provides useful information with which decisions can be made for a variety of applications, such as for recruitment, selection, development, opportunity for training, amongst numerous others. As part of the test administration process, candidates are asked to complete two forms prior to the start of the assessment session. The first is a Biographical Data Form[2] which requires the candidate to provide useful biographical and demographic information which is used by SHL for research purposes, such as for the generation of up to date norm information. Furthermore, this information can be used to investigate a wide range of research questions in order to maintain the psychometric integrity and usefulness of the instruments, as well as keep it up to date within the current occupational testing context. The test administrator informs the candidates of the use of this form prior to its completion. The second form is a consent form[3] whereby the candidate acknowledges their voluntary participation in the process, and provides their permission for the results to be used for whatever the intended purpose may be, i.e. for recruitment purposes, or selection for training and development. There is also a second block which informs the candidate that should they be required to undertake these tests again, within a period of twelve to eighteen months, they can use the same results from that test administration session.

---

[2] Please refer to Appendix A for a copy of the Biographical Data Form
[3] Please refer to Appendix B for a copy of the Consent Form

However, they can elect to redo the assessment instruments should they wish. Importantly, SHL recommends reassessment should the period of eighteen months lapse, as there could be significant development of one‟s competency potential within this time frame.

3.4.1 Data Collection Tools

The research employed the use of two cognitive ability measures, namely Verbal Evaluation (VC1.1) and Interpreting Data (NC2.1), the Occupational Personality Questionnaire (OPQ32i), as well as the SJT, Scenarios.

3.5 Research Methodology

This study uses a quantitative analysis of the data obtained from the psychometric assessment instruments used.

3.5.1 Strengths of Quantitative Research

The main strengths of quantitative research lie in its precision and control (Burns, 2000). Control is achieved through the sampling and design, and precision is achieved through the reliable measurement through quantitative methods (Burns, 2000). Furthermore, quantitative measurement leads to statements about causation, i.e. the manipulation of a specific variable can be shown to have a direct causal effect on other variables (Burns, 2000). This has been argued to provide information that is scientific and thus accurate.

3.5.2 Limitations of Quantitative Research

According to Burns (2000), an important limitation of quantitative research is that it does not take into account peoples‟ ability to interpret their own experiences and thus construct their own meanings. It is also important to note that a quantitative approach cannot be completely objective. Subjectivity is involved in the creation of a problem worthy of investigation and in the interpretation of the results (Burns, 2000).

3.6 Quantitative Analysis Tool

The data was analysed using STATISTICA™. This is a statistics and analytics software package developed by StatSoft. STATISTICA™ provides data analysis, data management, data mining, and data visualisation procedures.

3.6.1 The Pearson Product Moment Correlation Coefficient

The most powerful and commonly used of the correlation coefficients is the Pearson Product-Moment Correlation Coefficient (PMCC and typically denoted as r). It is a measure of the correlation (or linear dependence) between two variables, and provides a value between +1 and -1 inclusive. It provides both a strength and measure of the direction of the relationship between the two variables (Cohen, 1988). The PMCC was used to analyse the scale interrelationships of Scenarios, VC1.1, NC2.1 and the OPQ32i. The PMCC was also used to analyse the correlations between the Scenarios Test with that of the three other instruments (viz. VC1.1, NC2.1 and OPQ32i).

3.7 Ethical Considerations

Social research should conform to four basic ethical principles (De Vaus, 2001). The first principle is that of voluntary participation. Participants should not be led to believe that they are required to participate in a study. Furthermore, they should know that they can withdraw at any point in time. Participants of this research were informed of the voluntary nature of involvement, and were provided the opportunity to withdraw at any stage.

Secondly, informed consent should be obtained after the researcher has explained the purpose of the study and its associated benefits, and the use of the data that will be obtained. An explanation of how the respondents were selected should also be provided. Although a basic description of the nature of the study is necessary, care should be taken not to provide detailed information about the study, i.e. hypotheses and theories that are being tested. This could distort the way people answer questions and so undermine the validity of the findings. This is the most important principle in conducting social research (Gomm, 2004). The participants in this study were selected as part of an organisational selection process to identify potential for managerial roles. Participants were provided with a detailed description of the purpose and process of the study, and were given detailed feedback of their assessment results with the opportunity to ask questions and receive answers.

The third ethical principle is that of doing no harm to the participants (De Vaus, 2001). This means that no individual should be asked to co-operate in research that results in a sense of self-denigration and embarrassment. Although it could be argued that the process of assessment may be stressful for the participant, it is imperative that the explanation,

administration and feedback to participants is conducted in a professional and confidential manner. This was conducted in accordance with professional ethical guidelines as set out by legislation and professional bodies, such as the HPCSA.

Fourthly, anonymity and confidentiality should be maintained (De Vaus, 2001). Individuals participating in research are entitled to expect that they cannot be identified as the source of any particular information. Participants in this study have total anonymity.

3.8   Conclusion

This chapter explicated the research aim, sample, the sampling method, the data collection tools, and the methodology employed to analyse the data. Lastly, some ethical standards in conducting research were noted, with specific reference to this research topic.

CHAPTER 4: RESEARCH RESULTS

4.1 Introduction

This section outlines the results of the assessment instruments namely Verbal Evaluation (VC1.1), Interpreting Data (NC2.1), the Occupational Personality Questionnaire (OPQ32i) and Scenarios which were administered to the group of candidates (N=92). This is presented in three sections. The first two sections provide a preliminary analysis of the data. The first section provides the results of the candidates" completion of the ability tests, the OPQ32i, and Scenarios, as well as the inter-correlations between the ability tests and the scales of the OPQ32i, as well as the scales of Scenarios in order to ascertain whether or not the scales of the instruments themselves measure discrete constructs. Lastly, correlations between Scenarios and each of the instruments are provided in order to determine whether or not Scenarios is measuring something unique which is not tapped into by the ability and personality instruments.

4.2 Preliminary Data Analysis of Inter-Correlations

Table 6 below lists the scores obtained by the 79 candidates who completed both the VC1.1 and NC2.1. Although the sample size for the Scenarios Test and OPQ32i was 92, 13 of the respondents completed different ability tests at a higher level of complexity to the ones utilised by this research.

The mean score for VC1.1 was 37.11, with a minimum of 15 and a maximum of 54. The standard deviation was 8.3 raw score points, and Cronbach"s Alpha reliability coefficient is 0.84. For NC2.1, the mean score was 18.72, with a minimum of 7 and a maximum of 32. The standard deviation was 5.34 raw score points, and Cronbach"s Alpha reliability coefficient is 0.83.

Table 6: Results of Cognitive Ability Test Scores

Ability Scores	Valid N	Mean	Minimum	Maximum	Standard Deviation	Cronbach Alpha
Verbal Evaluation	79	37.11	15	54	8.30	0.84
Interpreting Data	79	18.72	7	32	5.34	0.83

Table 7 below lists the results of the same 92 candidates concerning their scores on the 32 scales of the OPQ32i. The highest score average on the OPQ32i was for the Persuasive scale where the mean for the group of 92 candidates was 19.93. The lowest score on the OPQ32i scales was for the Worrying scale where the score mean was 6.32. The scale which showed the most difference between the 92 candidates" scores was the Controlling scale, where the standard deviation was the equivalent of 4.98 raw score units.

Table 7: Results of OPQ32i

OPQ32i Scales	Valid N	Mean	Minimum	Maximum	Standard Deviation	Cronbach Alpha
Persuasive	92	19.93	7	26	3.75	0.75
Controlling	92	15.65	0	26	4.98	0.82
Outspoken	92	14.15	5	23	4.22	0.71
Independent Minded	92	11.24	2	24	4.08	0.70
Outgoing	92	11.89	3	24	4.22	0.75
Affiliative	92	11.53	3	22	4.02	0.78
Socially Confident	92	15.01	3	21	3.47	0.63
Modest	92	11.28	2	25	4.76	0.83
Democratic	92	13.90	4	24	4.19	0.72
Caring	92	12.66	5	22	3.82	0.71
Data Rational	92	13.15	2	25	4.58	0.81
Evaluative	92	14.64	9	23	3.24	0.53
Behavioural	92	12.58	4	21	3.73	0.66
Conventional	92	10.12	2	18	3.62	0.67
Conceptual	92	11.40	2	24	4.26	0.71
Innovative	92	14.24	2	25	4.90	0.83
Variety Seeking	92	12.93	4	23	4.06	0.68
Adaptable	92	11.34	2	23	4.68	0.79
Forward Thinking	92	14.63	6	23	4.11	0.76
Detail Conscious	92	13.08	3	22	4.18	0.71
Conscientious	92	17.89	8	25	3.35	0.67
Rule Following	92	10.59	1	23	4.56	0.83
Relaxed	92	10.71	2	18	3.43	0.61
Worrying	92	6.32	0	20	4.12	0.81
Tough Minded	92	12.17	4	22	3.51	0.56
Optimistic	92	14.87	7	24	3.57	0.66
Trusting	92	8.41	0	18	4.28	0.82
Emotionally Controlled	92	8.90	1	19	3.52	0.69
Vigorous	92	13.64	6	23	3.45	0.62
Competitive	92	16.28	5	25	4.62	0.77
Achieving	92	19.28	9	25	3.18	0.56
Decisive	92	11.57	1	24	4.91	0.81

Table 8 below provides the results of the scores of the 92 candidates on the scales of the Scenarios SJT. The highest mean score of 77.42 was obtained for the scale Big Picture, and the lowest mean score of 27.08 for the scale Personal Recognition. The People Management scale had the highest standard deviation of 7.91 raw score points, and the Personal Recognition had the lowest standard deviation of 4.74 raw score points.

