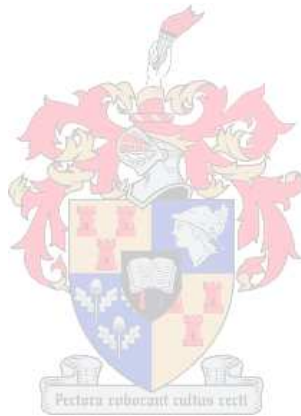


**THE DEVELOPMENT OF AN EXPERIMENTAL INTEGRITY INSTRUMENT FOR
VARIOUS CULTURAL GROUPS AS CONCEPTUALISED FROM THE SOUTH
AFRICAN PERSONALITY INVENTORY (SAPI) PROJECT**

By

Megon Lötter



Thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF COMMERCE

In the field of

INDUSTRIAL PSYCHOLOGY

at

THE UNIVERSITY OF STELLENBOSCH

SUPERVISOR: PROF DEON MEIRING

September 2010

DECLARATION

I, the undersigned, hereby declare that this thesis is my own original work and that all sources have been accurately reported and acknowledged, and that this document has not previously, in its entirety nor in part, been submitted at any university in order to obtain an academic qualification.



Megon Lötter

1 September 2010

ABSTRACT

Megon Lötter, M Comm (Stellenbosch University)

**THE DEVELOPMENT OF AN EXPERIMENTAL INTEGRITY INSTRUMENT FOR
VARIOUS CULTURAL GROUPS AS CONCEPTUALISED FROM THE SOUTH
AFRICAN PERSONALITY INVENTORY (SAPI) PROJECT**

Supervisor: Prof Deon Meiring

An urgent need exists for the development of a locally, multicultural personality instrument for South Africa. The South African Personality Inventory (SAPI) project was launched with the specific aim of developing a comprehensive personality questionnaire for all eleven South African language groups that covers all major aspects of personality deemed relevant in the South African context.

The current study focused on developing an experimental instrument for the integrity cluster, one of the 9 SAPI clusters. This study forms part of the second phase of the SAPI project (quantitative phase). In this phase the experimental integrity instrument was administered to a sample of police reservists of the South African Police Service (SAPS; $N = 1023$).

Findings revealed that certain items should be removed (30 of 132 items were removed). The first-order factor analysis confirmed one factor per facet that should be retained (specifically: Honest, Loyal, Pretending, Responsible, Trustworthy, Truthful, and Fair) for most of the facets. The exceptions were the Morally Conscious facet where two factors emerged and the Discriminative facet where no significant factor emerged. With the exception of the Discriminative facet (low reliability coefficient) and the Fair facet (average reliability coefficient) all the facets demonstrated acceptable levels of reliability.

The study concluded that the underlying dimensionality of the data confirmed the structure of the integrity cluster and the experimental integrity instrument. This first draft instrument can thus be applied to multi-cultural groups.

Keywords: personality, cross-cultural groups, South African Personality Inventory (SAPI) project, integrity, equivalence, bias, reliability, South Africa

OPSOMMING

Megon Lötter, M Comm (Universiteit van Stellenbosch)

DIE ONTWIKKELING VAN 'N EKSPERIMENTELE INTEGRITEITSINSTRUMENT VIR VERSKEIE KULTURELE GROEPE SOOS GEKONSEPTUALISEER VANUIT DIE SUID-AFRIKAANSE PERSOONLIKHEIDSINSTRUMENT (SAPI) PROJEK

Studieleier: Prof Deon Meiring

Suid-Afrika het 'n dringende behoefte aan die ontwikkeling van plaaslike, multikulturele persoonlikheidstoetse. Die Suid-Afrikaanse Persoonlikheidsinstrument projek is geloods met die spesifieke doel om 'n volledige persoonlikheidsvraelys vir al elf Suid-Afrikaanse taalgroepe te ontwikkel, wat betrekking het op alle belangrike aspekte van persoonlikheid wat relevant is tot die Suid-Afrikaanse konteks.

Die fokus van die huidige studie was om 'n eksperimentele instrument op een van die SAPI se 9-kluster modelle te ontwikkel, naamlik die integriteitskluster. Dit vorm deel van die tweede fase van die SAPI projek (kwantitatiewe fase) waar die eksperimentele integriteitsinstrument op 'n steekproef van intreevlak polisie-kandidate van die Suid-Afrikaanse Polisie (SAPD), afgelê is ($N = 1023$).

Die bevindinge het getoon dat sekere items verwyder moes word (30 van 132 items is verwyder). Die eerste-orde faktor analise het bevestig dat een faktor per faset behou moet word (meer spesifiek, Eerlik, Lojal, Skynheilig, Verantwoordelik, Betroubaar, Waarheid, en Regverdig). Twee faktore het na vore gekom in die Moreel Bewuste faset en geen betekenisvolle faktor was verkry vir die Diskriminerende faset nie. Alle fasette het aanvaarbare vlakke van betroubaarheid geopenbaar, behalwe vir 'n lae betroubaarheidskoeffisiënt vir die Diskriminerende faset, en 'n gemiddelde betroubaarheidskoeffisiënt vir die Regverdigheidsfaset.

Die gevolgtrekking was dat die onderliggende dimensionaliteit van die data die struktuur van die integriteitskluster en die eksperimentele integriteitsinstrument, bevestig het, asook dat die eerste proef-instrument toegepas kan word op multikulturele groepe.

Sleutelwoorde: persoonlikheid, kruiskulturele groepe, Suid-Afrikaanse Persoonliheidsinstrument (SAPI) projek, integriteit, ekwivalensie, bevooroordeeldheid, betroubaarheid, Suid-Afrika

ACKNOWLEDGEMENTS

All glory, honour and praise to my Almighty God for blessing me with the opportunity to have come this far, and for giving me strength and patience while guiding me through the completion of my thesis.

I would also like to thank the following individuals:

- The three people dearest to my heart, my parents, Nico and Mignon and my brother, Ludolf, for your enduring love and moral support.
- Prof. Deon Meiring, my supervisor, for your guidance and encouragement and for being an excellent mentor. I am glad that I did not undertake this journey any differently or with anyone else. Also, thank you for all the arrangements and assistance with regards to the SAPS sample group.
- The SAPI team for your guidance and support: Prof. Fons van de Vijver, Prof. Deon de Bruin, Prof. Ian Rothmann, Prof. Michael Temane, Velichko Valchev, Dr. Alewyn Nel, Dr. Carin Marais, Dr. Karina de Bruin, Byron Adams, Elke Chrystal, Charnelle Janse van Rensburg, Talitha Oosthuizen, Antoinette Labuschagne, Angelique Flattery, and Petrus van der Linde.
- The SAPI project's SANPAD financial assistance.
- Carol Saccaggi for the final editing.
- My family and friends for their love, support and encouragement.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
OPSOMMING.....	iv
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
CHAPTER ONE: INTRODUCTION AND OBJECTIVES OF THE CURRENT STUDY	
1.1 INTRODUCTION.....	1
1.2 RESEARCH OBJECTIVES.....	4
1.2.1 General Objective.....	5
1.2.2 Specific Objectives.....	5
1.3 CHAPTER OUTLINE.....	5
CHAPTER TWO: LITERATURE REVIEW	
2.1 INTRODUCTION.....	7
2.2 CROSS-CULTURAL PERSONALITY ASSESSMENT.....	7
2.3 THE CONCEPT OF INTEGRITY.....	10
2.3.1 Defining Integrity.....	10
2.3.2 Models of Integrity Measurement and the SAPI.....	12
2.4 MEASUREMENT OF INTEGRITY.....	17
2.5 DEVELOPMENT OF A MEASURING INSTRUMENT.....	21
2.5.1 Faking and Social Desirability.....	27
2.6 SUMMARY.....	32
CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY	
3.1 INTRODUCTION.....	33
3.2 RESEARCH DESIGN.....	33
3.3 PARTICIPANTS AND PROCEDURE.....	34
3.3.1 Pilot Study.....	34

3.3.2	Experimental Integrity Sample.....	36
3.4	MEASURING INSTRUMENT.....	38
3.4.1	Phase 1: Aim of the Instrument and Construct Domain.....	38
3.4.2	Phase 2: Writing of the Items.....	39
3.4.3	Phase 3: Inclusion of Other Items.....	40
3.4.4	Phase 4: Development of Social Desirability Items.....	40
3.4.5	Phase 5: Assembling of the Experimental Integrity Instrument.....	42
3.5	STATISTICAL ANALYSIS.....	42
3.6	CONCLUSION.....	44