Table 8: Results from Scenarios

Scenarios Scales					
	Valid N	*Mean*	*Minimum*	*Maximum*	*Standard Deviation*
Managing Objectives	92	36.15	19	56	7.31
People Management	92	35.83	13	55	7.91
Reputation Management	92	32.30	18	46	6.39
Big Picture	92	77.42	66	93	5.68
One-To-One	92	71.61	55	85	5.96
Personal Recognition	92	27.08	17	38	4.74
Delegative	92	53.88	42	66	5.43
Team	92	57.37	44	71	4.98
Company Protocol	92	75.77	58	96	6.83
TOTAL: Managerial Judgement	92	104.28	67	149	15.90

Table 9 below reveals that the correlation coefficient for VC1.1 and NC2.1 is 0.6302, indicating that they are significantly correlated at the 1% level. In terms of the sample, this means that in general those who were high scorers on one instrument, tended to score high on the other instrument too. Conversely, low scorers on one instrument tended to be low scorers on the other. Even though the VC1.1 and NC2.1 measure different abilities, there is still a general factor of intelligence embedded within both. This finding is similar to that found in other studies examining the inter-correlation of ability tests (Wilson, Baron & Borkowski, 1991).

Table 9: Cognitive Ability Test Inter-Correlations

Ability Tests	*Verbal Evaluation (VC1.1)*	*Interpreting Data (NC2.1)*
Verbal Evaluation (VC1.1)	1	0.6302**
Interpreting Data (NC2.1)	0.6302**	1

** $p \leq 0.01$

Table 10 below provides the inter-correlation scores between the 32 scales of the OPQ32i. Inter-correlations are an important consideration for multi-scale instruments such as the OPQ32 questionnaires as they indicate how closely related different constructs are to each other.

OPQ32 interpretation is generally based on an understanding of how individual scales work in combination or „link" with each other. It is therefore important to understand whether a particular combination of scales is commonplace or relatively unusual in the same way as it is useful to know whether an individual"s score on a single scale is typical or extreme in comparison to others (Bartram et al, 2006). For example, the scales of Socially Confident and Worrying are negatively correlated, so we would expect that an individual who is Socially Confident is less likely to score high on the Worrying scale.

The number of scales in the OPQ32 model means that they cannot be completely independent of each other. The design philosophy of the questionnaire is to produce a multi-faceted profile of behavioural style (Bartram et al, 2006). This means that broad domains are split into multiple scales which will inevitably inter-correlate to some extent (Bartram et al, 2006). Although this can lead to some redundancy in purely statistical terms, it allows greater depth and detail of interpretation for profiles, particularly where an individual is different from the general trend, implied by the correlation (Bartram et al, 2006). Other examples include the fact that people who score high on the scale of Socially Confident tend to be more Outgoing and extrovert in style, and most profiles will show similar scores on these scales (Batram et al, 2006). The OPQ32i is negatively biased with correlations generally lower (less positive or more negative) than the OPQ32n. This is because the ipsative format places a constraint on the scale variances, forcing raw scale scores sum to a constant for any individual (Bartram et al, 2006). Therefore, a dependence exists between the different scale scores by constraining the overall sum of scores. As a consequence, the mean scale inter-correlation of an ipsative measure is always negative, and approaches zero as the number of scales increases. It has also been argued that normative correlations may be slightly positively inflated due to response biases (Saville & Willson, 1991). The correlations between scales of the normative version (OPQ32n) range from –0.51 to 0.56, with the vast majority falling between –0.2 and 0.2 (Bartram et al, 2006). For the ipsative questionnaire, scale inter-correlations range from –0.36 to 0.58 with many very small indeed (Bartram et al, 2006). This suggests a generally high degree

of independence for the scales, despite the large number of relatively narrow scales included. However, raw score correlations can underestimate true relationships where scale reliabilities are relatively low.

In summation, the provision of inter-correlation scores determines whether or not the personality attribute scales of the OPQ32i provide information concerning different aspects of personality. High and many inter-correlations would not be preferable as this would indicate that the questionnaire could be significantly reduced in item number. Thus the same or similar results could be obtained within a shorter administration time.

The results indicate that the OPQ32i demonstrates few scale inter-correlations, providing evidence towards its scientific psychometric integrity. Therefore the 32 scales provide unique insight into varying attributes of personality.

Table 10: Occupational Personality Questionnaire (OPQ32i) Scale Inter-Correlations

	Relaxed	Decisive	Worrying	Tough Minded	Optimistic	Trusting	Emotionally Controlled	Vigorous	Competitive	Achieving	Persuasive	Caring	Controlling	Outspoken	Independent Minded	Outgoing
Relaxed	1.0000	0.1523	-0.2515*	0.2795**	0.0785	-0.0755	0.0814	-0.1380	0.0435	-0.1665	0.2355*	0.0176	-0.0550	0.0959	-0.0806	-0.1308
Decisive	0.1523	1.0000	-0.2058*	0.0944	-0.0089	-0.1141	0.0115	0.0711	-0.0609	-0.0687	-0.2064*	-0.1692	0.1206	0.1757	0.0980	-0.1419
Worrying	-0.2515*	-0.2058*	1.0000	-0.1611	-0.0427	-0.0821	0.3331**	-0.0290	0.0051	-0.1123	-0.1693	-0.2062*	-0.2998**	-0.2391*	0.1549	0.0437
Tough Minded	0.2795**	0.0944	-0.1611	1.0000	0.0352	-0.0743	0.1306	0.0161	-0.2248**	-0.1117	0.3203**	-0.0802	-0.0815	-0.0174	-0.0298	-0.0366
Optimistic	0.0785	-0.0089	-0.0427	0.0352	1.0000	0.2564*	-0.0973	-0.1874	0.1441	-0.1301	-0.0934	0.0911	-0.0403	0.0225	-0.0469	0.0158
Trusting	-0.0755	-0.1141	-0.0821	-0.0743	0.2564*	1.0000	-0.1023	-0.2038	-0.1414	-0.2390*	-0.1400	0.4046**	0.1161	-0.0218	-0.2158*	-0.0492
Emotionally Controlled	0.0814	0.0115	0.3331**	0.1306	-0.0973	-0.1023	1.0000	-0.1160	-0.2227*	-0.1555	-0.0772	-0.0107	-0.4069**	-0.2583*	0.0729	-0.2386*
Vigorous	-0.1380	0.0711	-0.0290	0.0161	-0.1874	-0.2038	-0.1160	1.0000	0.1442	0.2662**	-0.0919	-0.1644	0.0924	-0.0158	-0.0688	0.0818
Competitive	0.0435	-0.0609	0.0051	-0.2248*	0.1441	-0.1414	-0.2227*	0.1442	1.0000	0.1760	-0.0376	-0.2620**	0.2467*	0.1456	0.1615	0.2798**
Achieving	-0.1665	-0.0687	-0.1123	-0.1117	-0.1301	-0.2390*	-0.1555	0.2662**	0.1760	1.0000	0.1010	-0.1214	0.1394	-0.1514	0.0404	0.0907
Persuasive	0.2355*	-0.2064*	-0.1693	0.3203**	-0.0934	-0.1400	-0.0772	-0.0919	-0.0376	0.1010	1.0000	-0.1899	-0.0743	-0.1030	-0.1392	0.0288
Caring	0.0176	-0.1692	-0.2062*	-0.0802	0.0911	0.4046**	-0.0107	-0.1644	-0.2620**	-0.1214	-0.1899	1.0000	-0.0265	0.0148	-0.1184	0.0332
Controlling	-0.0550	0.1206	-0.2998**	-0.0815	-0.0403	0.1161	-0.4069**	0.0924	0.2467*	0.1394	-0.0743	-0.0265	1.0000	0.2592*	-0.0814	0.0710
Outspoken	0.0959	0.1757	-0.2391*	-0.0174	0.0225	-0.0218	-0.2583*	-0.0158	0.1456	-0.1514	-0.1030	0.0148	0.2592*	1.0000	0.3130**	0.2853**
Independent Minded	-0.0806	0.0980	0.1549	-0.0298	-0.0469	-0.2158*	0.0729	-0.0688	0.1615	0.0404	-0.1392	-0.1184	-0.0814	0.3130**	1.0000	0.1607
Outgoing	-0.1308	-0.1419	0.0437	-0.0366	0.0158	-0.0492	-0.2386*	0.0818	0.2798**	0.0907	0.0288	0.0332	0.0710	0.2853**	0.1607	1.0000

	Relaxed	Decisive	Worrying	Tough Minded	Optimistic	Trusting	Emotionally Controlled	Vigorous	Competitive	Achieving	Persuasive	Caring	Controlling	Outspoken	Independent Minded	Outgoing
Affiliative	-0.1527	-0.1901	0.1858	-0.1803	0.0798	0.0464	-0.1236	-0.0043	0.1863	0.0044	-0.0334	0.2531*	-0.1070	0.1766	0.1449	0.5024**
Socially Confident	0.0872	-0.2152*	-0.1831	0.1778	0.0915	-0.0380	-0.2449*	0.1132	-0.0358	0.1509	0.2900**	0.1381	-0.0768	-0.0505	-0.2970**	0.3631**
Modest	-0.1692	0.0147	0.2566*	0.0569	-0.1697	-0.1539	0.4430**	-0.0472	-0.3442**	-0.0227	-0.1849	-0.0116	-0.3463**	-0.2867**	0.2319*	-0.2590*
Democratic	-0.0710	-0.1170	-0.1567	0.0918	-0.0376	0.3001**	0.0105	-0.0640	-0.3992**	-0.1768	-0.1069	0.2677**	-0.0122	0.0071	-0.2736**	-0.2273*
Data Rational	-0.0300	-0.1608	-0.0165	-0.2824**	-0.0485	0.0343	-0.1417	-0.1766	0.1466	-0.1772	-0.1212	0.0219	0.1766	-0.1293	-0.1410	-0.1478
Detail Conscious	0.0039	-0.1675	0.0330	-0.2355*	-0.3004**	-0.1533	0.1544	-0.0811	-0.1826	-0.0074	0.1027	-0.0976	-0.1414	-0.2201*	-0.1262	-0.4242**
Conscientious	-0.1855	-0.1023	-0.1310	-0.1741	-0.2600*	-0.2003	-0.1295	0.1693	-0.1619	0.2551*	0.1079	0.0151	-0.0991	-0.2499*	-0.1524	-0.0949
Rule Following	-0.1162	-0.2163*	0.2794**	-0.1412	-0.0810	-0.0863	0.2956**	-0.1520	-0.1082	-0.1971	0.0389	0.0956	-0.3868**	-0.2460*	-0.1342	-0.0933
Evaluative	-0.1402	0.1531	-0.2250*	0.0210	0.0197	-0.0019	-0.2914**	-0.0028	-0.1025	0.0632	-0.0300	-0.2081*	0.1994	0.0708	-0.0641	-0.2008
Behavioural	-0.1267	-0.1469	-0.2841**	0.1342	-0.2517*	0.1280	-0.1866	0.1331	-0.2231*	0.0731	0.0200	0.1620	0.0950	-0.1370	-0.2938**	-0.1280
Conventional	-0.0405	-0.2238*	0.3337**	-0.0510	-0.0234	-0.0131	0.2994**	-0.1942	-0.1242	-0.2126*	-0.0415	0.0793	-0.3446**	-0.1250	0.0464	-0.2381*
Conceptual	0.1813	0.2838**	-0.0868	0.0291	-0.0825	-0.0556	-0.1015	0.0450	-0.2175*	-0.1446	-0.1106	-0.2053*	-0.0478	0.0822	-0.0379	-0.1543
Innovative	0.0566	0.0999	-0.1626	0.0967	0.1011	-0.0336	-0.0981	-0.0897	-0.0725	-0.0431	0.0739	-0.1356	0.1153	-0.2098*	-0.1906	-0.2697**
Variety Seeking	-0.0598	0.0912	0.2060*	-0.1320	0.0259	-0.1640	0.0857	0.0195	0.1152	0.1171	-0.1578	-0.2526*	-0.0049	-0.0187	0.2791**	0.1267
Adaptable	-0.1856	-0.1327	0.3287**	-0.1449	-0.0217	-0.0306	0.1703	-0.0502	-0.0497	0.0267	0.0921	-0.0736	-0.2918**	-0.2120*	-0.1362	0.0392
Forward Thinking	0.0164	0.0507	-0.2476*	-0.0343	-0.0033	0.0088	-0.0527	-0.0744	0.0119	0.0181	-0.0201	-0.1691	0.0452	-0.0639	-0.1749	-0.2603*

	Affiliative	Socially Confident	Modest	Democratic	Data Rational	Detail Conscious	Conscientious	Rule Following	Evaluative	Behavioural	Conventional	Conceptual	Innovative	Variety Seeking	Adaptable	Forward Thinking
Relaxed	-0.1527	0.0872	-0.1692	-0.0710	-0.0300	0.0039	-0.1855	-0.1162	-0.1402	-0.1267	-0.0405	0.1813	0.0566	-0.0598	-0.1856	0.0164
Decisive	-0.1901	-0.2152*	0.0147	-0.1170	-0.1608	-0.1675	-0.1023	-0.2163*	0.1531	-0.1469	-0.2238*	0.2838**	0.0999	0.0912	-0.1327	0.0507
Worrying	0.1858	-0.1831	0.2566*	-0.1567	-0.0165	0.0330	-0.1310	0.2794**	-0.2250*	-0.2841**	0.3337**	-0.0868	-0.1626	0.2060*	0.3287**	-0.2476*
Tough Minded	-0.1803	0.1778	0.0569	0.0918	-0.2824**	-0.2355*	-0.1741	-0.1412	0.0210	0.1342	-0.0510	0.0291	0.0967	-0.1320	-0.1449	-0.0343
Optimistic	0.0798	0.0915	-0.1697	-0.0376	-0.0485	-0.3004**	-0.2600*	-0.0810	0.0197	-0.2517*	-0.0234	-0.0825	0.1011	0.0259	-0.0217	-0.0033
Trusting	0.0464	-0.0380	-0.1539	0.3001**	0.0343	-0.1533	-0.2003	-0.0863	-0.0019	0.1280	-0.0131	-0.0556	-0.0336	-0.1640	-0.0306	0.0088
Emotionally Controlled	-0.1236	-0.2449*	0.4430**	0.0105	-0.1417	0.1544	-0.1295	0.2956**	-0.2914**	-0.1866	0.2994**	-0.1015	-0.0981	0.0857	0.1703	-0.0527
Vigorous	-0.0043	0.1132	-0.0472	-0.0640	-0.1766	-0.0811	0.1693	-0.1520	-0.0028	0.1331	-0.1942	0.0450	-0.0897	0.0195	-0.0502	-0.0744
Competitive	0.1863	-0.0358	-0.3442**	-0.3992**	0.1466	-0.1826	-0.1619	-0.1082	-0.1025	-0.2231*	-0.1242	-0.2175*	-0.0725	0.1152	-0.0497	0.0119
Achieving	0.0044	0.1509	-0.0227	-0.1768	-0.1772	-0.0074	0.2551*	-0.1971	0.0632	0.0731	-0.2126*	-0.1446	-0.0431	0.1171	0.0267	0.0181
Persuasive	-0.0334	0.2900**	-0.1849	-0.1069	-0.1212	0.1027	0.1079	0.0389	-0.0300	0.0200	-0.0415	-0.1106	0.0739	-0.1578	0.0921	-0.0201
Caring	0.2531*	0.1381	-0.0116	0.2677*	0.0219	-0.0976	0.0151	0.0956	-0.2081*	0.1620	0.0793	-0.2053*	-0.1356	-0.2526*	-0.0736	-0.1691
Controlling	-0.1070	-0.0768	-0.3463**	-0.0122	0.1766	-0.1414	-0.0991	-0.3868**	0.1994	0.0950	-0.3446**	-0.0478	0.1153	-0.0049	-0.2918**	0.0452
Outspoken	0.1766	-0.0505	-0.2867**	0.0071	-0.1293	-0.2201*	-0.2499*	-0.2460*	0.0708	-0.1370	-0.1250	0.0822	-0.2098*	-0.0187	-0.2120*	-0.0639
Independent Minded	0.1449	-0.2970**	0.2319*	-0.2736**	-0.1410	-0.1262	-0.1524	-0.1342	-0.0641	-0.2938**	0.0464	-0.0379	-0.1906	0.2791**	-0.1362	-0.1749
Outgoing	0.5024**	0.3631**	-0.2590*	-0.2273*	-0.1478	-0.4242**	-0.0949	-0.0933	-0.2008	-0.1280	-0.2381*	-0.1543	-0.2697**	0.1267	0.0392	-0.2603*

	Affiliative	Socially Confident	Modest	Democratic	Data Rational	Detail Conscious	Conscientious	Rule Following	Evaluative	Behavioural	Conventional	Conceptual	Innovative	Variety Seeking	Adaptable	Forward Thinking
Affiliative	1.0000	0.2256*	-0.1427	-0.1346	-0.0964	-0.2050*	0.0100	0.1128	-0.2229*	-0.1202	0.0257	-0.2570*	-0.4482**	-0.1802	0.0546	-0.3200**
Socially Confident	0.2256	1.0000	-0.1379	-0.0703	-0.3040**	-0.1190	0.1711	-0.1207	-0.0896	0.2220*	-0.1156	-0.2227*	-0.1114	-0.0444	0.0505	-0.1699
Modest	-0.1427	-0.1379	1.0000	-0.0206	-0.0479	0.0806	0.0742	0.2049*	-0.0660	-0.0006	0.2668**	-0.1541	-0.1833	0.1817	0.1204	-0.2005
Democratic	-0.1346	-0.0703	-0.0206	1.0000	0.0794	0.0651	-0.1034	-0.0096	0.1976	0.2472*	-0.1174	0.0300	0.0071	-0.2332*	-0.1262	0.1843
Data Rational	-0.0964	-0.3040**	-0.0479	0.0794	1.0000	0.1619	-0.0913	0.0378	0.0489	0.0663	-0.1304	0.1496	0.1131	-0.1355	-0.0717	0.0456
Detail Conscious	-0.2050	-0.1190	0.0806	0.0651	0.1619	1.0000	0.4483**	0.3258**	0.0726	0.0014	0.2512*	-0.0091	0.0281	-0.2704**	-0.0058	0.0732
Conscientious	0.0100	0.1711	0.0742	-0.1034	-0.0913	0.4483**	1.0000	0.1840	0.1036	0.1817	0.0825	-0.1947	-0.0707	-0.0990	0.0696	-0.0667
Rule Following	0.1128	-0.1207	0.2049*	-0.0096	0.0378	0.3258**	0.1840	1.0000	-0.2624*	-0.2567*	0.5715**	-0.1572	-0.2781**	-0.2604*	0.1978	-0.0762
Evaluative	-0.2229	-0.0896	-0.0660	0.1976	0.0489	0.0726	0.1036	-0.2624*	1.0000	0.0891	-0.1845	0.1739	0.1377	-0.1113	-0.3044**	0.2736**
Behavioural	-0.1202	0.2220*	-0.0006	0.2472*	0.0663	0.0014	0.1817	-0.2567*	0.0891	1.0000	-0.1150	-0.1033	0.1048	-0.2043	-0.0094	0.1172
Conventional	0.0257	-0.1156	0.2668**	-0.1174	-0.1304	0.2512*	0.0825	0.5715**	-0.1845	-0.1150	1.0000	-0.0765	-0.1838	-0.1945	-0.0082	-0.0973
Conceptual	-0.2570	-0.2227*	-0.1541	0.0300	0.1496	-0.0091	-0.1947	-0.1572	0.1739	-0.1033	-0.0765	1.0000	0.3299**	-0.0741	-0.0824	0.0933
Innovative	-0.4482	-0.1114	-0.1833	0.0071	0.1131	0.0281	-0.0707	-0.2781**	0.1377	0.1048	-0.1838	0.3299**	1.0000	0.0837	-0.2217*	0.2794**
Variety Seeking	-0.1802	-0.0444	0.1817	-0.2332*	-0.1355	-0.2704**	-0.0990	-0.2604*	-0.1113	-0.2043	-0.1945	-0.0741	0.0837	1.0000	0.1308	0.0031
Adaptable	0.0546	0.0505	0.1204	-0.1262	-0.0717	-0.0058	0.0696	0.1978	-0.3044**	-0.0094	-0.0082	-0.0824	-0.2217*	0.1308	1.0000	-0.2201*
Forward Thinking	-0.3200	-0.1699	-0.2005	0.1843	0.0456	0.0732	-0.0667	-0.0762	0.2736**	0.1172	-0.0973	0.0933	0.2794**	0.0031	-0.2201*	1.0000