CHAPTER FOUR: RESEARCH RESULTS

4.1	INTRODUCTION.....	45
4.2	DATA PREPERATION.....	45
4.2.1	Unexpected Responses.....	45
4.2.2	Missing Values.....	45
4.2.3	Descriptive Statistics in terms of Skewness and Kurtosis.....	46
4.3	HIERARCHICAL STRUCTURE OF THE DATA.....	49
4.4	ITEM CORRELATIONS WITH FACETS.....	49
4.5	FIRST-ORDER FACTOR ANALYSIS.....	52
4.5.1	The Factor Analysis Results of the Honest Facet.....	53
4.5.2	The Factor Analysis Results of the Loyal Facet.....	54
4.5.3	The Factor Analysis Results of the Morally Conscious Facet.....	55
4.5.4	The Factor Analysis Results of the Pretending Facet.....	57
4.5.5	The Factor Analysis Results of the Responsible Facet.....	58
4.5.6	The Factor Analysis Results of the Trustworthy Facet.....	59
4.5.7	The Factor Analysis Results of the Truthful Facet.....	60
4.5.8	The Factor Analysis Results of the Discriminative Facet.....	61
4.5.9	The Factor Analysis Results of the Fair Facet.....	62
4.5.10	The Factor Analysis Results of the Integrity Cluster.....	63
4.6	RELIABILITY.....	63
4.7	SECOND-ORDER FACTOR ANALYSIS.....	64
4.8	CONCLUSION.....	64

CHAPTER FIVE: DISCUSSION AND RESULTS

5.1 INTRODUCTION..... 66

5.2 SUMMARY OF FINDINGS..... 69

5.3 FUTURE RECOMMENDATIONS..... 71

5.4 PRACTICAL IMPLICATIONS..... 72

5.5 CONCLUDING REMARKS.....73

REFERENCES.....74

LIST OF TABLES

Table Number	Page
Table 2.1	Similarities between the SAPIs Integrity Cluster and Other Personality Measures..... 17
Table 3.1	Demographics of the Pilot Sample ($N = 174$).....35
Table 3.2	Demographics of the Experimental Integrity Instrument Sample ($N = 1023$).....37
Table 3.3	Definitions of the Integrity Cluster.....38
Table 3.4	Additional Personality Instrument Items Transformed..... 40
Table 3.5	MCSD Items Modified..... 41
Table 3.6	Distribution of Integrity Items.....42
Table 4.1	Descriptive Statistics of the 132 Items ($N = 1023$).....47
Table 4.2	Item Correlations with Facets ($N = 1023$)..... 50
Table 4.3	Eigenvalues for the Honest Facet ($N = 1023$)..... 53
Table 4.4	Eigenvalues for the Loyal Facet ($N = 1023$)..... 54
Table 4.5	Eigenvalues for the Morally Conscious Facet ($N = 1023$)..... 55
Table 4.6	Component Loadings of Items Comprising the Morally Conscious Facet, Excluding Items 28 and 35..... 56
Table 4.7	Eigenvalues for the Pretending Facet ($N = 1023$)..... 57
Table 4.8	Eigenvalues for the Responsible Facet ($N = 1023$)..... 58
Table 4.9	Eigenvalues for the Trustworthy Facet ($N = 1023$).....59
Table 4.10	Eigenvalues for the Truthful Facet ($N = 1023$)..... 60
Table 4.11	Eigenvalues for the Discriminative Facet ($N = 1023$)..... 61
Table 4.12	Eigenvalues for the Fair Facet ($N = 1023$)..... 62
Table 4.13	Eigenvalues for the Integrity Cluster ($N = 1023$)..... 63
Table 4.14	Descriptive Statistics, Skewness, Kurtosis, and Cronbach Alpha Coefficients ($N = 1023$)..... 64

LIST OF FIGURES

Figure Number		Page
Figure 4.1	Scree-plot for the Honest facet.....	53
Figure 4.2	Scree-plot for the Loyal facet.....	54
Figure 4.3	Scree-plot for the Morally Conscious facet.....	55
Figure 4.4	Scree-plot for the Pretending facet.....	57
Figure 4.5	Scree-plot for the Responsible facet.....	58
Figure 4.6	Scree-plot for the Trustworthy facet.....	59
Figure 4.7	Scree-plot for the Truthful facet.....	60
Figure 4.8	Scree-plot for the Discriminative facet.....	61
Figure 4.9	Scree-plot for the Fair facet.....	62
Figure 4.10	Scree-plot and parallel analysis plot for the Integrity cluster.....	63

CHAPTER ONE

INTRODUCTION AND OBJECTIVES OF THE CURRENT STUDY

1.1 INTRODUCTION

The search for psychologically universal generalisable traits has been one of the major goals of cross-cultural psychology (Schnabel, Banse, & Asendorpf, 2006). Historically the development of psychological tests within South Africa has been largely shaped by the Apartheid political dispensation before the new Government of National Unity in 1994 (Foxcroft, 1997, 2004). From 1920-1960 these measures were either developed for specific categories of White people or adapted from international measures. The majority of these instruments were from typical Anglo-Saxon countries (Meiring, Van de Vijver, Rothmann, & Barrick, 2005). This created problems related to the use of such measures for different ethnic groups. Such difficulties include the influence of cultural, environmental and temperamental factors as well as the cultural appropriateness of the measures (Biesheuvel, 1943). Therefore, this specific time period was characterised by focussing on standardised measures for White people only and misusing measures that were standardised for a specific group by administering it onto another group without investigating issues such as bias and the inappropriateness for the second group (Meiring, 2007). Also, conclusions were made about differences between groups without considering the impact of cultural, socioeconomic, environmental, and educational factors on test performance (Meiring, 2007).

Following the 1994 democratic elections the country instituted a new constitution. Stronger demands for the cultural appropriateness of psychological tests culminated in the promulgation of the Employment Equity Act 55 of 1998, Section 8 (Government Gazette, 1998), which stipulates the following:

Psychological testing and other similar assessments are prohibited unless the test or assessment being used (a) has been scientifically shown to be valid and reliable, (b) can be applied fairly to all employees; and (c) is not biased against any employee or group. (p. 9)

The onus of proof has shifted to psychologists using these instruments, who now have to indicate that they adhere to the regulations of the Employment Equity Act (Taylor, 2008). The transformation of South African society, the integration of schools, universities, the workplace

and society in general since 1994 has created an urgent need for local measuring instruments that not only meet the Employment Equity Act requirements, but that can also be used for all the cultural and language groups in South Africa (Taylor, 2008). In addition, test takers have learned to voice their rights, since fairness is an imperative standard to all cultural (language) groups. Psychology as a profession is being held accountable for the (im)proper usage of instruments (Robins, Fraley, & Krueger, 2007). The Employment Equity Act creates a daunting task for psychology as a profession, as it loads the burden of proof onto the profession and the psychologists using psychological measures.

The South African landscape is further complicated by the current skills shortage, which makes it imperative to retain staff (Claassen, 1997). In order to maintain the workforce, it is important to ensure that the person who is being employed will remain in the same position for a certain period of time. Personality assessment can thus assist in the pursuit of workforce retention. Hough and Oswald (2008) cite various researchers' investigations of personality and performance in the workplace that predict overall job performance. Investigations concerning personality and counterproductive work behaviours are also relevant, such as personality variables that predict counterproductive work behaviour (CWB) and personality-based integrity tests that predict CWB and absenteeism (Hough & Oswald, 2008).

According to Ones and Viswesvaran (2001), personality predictors used in personnel selection can be divided into two categories, namely measures of normal adult personality and measures of personality at work. Measures of normal adult personality are measures that accurately describe individual differences in personality and are used for personnel screening and selection (Ones & Viswesvaran, 2001). The measures of personality at work accurately predict individual differences in work behaviours of interest. These measures include integrity tests, violence scales, drug and alcohol scales, sales potential scales and managerial potential scales (Ones & Viswesvaran, 2001).

Integrity tests, also referred to as honesty scales, are considered prototypical criterion-focused occupational personality scales (Ones & Viswesvaran, 2001). Integrity tests are specifically developed to assess the integrity, honesty and dependability of applicants, thereby facilitating the prediction of job performance and counterproductive behaviours on the job, including theft, disciplinary problems, absenteeism and future on the job dishonest behaviours (Ones, Viswesvaran, & Schmidt, 2003).

Psychological tests have been translated and adapted into different languages worldwide (Hambleton, 2001; Schmidt, Kihm, & Robie, 2000; Van de Vijver & Hambleton, 1996). Personality tests that are developed in a specific country and transported to other countries are widely known as the 'imposed etic' strategy (Berry, 1969). There is a growing focus on the generalisability and efficiency of these test instruments. The use of such instruments is also being questioned (Butcher, 2006). Sue (1983) criticises the predominance of the etic approach in psychology because it disadvantages the emic approach. According to Sue (1983), the etic approach accentuates the universal 'core similarities' in all human beings universally, whereas the emic approach utilises 'a cultural-specific orientation' relevant to the local context.