* p ≤ 0.05

** p ≤ 0.01

Table 11 below provides the inter-correlation scores for Scenarios. The results reveal that there are quite a few inter-correlations at both the 99% and 95% level. This could indicate that although Scenarios seeks to measure a number of different constructs relating to Managerial Judgement, there may be an underlying umbrella component which is being measured, such as tacit (or practical) knowledge. Therefore, although Scenarios provides information relating to various job related constructs, these may be grouped to indicate a measurement of a single underlying construct. It is the opinion of the author that it is possible that the scales of Scenarios are all tapping into (at least to some extent) cognitive ability, personality, job experience, or some combination of these three variables, which could be an explanation for the significant correlations found between the scales. Despite this, Scenarios provides an objective and scientific measurement of a range of individual difference variables those of which may be critical for job success.

Table 11: Scenarios Scale Inter-Correlations

	M.O.	P.Mgt	R.M	TOTAL	Big Picture	One-To-One	P. R.	Delegative	Team	Company Protocol
Managing Objectives	1.0000	0.2969**	0.2785**	0.7192**	-0.5925**	-0.2597*	-0.0235	-0.5988**	-0.2513*	-0.3598**
People Management	0.2969**	1.0000	0.3536**	0.7759**	0.0242	-0.8232**	-0.3817**	-0.4677**	-0.7689**	-0.3987**
Reputation Management	0.2785**	0.3536**	1.0000	0.7057**	0.1306	-0.4284**	-0.5236**	-0.4960**	-0.0651	-0.6778**
TOTAL	0.7192**	0.7759**	0.7057**	1.0000	-0.2078*	-0.7009**	-0.4110**	-0.7071**	-0.5240**	-0.6360**
Big Picture	-0.5925**	0.0242	0.1306	-0.2078*	1.0000	-0.1440	-0.2875**	-0.1264	0.1553	0.0422
One-To-One	-0.2597*	-0.8232**	-0.4284**	-0.7009**	-0.1440	1.0000	0.4215**	0.5229**	0.3276**	0.4392**
Personal Recognition	-0.0235	-0.3817**	-0.5236**	-0.4110**	-0.2875**	0.4215**	1.0000	0.3098**	0.1562	0.0920
Delegative	-0.5988**	-0.4677**	-0.4960**	-0.7071**	-0.1264	0.5229**	0.3098**	1.0000	0.2285*	0.5131**
Team	-0.2513*	-0.7689**	-0.0651	-0.5240**	0.1553	0.3276**	0.1562	0.2285**	1.0000	0.2003
Company Protocol	-0.3598**	-0.3987**	-0.6778**	-0.6360**	0.0422	0.4392**	0.0920	0.5131**	0.2003	1.0000

* $p \leq 0.05$

** $p \leq 0.01$

4.3   Correlation Data Analysis

This section provides an analysis of the correlations between Scenarios and the ability tests results (VC1.1 and NC2.1), the scales of the OPQ32i, as well as the weighted scores of the competencies of the IMC.

Table 12 below provides the correlation between the 10 scales of Scenarios and the two ability tests. The results reveal that only the scale of Reputation Management of Scenarios correlates negatively with both VC1.1 ($p \leq 0.01$) and NC2.1 ($p \leq 0.05$).

Table 12: Correlations between Scenarios and the Cognitive Ability Tests (VC1.1 and NC2.1)

	Verbal Evaluation (VC1.1)	*Interpreting Data (NC2.1)*
Managing Objectives	-0.1050	-0.0561
People Management	0.0074	0.0312
Reputation Management	-0.3089**	-0.2693*
Managerial Judgement	-0.1688	-0.1180
Big Picture	0.0466	0.0227
One-To-One	0.0696	-0.0188
Personal Recognition	0.2076	0.1692
Delegative	0.1859	0.0855
Team	-0.1117	-0.0581
Company Protocol	0.0965	0.0934

* $p \leq 0.05$
** $p \leq 0.01$

Table 13 below provides the correlation between the 10 scales of Scenarios and the 32 scales of the OPQ32i. The Managing Objectives scale has a negative correlation with the following scales of the OPQ32i ($p \leq 0.05$): Controlling, Evaluative and Behavioural. Furthermore, the Managing Objectives scale correlates positively with Competitive and Rule Following, and correlates negatively with the Democratic of the scale of the OPQ32i ($p \leq 0.01$).

The People Management scale of Scenarios correlates positively with the Tough Minded, Rule Following and Adaptable scales of the OPQ32i ($p \leq 0.05$). This same scale of Scenarios has a negative correlation with the Controlling scale of the OPQ32i ($p \leq 0.01$).

The Reputation Management scale of Scenarios reveals a positive correlation with the Persuasive scale of the OPQ32i ($p \leq 0.05$), and a positive correlation with Relaxed ($p \leq 0.01$).

The Total scale (which combines Managing Objectives, People Management and Reputation Management) correlates positively with the following OPQ32i scales ($p \leq 0.05$): Persuasive and Adaptable, and negatively with Democratic, Evaluative and Behavioural at the same confidence level. It correlates negatively with Controlling and positively with Rule Following ($p \leq 0.01$).

The Big Picture scale of Scenarios has a positive correlation with Democratic and Behavioural, and a negative correlation with Rule Following ($p \leq 0.05$). This same scale correlates negatively with the Competitive scale of the OPQ32i ($p \leq 0.01$).

The One-to-One scale of Scenarios shows a negative correlation with the Tough Minded, Rule Following and Adaptable scales of the OPQ32i ($p \leq 0.05$). The One-to-One scale has a positive correlation with the Controlling scale of the OPQ32i ($p \leq 0.01$).

The Personal Recognition scale of Scenarios correlates negatively with the scales of Tough Minded and Adaptable of the OPQ32i ($p \leq 0.05$). This same scale of Scenarios correlates positively with the Data Rational scale of the OPQ32i ($p \leq 0.01$).

The Delegative scale of Scenarios correlates negatively with the Persuasive and Rule Following scales of the OPQ32i ($p \leq 0.05$). There is a positive correlation with the Controlling scale at this same confidence level. It does not correlate to any of the OPQ32i scales at the 99% confidence level.

The Team scale of Scenarios reveals a positive correlation with the Data Rational scale of the OPQ32i ($p \leq 0.05$) and does not correlate to any of the OPQ32i scales at the 99% level of confidence.

The Company Protocol scale of Scenarios correlates negatively with the Relaxed OPQ32i scale ($p \leq 0.05$), and does not correlate with any other scale of the OPQ32i at the 99% level.

Table 13: Correlations between Scenarios and the Occupational Personality Questionnaire (OPQ32i)

	Relaxed	Decisive	Worrying	Tough Minded	Optimistic	Trusting	Emotionally Controlled	Vigorous	Competitive	Achieving	Persuasive	Caring
Managing Objectives	0.0562	-0.0618	0.0586	-0.0456	0.0564	-0.2021	0.1673	-0.0662	0.2826**	-0.1275	0.1710	-0.1621
People Management	-0.0266	-0.0532	0.1042	0.2286*	0.1136	0.0336	0.1495	0.0657	0.0008	0.0413	0.1636	0.0713
Reputation Management	0.3735**	0.0540	0.0188	0.1144	0.0109	-0.0071	0.0957	-0.0971	0.0793	-0.1134	0.2350*	0.0381
TOTAL	0.1627	-0.0332	0.0863	0.1387	0.0868	-0.0790	0.1897	-0.0368	0.1621	-0.0837	0.2544*	-0.0238
Big Picture	0.0945	0.0374	-0.0733	0.1993	-0.0222	0.1982	-0.1250	0.0034	-0.2836**	-0.0006	0.0091	0.1669
One-To-One	-0.0799	-0.0006	-0.1218	-0.2100*	-0.0401	0.0064	-0.1988	0.0129	0.0168	0.0146	-0.1384	-0.0981
Personal Recognition	-0.1969	-0.1105	-0.1132	-0.2429*	-0.0683	0.1062	0.0354	-0.1145	-0.0512	0.0605	-0.1427	0.0720
Delegative	-0.1260	0.0446	-0.0120	-0.1438	-0.0076	0.0660	-0.0070	0.1437	-0.0977	0.0529	-0.3122*	0.0410
Team	0.1680	0.0579	-0.0287	-0.1446	-0.0899	-0.0227	-0.0556	-0.1455	-0.0065	-0.1342	-0.1276	0.0598
Company Protocol	-0.2185*	0.0144	0.0365	-0.0874	0.0632	0.1073	-0.1758	0.0263	-0.0694	0.0723	-0.1892	-0.0178

	Controlling	Outspoken	Independent Minded	Outgoing	Affiliative	Socially Confident	Modest	Democratic	Data Rational	Detail Conscious	Conscientious	Rule Following
Managing Objectives	-0.2210*	0.0271	0.0858	-0.0334	0.0402	-0.0616	0.0227	-0.2773**	0.1620	0.1032	-0.0989	0.2926**
People Management	-0.2834**	-0.0595	-0.1013	0.0410	-0.0105	-0.0003	-0.0690	-0.1274	-0.1636	-0.0405	-0.0206	0.2536*
Reputation Management	-0.1460	0.0542	0.0318	-0.0967	-0.0957	-0.1400	-0.0736	-0.1304	-0.0475	0.0473	-0.1477	0.1078
TOTAL	-0.3012**	0.0046	0.0018	-0.0338	-0.0252	-0.0848	-0.0535	-0.2432*	-0.0260	0.0463	-0.1151	0.3039**
Big Picture	0.1111	0.0042	-0.0994	-0.0642	-0.0975	0.0701	-0.1231	0.2320*	-0.1141	-0.0194	0.0781	-0.2046*
One-To-One	0.2835**	0.0811	0.0536	0.0508	0.0189	0.0379	0.0094	0.1279	0.0888	0.0744	0.0638	-0.2264*
Personal Recognition	0.1578	-0.0798	0.1054	0.0351	0.0359	-0.1151	0.0955	0.1899	0.3000**	0.0513	0.0151	-0.1278
Delegative	0.2361*	-0.0088	-0.0076	0.0451	-0.0227	-0.0414	0.1034	0.1296	-0.0324	-0.0742	0.0023	-0.2135*
Team	0.1573	0.0088	0.1022	-0.0813	-0.0149	-0.0359	0.0567	0.0887	0.2116*	-0.0209	-0.0502	-0.1500
Company Protocol	0.1937	0.0184	-0.0797	0.0400	-0.0215	0.1700	0.0230	0.0461	-0.0689	-0.0864	0.1228	-0.1776

	Evaluative	Behavioural	Conventional	Conceptual	Innovative	Variety Seeking	Adaptable	Forward Thinking
Managing Objectives	-0.2163*	-0.2597*	0.1999	-0.0401	-0.0545	-0.0178	0.1132	0.0107
People Management	-0.1453	-0.0703	0.1151	-0.0645	-0.1002	-0.1147	0.2410*	-0.0348
Reputation Management	-0.0998	-0.1310	0.0692	0.0439	-0.0245	-0.1064	0.1892	-0.1111
TOTAL	-0.2118*	-0.2070*	0.1769	-0.0329	-0.0847	-0.1080	0.2479*	-0.0570
Big Picture	0.1804	0.2161*	-0.1959	0.0865	0.0054	-0.0493	0.0119	-0.0953
One-To-One	0.0933	0.0028	-0.0441	0.0967	0.0634	0.0471	-0.2288*	0.0456
Personal Recognition	0.1013	0.0447	-0.1446	-0.1682	0.0219	0.1957	-0.2386*	0.1904
Delegative	0.1219	0.0507	-0.1458	0.0287	0.1528	0.1214	-0.1161	-0.0517
Team	0.1118	0.0411	-0.1402	-0.0019	0.1310	0.1240	-0.2034	0.0041
Company Protocol	0.0812	0.0958	-0.0504	-0.0172	0.1381	0.0974	-0.1204	0.0130

* $p \leq 0.05$

** $p \leq 0.01$

4.4   Conclusion

This section has provided a surface presentation of the correlation results obtained from the sample population of 92 candidates who completed the OPQ32i and Scenarios SJT, and 79 of the 92 candidates who completed the VC1.1 and NC2.1. This has been in the form of demonstrating the positive and negative correlations between the ten scales of Scenarios and each of the other instruments used.

CHAPTER 5: DISCUSSION OF THE RESULTS

5.1   Introduction

This section provides an analysis and explanation of the data obtained from the correlations between Scenarios and the ability tests and personality questionnaire in consideration with the literature surrounding SJT research.

5.2   Discussion of the Results in Table 12 (Correlations between Scenarios and the Cognitive Ability Tests)

The results reveal that only the scale of Reputation Management of Scenarios correlates negatively with both VC1.1 (p ≤ 0.01) and NC2.1 (p ≤ 0.05). Reputation Management measures the degree to which people place the organisation‟s reputation above their own and/or the team‟s needs (SHL, 2006). The results indicate that those individuals who are less concerned about the management of the organisation‟s reputation at the expense of their own (or team‟s) scored lower on the ability tests. The converse of this would be those individuals who considered the organisation‟s reputation were more likely to score higher on the VC1.1 and NC2.1. This result differs from that found by Howard & Choi (2004). Their study revealed that there was no correlation between the Scenarios scale of Reputation Management and the ability tests administered. The ability tests comprised of two numerical reasoning and one verbal reasoning test as well as the Watson Glaser. The Managing Judgement scale of Scenarios correlated to the Watson Glaser (timed) instrument (r=0.14) and the verbal reasoning test (r=0.25) at the 95% level of confidence (Howard & Choi, 2004). The Managing Objectives scale of Scenarios correlated with the Watson Glaser (Timed) test at the 98% confidence level (r=0.20) and the Watson Glaser (Untimed) at the 95% confidence level (r=0.18) (Howard & Choi, 2004). The People Management Scenarios‟ scale correlated with the verbal reasoning test at the 95% confidence level (r=0.22) (Howard & Choi, 2004).

In consideration of the influence of task and contextual knowledge on SJT performance, the results of this study may indicate that those test items of Scenarios designed to measure Reputation Management are tapping into task knowledge in relation to job performance.

Significantly, the composite scale score of Managerial Judgement did not correlate with either the verbal cognitive ability test or the numerical cognitive ability test. Taxonomies of managerial performance generally indicate that cognitive ability forms a critical foundation for effective managerial performance. Despite this assumption, the lack of correlation is congruent with other studies which have sought to determine the relationship between SJTs to that of cognitive ability (Bergman et al, 2006; Mumford et al, 2008; Weekley & Jones, 1999; Mullins & Schmitt, 1998). In particular, Mumford et al"s (2008) SJT of Team Role Knowledge not only demonstrated low correlations with mental ability, but also indicated that TRT scores provided incremental validity beyond cognitive ability.

However, in consideration of Fertig"s (2009) study, the incremental validity of the SJT in measuring Managerial Judgement over and above that of cognitive ability was questionable. Therefore, although this particular SJT has been found to demonstrate low correlations with cognitive ability, it would be useful to establish the predictive validity of the instruments used in relation to the construct of Managerial Judgement. This could only be determined by examining performance related information of the sample used in this study through rating scores of on-the-job performance. Due to the fact that the assessment centre was used for selection purposes (as opposed to a developmental initiative) the sample size would be far too small to undertake such a validation exercise. Moreover, even if the sample size was large enough (N>50), one has to consider the effects of the restriction of range due to the fact that only high scorers would be successful.

Furthermore, despite the fact that the SJT of Managerial Judgement has been shown to have limited correlations with verbal and numerical cognitive ability through a divergent correlation study, it is also crucial that convergent validity be established with other measures of Managerial Judgement. This could also be determined by establishing concurrent and predictive validity to criterion measures of job performance.

The results of this study indicate that Scenarios would provide additional information about a candidate over and above that of verbal and numerical ability tests at a junior managerial level, and would therefore not be considered an additional unnecessary

expense in a battery of assessments. However, it has been suggested that knowledge is more proximal to performance than cognitive ability (Mumford et al, 2008). This would imply that job knowledge is an important contributor to job performance as opposed to cognitive ability alone. SJTs are developed with a consideration to the context of application. Therefore, knowledge and familiarity of the job context would be a crucial determinant of successful performance as opposed to cognitive ability alone. This would suggest that successful performance on an SJT designed to measure Managerial Judgement relies on managerial experience and exposure as opposed to cognitive ability.

5.3   Discussion of the Results in Table 13 (Correlations between Scenarios and the OPQ32i)

This study revealed that the following scales of the OPQ32i demonstrate no correlation with any of the scales of the Scenarios SJT: Decisive, Worrying, Optimistic, Trusting, Emotionally Controlled, Vigorous, Achieving, Caring, Outspoken, Independent Minded, Outgoing, Affiliative, Socially Confident, Modest, Detail Conscious, Conscientious, Conventional, Conceptual, Innovative, Variety Seeking, and Forward Thinking.

The results have also shown that the Team, Company Protocol and Reputation Management scales of Scenarios show the least correlation with the OPQ32i (correlation with one, one and two scales of the OPQ32i respectively). In consideration of the correlation between the Reputation Management scale with the cognitive ability test, the non-correlation with personality would support the finding that the items tapping into Reputation Management are more heavily dependent on task knowledge as opposed to contextual knowledge. This would indicate that items measuring Reputation Management require activities that support and maintain the technical core of managerial-prescribed functions.