Nel (2008) states that there are more etic studies than emic studies in personality psychology literature. However, the combined etic-emic approach is the most popular in current personality research. The development of the Chinese Personality Assessment Inventory (CPAI) is a rare example of an indigenous measure receiving international attention (Cheung et al., 1996). The CPAI is a combined etic-emic measure that includes both universal and indigenous constructs by adopting the convergence approach in cross-cultural psychology (Van de Vijver & Leung, 1997). This indigenous instrument covers personality characteristics for normal samples but includes a diagnostic assessment of Chinese people based on previous collaboration on the translation of the Chinese MMPI (Minnesota Multiphasic Personality Inventory) (Cheung et al., 1996). The personality constructs included in the CPAI were "derived from personality adjectives or person descriptions of everyday life" (Nel, 2008; p. 27). These adjectives and descriptions included Chinese novels, Chinese proverbs, reviews of Chinese literature and the completion of self/other description surveys by both professional and ordinary people.

Similarly in 2005 the South African Personality Inventory (SAPI) project was undertaken (Meiring, 2007). The project aims to develop a single, unified personality inventory for South Africa that takes into consideration both universal and unique personality factors across the eleven language groups (Meiring, 2007). One of the main objectives of the first phase of the SAPI project was to derive authentic, relevant and accurate personality-descriptive terms from each of the eleven official languages (Nel, 2008). These descriptive terms were derived using information collected from semi-structured interviews that were conducted in each of the eleven language groups (Nel, 2008). Participants were asked to give a description of a particular person(s) in behavioural or trait terms (Nel, 2008). These descriptive terms were then analysed in order to understand the inherent labels associated with certain behaviour and traits in a specific language group (Nel, 2008). Nel (2008) identified more than 50 000 personality-

descriptive terms. These were then reduced to 190 personality dimensions by means of cluster analysis. The analysis included the grouping of synonyms and antonyms, the use of dictionaries, literature and knowledge concerning the content (Nel, 2008). The 190 dimensions were also divided into those that are common (shared by all eleven language groups), semi-common (shared by seven to ten of the language groups), semi-specific (shared by two to six of the language groups) and language-specific (unique to a particular language group) (Nel, 2008). Nel (2008) discovered that 79 dimensions were common, 70 dimensions were semi-common, 32 dimensions were semi-specific and only 11 dimensions were language-specific. Most of the personality dimensions seem to be shared by the language groups rather than being unique to a specific language group (Nel, 2008). The 190 dimensions were then further subjected to cluster analysis in order to build an indigenous personality structure (Nel, 2008). Clustering produced 37 sub-clusters, which consisted of two to ten dimensions and nine overall clusters consisting of two to six sub-clusters (Nel, 2008). These nine clusters were labelled Extraversion, Soft-heartedness, Conscientiousness, Emotional stability, Intellect, Openness, Integrity, Relationship harmony and Facilitating (Nel, 2008). Although universal aspects are evident within these clusters many indigenous aspects also exist within the structure (Nel, 2008).

According to Meiring (2007), the first phase of the SAPI project used an indigenous approach to identify culturally and linguistically adequate personality descriptive terms for all eleven languages. This part of the project is completed and the work on the second phase has begun. In the second phase the descriptive terms are to be converted into a validated quantitative inventory (Meiring, 2007). The focus of the current study was on developing a comprehensive inventory for Integrity, one of the nine identified clusters. The Integrity cluster, as conceptualised in the first phase of the SAPI project, consists of facets labelled honesty, integrity, loyalty, morally conscious, pretending, responsible, trustworthy, trustful, discriminative and fair. The experimental instrument consisted of a pool of scales tapping integrity items that was administered to a multi-cultural group. This study was explorative in nature and aimed to contribute to the larger quantitative phase of the SAPI project.

1.2 RESEARCH OBJECTIVES

The aim of the present study was to develop a measuring instrument for the personality construct of Integrity. Items were generated from the various sub-facets scales of the integrity instrument. In order to generate these items the sub-facet scales were first defined. A pilot

study was conducted using the experimental integrity instrument. The items of the integrity instrument were then analysed. Finally, a first draft instrument was developed that can be applied to multi-cultural groups.

The research objectives consisted of both a general objective and specific objectives.

1.2.1 General Objective

The general objective of this study was to develop an experimental integrity instrument for the integrity construct of the SAPI project.