Scales which demonstrate the most correlation with the OPQ32i include the Total Managerial Judgement and Managing Objectives scales (correlating with a total of seven and six scales respectively), but even this is not high enough to warrant a sufficient degree of overlap between the two instruments. In consideration of the Managing Objectives scale of Scenarios, this scale may require an understanding of

contextual knowledge in relation to the vignettes proposed by the relevant items. This would indicate that the knowledge required for these items are not typically written in a job description, but are nonetheless considered an important component of job performance. In particular, the literature has revealed that an element of contextual knowledge requires endorsing, supporting and defending organisational objectives, a critical determinant of Managing Objectives (Bess, 2001). The findings also indicate that the overall measure of Managerial Judgement may be more reliant on contextual knowledge rather than task knowledge. The emphasis on contextual knowledge would be in congruence with a low fidelity SJT measure in that the scenarios would be designed to be reflective of most organisational contexts. Should scenarios be focussed on the particular nuances of a specific organisation, the instrument would have little relevance for other workplace contexts, thus rendering it inappropriate for a wide range of users.

A study conducted by Howard & Choi (2004) revealed that eight of the scales of the Occupational Personality Questionnaire correlated significantly with the Managerial Judgement scale and the sub-scales of the Scenarios Test. The results of this study suggested that high scorers on the Scenarios Test tend to be more independent, affiliative, artistic, innovative, optimistic, achieving, decisive and less practical (Howard & Choi, 2004). The results of this study indicate that high scorers on the Scenarios Test tend to be more persuasive, controlling, democratic, rule following, evaluative, behavioural and adaptable. The results therefore do not mirror those recorded in the earlier study conducted by Howard & Choi (2004). This may be due largely to the small sample size, and thus it would be recommended that a similar study be conducted on a larger sample.

A further finding by Howard & Choi (2004) was the lack of correlation between the Reputation Management sub-scale of the Scenarios Test with any aspect of personality. The results of this study revealed the correlation between Reputation Management and the Relaxed and Persuasive scales of the OPQ32i. As indicated above, it would be recommended that the study be repeated with a larger more representative sample.

Although there is limited research investigating the relationship between SJTs and personality, there are studies which report a relationship (Bess, 2001; McDaniel et al, 2001; Mullins & Schmitt, 1998; Smith & McDaniel, 1998). Bess (2001) determined that only personality revealed a relationship with the SJT, linking this finding to the requirement of contextual knowledge evident in the SJT items. Strahan et al (2005) found that personality is not a strong determinant of performance on the SJT with only Agreeableness showing a significant correlation.

There are also studies which report similar results to the current research. Bergman et al (2006) determined that no correlation exceeded 0.32 in absolute value, suggesting that the SJT, Leadership Skills Assessment (LSA), did not demonstrate a strong relationship with either cognitive ability or personality.

The results have indicated that despite the correlations between Scenarios and a few of the scales of the OPQ32i, the administration of the Scenarios Test in a battery would help to shed further light on aspects of an individual that are not revealed by personality assessment alone.

5.4   General Discussion of Research Results

The difficulties with investigating the construct validity of SJTs have been elucidated, with particular reference to common pitfalls in SJT test development. This concerns the fact that researchers tend to report results and frame their data in method rather than construct terms and that SJTs are often designed with the criterion in mind. The particular SJT employed in this research was developed with a key consideration to the construct under investigation, viz. Managerial Judgement. Furthermore, although Christian et al (2010) argue that using a composite scale score may contaminate the construct under investigation, this would not apply to Scenarios SJT as the composite score is comprised of six style scales all designed to measure the construct of Managerial Judgement.

The study undertaken by Bergman et al (2006) closely mirrors that undertaken by this research in respect of the construct under investigation as well as the development of the SJT, i.e. Bergman et al"s (2006) SJT sought to measure Leadership Skills whereas the SJT employed by this research measures Managerial Judgement. Bergman et al"s

(2006) LSA SJT was developed using expert-based scoring keys, which also formed part of the developmental stages of the Scenarios SJT. In other words, subject matter experts (SMEs) determine the items" relevance to the criterion. Both studies indicate low correlations of the SJT with cognitive ability and personality measures. These studies reveal the importance of accurately defining the construct purported to be measured by the SJT. The obvious detriment of ignoring a rigorous definition of the construct under scrutiny is the inability to cross-compare research results.

On review of the literature and related findings of this research study, perhaps what has been highlighted is the pressing need to provide stricter theoretical underpinnings for the development of SJTs. A thorough consideration to the particular construct/s required to be measured needs to inform test development, including items that may tap into task knowledge or contextual knowledge. However, it is acknowledged that this may be difficult when the primary emphasis is on the criterion requiring investigation. For example, in the context of assessment centres in the workplace, the context of job performance and various critical incidents inform the content of the development of the SJT as opposed to discrete or distinct constructs. However, this does not necessarily indicate that SJTs developed with the criterion in mind measure a construct in and of themselves, such as tacit knowledge or practical intelligence. Quite possibly, the SJTs may be tapping into a wide range of constructs, signifying the importance of establishing convergent and divergent validity with other instruments. In the context of assessment centres within the workplace, the obvious emphasis is on the utility of the test with a firm consideration to the cost component. Not only should the SJT provide evidence over and above that of other measures used, should an SJT be developed with the criterion of job performance in mind, organisations would do well to establish concurrent or predictive validity in comparison with that of other measures used (should sample sizes allow for such an undertaking).

In line with Bess" (2001) study into the differentiation between task and contextual knowledge in determining SJT performance, it may be worthwhile to investigate the items of the Scenarios SJT to determine their emphasis on one or both of the knowledge domains (task knowledge or contextual knowledge). This may help to shed some insight into the correlations between some of the scales of Scenarios with that of the cognitive ability tests, as well as the correlations with some of the scales of

the personality questionnaire. This would also help to shed insight into how the construct of Managerial Judgement is measured by an SJT in terms of possible personality influencers and cognitive ability determinants.

A further important consideration is the fact that some constructs may be more contaminated by cognitive ability and/or personality than other constructs. Those that are would demonstrate more significant relationships. It therefore becomes crucial to consider the construct validation of each SJT in isolation. One cannot draw on SJT validation studies without a thorough consideration of the construct under investigation. As indicated earlier, this points to the necessity of framing SJT test development within method measurement terms, as opposed to a broad determination of "practical intelligence". In this manner, SJTs become hollow instruments which can measure a variety of constructs, those of which may or may not be dependent upon cognitive ability or personality.

5.5   Conclusion

The Corporate Protocol and the Team scales of Scenarios demonstrate the least correlation with any of the other instruments and assessment techniques used. Reputation Management is also one of the scales which demonstrated the lower correlation with the OPQ32i, however, it was the only scale to correlate with both ability tests. In general, the results have indicated that Scenarios demonstrates limited correlation with the verbal and numerical ability tests, and personality questionnaire used. Thus, Scenarios demonstrates low test inter-correlation with the assessment tools used in this research.

It is quite possible that SJTs reflect the vehicle for a method to measure a variety of different constructs, i.e. Scenarios measures the construct of Managerial Judgement, or one‟s judgement in situations which require leadership skills. In contrast, an SJT could be designed to measure the competency of communication. Therefore, one does not necessarily measure an underlying tacit knowledge. Moreover, the constructs measured may or may not correlate with cognitive ability or personality. This would have to be determined for each SJT measuring its purported construct/s. This is a finding demonstrated by other research studies, such as Trippe (2002) where the multitrait-multimethod matrix (MTMMM) was used to examine the construct validity

of an SJT. This method allows one to measure more than one construct using more than one method. Correlations among measures of the same trait using different methods provide an indication of how well measures of the trait converge, and thus provides evidence that they are measuring the intended constructs (Trippe, 2002). To this end, an SJT was developed by the researcher to measure three FFM personality traits, Conscientiousness, Agreeableness, and Openness to Experience. SJT scales designed to measure these factors should exhibit convergent validity, whereas discriminant validity should be demonstrated for the remaining two factors of Extraversion and Neuroticism. The study concluded that SJTs can indeed function as a modality for measuring any number of constructs (Trippe, 2002). The study further revealed that it is possible for SJTs to be developed to measure specific constructs (Trippe, 2002). Furthermore, Trippe (2002) indicated that SJTs can be conceptualised as a measurement modality rather than indicators of a new and unique construct. However, Trippe (2002) cautions that it is not outside the realm of possibility that the SJT method factor is actually some kind of "judgement" construct, i.e. judgement of communication within particular situations, or judgement of leadership requirements within certain scenarios.

CHAPTER 6: RECOMMENDATIONS AND CONCLUSION

6.1   Introduction

This section will investigate possible areas for future research concerning the construct validity of SJTs in general. Limitations regarding this specific research will be highlighted with a view to correcting for these in future investigations.

6.2   Areas for Future Research

The identification and assessment of practical problem solving skills represent promising returns for the study of practical intelligence and tacit knowledge. However, despite this benefit, further research needs to be generated building on the assumption that SJTs measure constructs related to experience and performance, and are distinct from measures of cognitive ability and personality. Should an SJT be developed to measure a particular construct, it is critical that it is constructed with a theoretical consideration of the construct in mind. This would inform test validation through an investigation of divergent and convergent validity evidence with other instruments. Alternatively, in the context of the workplace, should an SJT be designed with a criterion in mind, concurrent and predictive validity should be established with job performance. The SJT utilised in the research was designed to measure the particular construct of Managerial Judgement which also relates to the criterion of managerial performance within work contexts. Therefore, future research to validate the instrument could investigate the incremental validity of Scenarios over and above that of cognitive ability and personality measures. Moreover, integrity of the construct under investigation could be supported through convergent validity studies with other tests purporting to measure Managerial Judgement.

The above recommendations are supported by Bess" (2001) study. It is stated that should future research confirm the dimensionality suggested by the differences between task and contextual knowledge, then the use of SJTs will need to become more tailored in both their development and use. Specifically, test users will have to better identify what constructs they wish to measure.

A further consideration for future research concerns whether or not SJTs are prone to practice effects. As the particular SJT employed in this research is only available in its

current form, should it be used in high-stakes context (such as recruitment and selection) it is critical that knowledge of the items will not affect performance on subsequent administrations. Therefore, future research should examine the effects of practice and coaching on SJT scores and validity (Lievens, Buyse & Sackett, 2005).

6.3   Limitations of the Study

A larger more representative sample should be used with specific participant demographic information collected. The sample was not large enough to test for race effects. This could help to determine what, if any, adverse impact exists within the instrument. In addition, future research should also obtain participant data such as job performance information, which could help to establish criterion-related validation of the instrument. Despite the fact that this study did not focus on determining the correlation between job performance and scores on the various instruments, this measure could also help to determine how the various instruments predict job performance.

As it has been acknowledged that limited research has investigated the correlation of SJTs with personality, it would be worthwhile to investigate the relationship between Scenarios with that of the Five Factor Model of Personality. This would assist in the cross-comparison with other studies which have examined this link. Examining the relationship of the SJT to scales, or traits, measured by the personality questionnaire has allowed for a limited comparison with other research studies.

A further limitation of the study concerns an understanding of the items of the Scenarios test in relation to task and contextual knowledge. The relationships found by this study (albeit low correlations) could help to frame the hypotheses for future studies, i.e. that those SJT scales correlating with cognitive ability are dependent upon task knowledge, whereas those SJT scales which reveal correlations with personality are largely dependent upon contextual knowledge.

6.4   Conclusion

This research study has focused on investigating whether or not a specific SJT, Scenarios, measures something unique to that of three specific measures of ability (measured through the VC1.1 and NC2.1) and personality (measured by the OPQ32i).

The results indicated that Scenarios does tap into a unique construct not entirely measured by that of ability and personality.

REFERENCES

Bartram, D. (2004). Assessment in Organisations: Special Issue in Advances in Testing Methodology from an International Perspective, *Applied Technology,* 53(2), 237 - 259.

Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, 90, 1185-1203.

Bartram, D. & Brown, A. (2005). *Great Eight Factor Model OPQ32 Report*. Surrey: SHL Group Plc.

Batram, D., Brown, A., Fleck, S., Inceoglu, I., & Ward, K. (2006). *OPQ32: Manual and Users' Guide.* Surrey: SHL Group Plc.

Bartram, D., Robertson, L. T., & Callinen, M. (2002). Introduction: A Framework for Examining Organisational Effectiveness. In I. T. Robertson, M. Callinen & D. Bartram (Eds.), *Organisational Effectiveness: The role of Psychology* (pp. 1-11). Chichester: Wiley.

Bergman, M.E., Drasgow, F., Donovan, M.A., Henning, J.B. & Juraska, S.E. (2006). Scoring Situational Judgement Tests: Once You Get the Data, Your Troubles Begin. *International Journal of Selection and Assessment,* 14(3), 223-235.

Bess, T.L. (2001). *Exploring the Dimensionality of Situational Judgement: Task and Contextual Knowledge.* Unpublished Master"s Thesis, Virginia Polytechnic Institute and State University.

Bledow, R. & Frese, M. (2009). A Situational Judgement Test of Personal Initiative and its relationship to Performance. *Personnel Psychology,* 62(2), 229-258.

Bless, C. & Higson-Smith, C. (1995). *Fundamentals of Social Research Methods: An African Perspective* (2nd ed.). Kenwyn: Juta.

Borman, W.C. & Motowildo, S.J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance,* 10, 99-102.

Burke, E. (2006). *Better Practice for Unsupervised Online Assessment.* SHL White paper 2006. Surrey: SHL Group Plc.

Burns, R.B. (2000). *Introduction to Research Methods* (4th edition). London: Sage.

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.

Chamorro-Premuzic, T. & Furnham, A. (2010). *The Psychology of Personnel Selection.* Cambridge: Cambridge University Press.

Chan, D., Borman, W., Lievans, F., McDaniel, M.A., Ployhart, R.E. & Schmitt, N. (2006). *Situational Judgement Tests: Construct Validity and Directions for Future Research.* Panel Discussion at the 21st Annual Conference for the Society of Industrial and Organisational Psychology, Dallas, Texas.

Chan, D. & Schmitt, N. (1997). Video-Based Versus Paper-and-Pencil Method of Assessment in Situational Judgement Tests: Sub-Group Differences in Test Performance and Face Validity Perceptions. *Journal of Applied Psychology,* 82, 143–159.

Chan, D. & Schmitt, N. (2002). Situational Judgement and Job Performance. *Human Performance*, 15 (3), 233–254.

Christensen, L.B. (2001). *Experimental Methodology* (8th ed.). London: Allyn & Bacon.

Clevenger, J. Pereira, G.M., Wiechmann, D., Schmitt, N. & Harvey, V.S. (2001). Incremental Validity of Situational Judgement Tests. *Journal of Applied Psychology,* 86, 410-417.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioural Science* (2nd ed.). New Jersey: Erlbaum.

Conner, L.A. (2007). *Evaluation of the Situational Judgement Test*. Unpublished Master"s Thesis, University of North Texas.

Creighton, P. & Scott, N. (2006). An Introduction to Situational Judgement Inventories. *Selection and Development Review,* 22, 3-6.

De Vaus, D. (2001). *Research Design in Social Research.* London: Sage.

Durrheim, K. (2002). Quantitative Measurement. In M. Terre Blanche & K. Durrheim (Eds), *Research in Practice: Applied Methods for the Social Sciences* (pp. 72-95). Cape Town: University of Cape Town Press.

Fertig, S. (2009). *The Incremental Validity of a Situational Judgement Test (SJT) Relative to Personality and Cognitive Ability to Predict Managerial Performance.* Unpublished Master"s Thesis, Stellenbosch University, Stellenbosch.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic Efforts in the Description of Leader Behaviour: A Synthesis and Functional Interpretation. *Leadership Quarterly*, 2(4), 245-287.

Gessner, T.L. & Klimoski, R.J. (2006). Making Sense of Situations. In J.A. Weekley & R.E. Ployhart (Eds), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 13-38). New Jersey: Lawrence Erlbaum Associates.

Gomm, R. (2004). *Social Research Methodology: A Critical Introduction.* New York: Palgrave Macmillan.

Graziano, A.M. & Raulin, M.L. (2000). *Process Inquiry: Research Methods.* London: Allyn & Bacon.

Gregory, R.J. (1996). *Psychological Testing: History, Principles and Applications* (2nd ed.). Boston, MA: Allyn & Bacon.

Health Professions Council of South Africa Act No. 56 of 1974.

Hedland, J., Antonakis, J. & Sternberg, R. (2002). *Tacit Knowledge and Practical Intelligence: Understanding the Lessons of Experience.* United States Army Research Institute for the Behavioural and Social Sciences.

Howard, A. & Choi, M. (2004). *Scenarios: Technical Manual.* London: SHL Group Plc.

Hunter, J.E. & Hunter, R.F. (19984). Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin,* 96(1), 72-98.

Joiner, D. (2008). *WRIPAC Presentation,* presented in Sparks, Nevada, on the 3rd of October 2008.

Konradt, U., Hertel, G. & Joder, K. (2003). Web-Based Assessment of Call Centre Agents: Development and Validation of a Computerised Instrument, *International Journal of Selection and Assessment,* 11, 184-193.

Kurz, R. & Bartram, D. (2002). Competency and Individual Performance: Modelling the World of Work. In I.T. Robertson, M. Callinen & D. Bartram (Eds), *Organisational Effectiveness: The Role of Psychology* (pp. 227-257). Chichester: Wiley.

Lievens, F., Buyse, T. & Sackett, P.R. (2005). The Operational Validity of a Video-Based Situational Judgement Test for Medical College Admissions: Illustrating the Importance of Matching Predictor and Criterion Construct Domains, *Journal of Applied Psychology,* 90(3), 442-452.

Lindsay, A. (2000). *The Predictive Validity of the OPQ in Selecting Middle Managers to attend a Senior Management and Staff Training Course.* Unpublished Master's Thesis, University of Pretoria, Pretoria.

McDaniel, M.A. (2006). *Situational Judgement Tests: Construct Validity and Directions for Future Research.* Panel Member as Part of a Discussion at the 21st Annual Conference for the Society of Industrial and Organisational Psychology, Dallas, Texas.

McDaniel, M.A., Morgeson, F.P., Finnegan, E.B., Champion, M.A. & Braverman, E.P. (2001). Use of Situational Judgement Tests to Predict Job Performance: A Clarification of the Literature. *Journal of Applied Psychology,* 86, 730-740.

McDaniel, M.A. & Nguyen, N.T. (2001). Situational Judgement Tests: A Review of Practice and Constructs Assessed. *International Journal of Selection and Assessment, 9,* 103-113.

McDaniel, M.A. & Nguyen, N.T. (2003). Response Instructions and Racial Differences in a Situational Judgement Test. *Applied H.R.M. Research,* 8(1), 33-44.

Mead, A.D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458.

Motowildo, S.J., Dunette, M.D. & Carter, G.W. (1990). An Alternative Selection Procedure: The Low-Fidelity Simulation. *Journal of Applied Psychology,* 75, 640-647.

Motowildo, S.J., Hanson, M.A, & Crafts, J.L. (1997). Low Fidelity Simulations. In D.L. Whetzel & G.R. Wheaton (eds). *Applied Measurement Methods in Industrial Psychology* (pp.241-260). California: Davies-Black Publishing.

Motowildo, S.J. & Tippins, N. (1993). Further Studies of the Low-Fidelity Simulation in the Form of a Situational Inventory. *Journal of Occupational and Organisational Psychology,* 66, 337-344.

Mouton, J. & Marais, H.C. (1992). *Basic Concepts: Methodology of the Behavioural Sciences.* Pretoria: HSRC.

Muchinsky, P.M. (2000). *Psychology Applied to Work* (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Mullins, M.E. & Schmitt, N. (1998). Situational Judgement Testing: Will the Real Constructs Please Present Themselves? Paper presented at the 13th Annual Conference of the Society of Industrial Organisational Psychology, Dallas, TX.

Mumford, T.V., Van Iddekinge, C.H., Morgeson, F.P. & Campion, M.A. (2008). The Team Role Test: Development and Validation of a Team Role Knowledge Situational Judgement Test. *Journal of Applied Psychology,* 93(2), 250-267.

Neuman, W.L. (1997). *Social Research Methods: Qualitative and Quantitative Approaches* (3rd ed). London: Allyn & Bacon.

Nguyen, N.T. & McDaniel, M.A. (2003). Response Instructions and Racial Differences in a Situational Judgement Test. *Applied H.R.M. Research,* 8(1), 33-44.

Ones, D. & Viswesvaran, C. (1996). Bandwidth Fidelity dilemma in personality measurement for personnel selection. *Journal of Organisational Behaviour,* 17, 609-626.

Ployhart, R.E. (1999). *An Interactionist Approach to Assessing Personality in Work Contexts: Construct Validation of a Predictor of Customer Service Performance.* Unpublished Doctoral Dissertation. Michigan State University.

Ployhart, R.E. (2006). The Predictor Response Process Model. In J.A. Weekley & R.E. Ployhart (Eds), *Situational Judgement Tests: Theory, Measurement and Application* (pp.83-105). New Jersey: Lawrence Erlbaum Associates.

Ployhart, R.E. & Ehrhart, M.G. (2001). Effects of Response Instructions of the Criterion Related Validity, Construct Validity, and Reliability of Situational Judgement Tests. Paper presented at the Annual Conference Society for Industrial Organisational Psychology, San Diego, CA.

Ployhart, R.E. & Weekley, R.K. (19986). Situational Judgement: Some Suggestions for Future Science and Practice. In J.A. Weekley & R.E. Ployhart (Eds), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 345-350). New Jersey: Lawrence Erlbaum Associates.

Porr, W.B., Baughman, K.L., Ployhart, R.E. & McFarland, L.A. (2006). *Reactions to a Situational Judgement Test.* Poster presented at the 21st Annual Conference for the Society of Industrial and Organisational Psychology, Dallas, Texas.

Radhakrishna, R.B., Yoder, E.P. & Ewing, J.C. (2007). *Strategies for Linking Theoretical Framework and Research Types.* Proceedings of the 2007 AAAE Research Conference, 34, pp.692-694.

Rafilson, F. (1991). The Case for Validity Generalisation. *ERIC/TM Digest*, ED338699, Retrieved 26 November 11:00.

SHL. (2006). *Scenarios Profile Report.* Surrey: SHL Group Plc.

SHL. (2008a). *Reliability Study for the Critical Reasoning Test Battery.* Surrey: SHL Group Plc.

SHL. (2008b). *Test Administration and Occupational Testing Course Notes.* Surrey: SHL Group Plc.

Schmidt, F.L. & Hunter, J.E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin,* 124, 262-274.

Smith, K.C. & McDaniel, M.A. (1998). *Criterion and Construct Validity Evidence for a Situational Judgement Measure.* Paper presented at the 13th Annual Conference of the Society for Industrial and Organisational Psychology, Dallas, Texas.

South African Employment Equity Act 55 of 1998.

Stemler, S.E. & Sternberg, R.J. (2006). In J.A. Weekley & R.E. Ployhart (Eds), *Situational Judgement Tests: Theory, Measurement and Application* (pp. 107-131). New Jersey: Lawrence Erlbaum Associates.

Sternberg, R.J. & Hedlund, J. (2002). Practical Intelligence, g, and Work Psychology. *Human Performance*, 15 (1), 143 – 160.

Sternberg, R.J. & Wagner, R.K. (1986). *Practical Intelligence: Nature and Origins of Competence in the Everyday World.* Cambridge: Cambridge University Press.

Sternberg, R.J., Wagner, R.K., Williams, W.M. & Horvath, J.A. (1995). Testing Common Sense. *American Psychologist,* 50, 912-927.

Strahan, J., Fogarty, G.J. & Machin, A.M. (2005). Predicting Performance on a Situational Judgement Test: The Role of Communication Skills, Listening Skills, and Expertise. *Proceedings of the 40th Annual Conference of the Australian Psychological Society,* 323-327, Sydney, Australia.

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.

Thorndike, R.L. & Stein, S. (1937). An Evaluation of the Attempts to Measure Social Intelligence. *Psychological Bulletin,* 34, 275-285.

Trippe, D.M. (2002). *Construct Validity of SJTs.* Unpublished Master's Thesis, Virginia Polytechnic Institute and State University.

Wagner, R.K. & Sternberg, R.J. (1986). Practical Intelligence in Real World Pursuits: The Role of Tacit Knowledge. *Journal of Personality and Social Psychology,* 49, 436-458.

Weekley, J.A. & Jones, C. (1999). Further Studies of Situational Tests. *Personnel Psychology,* 52, 679-700.

Weekley, J.A. & Ployhart, R.E. (2005). Situational Judgement: 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 Judgement Tests: Theory, Measurement and Application* (pp. 1-10). New Jersey: Lawrence Erlbaum and Associates.

Wilson, P., Baron, H. & Borkowski, T. (1991). *Critical Reasoning Test Battery Manual and Users' Guide.* Surrey: SHL Group Plc.

www.humanassets.co.uk/pagecontent.asp, Retrieved March 2009.

Yukl, G. A. (2006). *Leadership in Organisations.* New Jersey: Prentice-Hall.

Appendix A: Biographical Data Form

SHL Biographical Data Form

TO BE COMPLETED BY ALL CANDIDTATES BEING ASSESSED

SURNAME | INITIALS

Important directions for completing the answer sheet
- Use only soft pencils
  Do not use ink or ballpoint pens
- Make heavy black marks that fill the circle completely

EXAMPLES

CORRECT: ● Ⓑ Ⓒ Ⓓ Ⓔ

WRONG: Ⓐ ✓ Ⓒ Ⓓ Ⓔ
WRONG: Ⓐ Ⓑ ✗ Ⓓ Ⓔ
WRONG: Ⓐ Ⓑ ◓ Ⓓ Ⓔ

- Fully erase any answers you wish to change
- Make no other marks on the answer sheet

Age, gender, educational level, home language and ethnic origin information are used for research and equity monitoring only.

TODAY DATE
DAY MONTH YEAR | AGE | GENDER

EDUCATIONAL LEVEL
Grade 10 / Std 8 or Below	◯
Grade 12 / Std 10	◯
Post Matric Certificate	◯
Degree / 3 year Diploma	◯
Post Graduate	◯

ETHNIC ORIGIN
African	◯
Coloured	◯
Indian	◯
White	◯
Other	◯

HOME LANGUAGE
(CHOOSE ONE ONLY)
Afrikaans	◯
English	◯
Ndebele	◯
North Sotho	◯
South Sotho	◯
Swazi	◯
Tsonga	◯
Tswana	◯
Venda	◯
Xhosa	◯
Zulu	◯
Other	◯

IDENTIFICATION NO.

PLACE OF RESIDENCE
| City / Town; Urban | ◯ |
| Rural; Non-urban | ◯ |

ENGLISH IS MY ...... LANGUAGE
1st	◯
2nd	◯
3rd	◯
4th	◯
5th	◯

The reproduction of this answer sheet by a duplicating machine, photocopying process, or any other method, is breaking the copyright law.

JOB TITLE 1 APPLYING FOR

JOB TITLE 2 APPLYING FOR

CURRENT JOB TITLE

NUMBER OF
YEARS IN
CURRENT
POSITION

OTHER

SHL

PO Box 1305, Groenkloof
Pretoria, 0027
South Africa
Tel: +27 (0) 12 346-5330
Fax: +27 (0) 12 346-5333
Email: info@shl.co.za
Internet: www.shl.co.za

© SHL Group plc. 2003

Biographical data form SA Version 28/03/03

REF: AS0073

BPC 44905 24/04/03

Appendix B: Consent Form

CONSENT FORM

I hereby state that I wish to undergo the assessment process, and that I am aware and understand that although I am not guaranteed a placement by participating in this process, I will be considered for roles applied for in [the company]. I understand and give consent that the results of this assessment may be used by suitably qualified professionals in the decision-making process, and acknowledge my awareness that restricted access to assessment results will be granted to the selected decision-making panel. All assessment results will be treated in the strictest confidence by this panel.

	Yes, I agree to the above		No, I do not wish to undergo the assessment process

(Please tick the appropriate box)

Should I not be successful in this round, I give consent to the use of the assessment results for other positions within [the company], where these particular tests or questionnaires are relevant on the basis of the job requirements, for a period of up to one year from this date. Should any additional or alternative tests or questionnaires be required, I will be given the opportunity to partake in such an assessment in order to be given a fair opportunity.

	Yes, I agree to the above		No, I do not wish to undergo the assessment process

(Please tick the appropriate box)

I claim a disability	YES	NO

(Please tick the appropriate box)

I also give consent to the use of assessment results for research purposes, providing identifying information is withheld

	Yes, I agree to the above		No, I do not wish the results to be used for research purposes

(Please tick the appropriate box)

If you answered "Yes" to the above statements, please provide details below and send to SHL at least 48 hours (or 4 working days) before your assessment.

Signature: _____

Name and surname (printed): _____

Date: _____