1.2.2 Specific Objectives

- To define the sub-facets' scales of the integrity instrument clearly and develop an experimental item pool for the various facets of the integrity instrument;
- To draft an experimental integrity instrument with an appropriate response format scale;
- To conduct a pilot study with the experimental integrity instrument;
- To follow a hierarchical approach in analysing the data with the aim of examining the underlying dimensionality of the data to confirm the structure of the integrity cluster; and
- To develop a first draft instrument that can be applied to multi-cultural groups.

1.3 CHAPTER OUTLINE

Chapter two provides an extensive review of the literature on cross-cultural personality assessment. The chapter defines the concept of integrity, discusses models of integrity measurement and explores the conceptualisation of integrity within the SAPI project. The chapter also describes various integrity measures that are being utilised internationally and in South Africa. The last section deals with the development of a measuring instrument and concludes with a discussion of the concepts of faking and social desirability.

In light of the SAPI integrity structure model, the research design and methodological approach are discussed in chapter three. In addition, the composition and nature of the sample, as well as the development of the integrity instrument are described. The statistical analysis used is also explained. The results of the analysis are presented in chapter four.

Chapter five presents the discussion and conclusions of the study. In that chapter the limitations of the study are highlighted and recommendations are made for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter contains a review of the literature on cross-cultural personality assessment with specific attention given to the concept of integrity. Various models of integrity measurement are inspected and a comparison is drawn between these models and the conceptual models of integrity within the SAPI project. The final section of the chapter focuses on the process of developing a measuring instrument.

2.2 CROSS-CULTURAL PERSONALITY ASSESSMENT

Cross-cultural psychology involves the systematic study of the relationships between the cultural context of human development and the behaviours that become established in the selection of individuals growing up in a particular culture (Berry, Poortinga, Pandey, Dasen, Saraswathi, Segall, & Kagitcibasi, 1996). Van de Vijver (2002) refers to cross-cultural assessment as “all issues arising in the application of psychological instruments, either in a single country in the assessment of migrant groups, or in the assessment of individuals from at least two countries” (p. 547). Van de Vijver (2002) holds that it is essential that the tests used must demonstrate their appropriateness to all cultural groups involved.

Different theoretical perspectives are employed in the cross-cultural assessment literature. The three dominant perspectives towards assessment are known as the cross-cultural, cultural and indigenous approaches (Church, 2001). The cross-cultural approach typically involves: (a) comparisons of multiple cultures in the search for cultural universals or culture-specific factors amidst universals; (b) treatment of culture, or quantitative variables related to ecology and culture, as variables outside the individual that can be used to predict behaviour; (c) use of traditional and relatively context-free psychometric scales and questionnaires; (d) concern about the cross-cultural equivalence of constructs and measures; and (e) a focus on individual differences (Church, 2001). The cultural psychological approach involves: (a) a focus on contextual descriptions of psychological phenomenon in one or more cultures, with less emphasis on, or expectations of, culture universals; (b) a theoretical emphasis on the dynamic and mutually constitutive nature of culture and psychological functioning; (c) an emphasis on qualitative, ethnographic, and interpretive research methods; and (d) a de-emphasis on

individual differences (Church, 2001). Finally, the indigenous approach focuses on the need to formulate theory, constructs, and methods that reflect indigenous cultural context (Church, 2001).

Church (2001) says that research regarding imported personality instruments generally investigates the universality of personality dimensions, examines the nomological networks of personality constructs across cultures or compares trait levels across cultures. In this regard the model that has received the most attention in contemporary personality psychology is the Five-Factor Model (FFM) (De Raad & Perugini, 2002; McCrae & John 1992; Paunonen & Ashton, 1998). The FFM is a hierarchical model of trait structure in which relatively narrow and specific characteristics are organised in five broad factors labelled Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness to Experience.

De Raad and Peabody (2005) recently contested the consistency of the FFM across cultures. A re-examination of the five-factor analytic solutions of six European studies found that the five factor results did not always support the FFM (De Raad & Peabody, 2005). The researchers' re-analysis of the existing data suggested that a three-factor model (Extraversion, Agreeableness, and Conscientiousness) provided a more appropriate fit (De Raad & Peabody, 2005). In a more recent study, De Raad et al. (2010) confirmed that only 3 personality description factors are replicated across many different languages (i.e. 14 trait taxonomies from 12 different languages) if they are independently derived by a psycholexical approach. These findings suggest that only the first three factors of the Big Five are robustly replicated. This suggests substantial universality of a Big Three personality factor model (De Raad et al., 2010).

Finding personality traits that are specific to a given culture creates further difficulties. In order to identify the major dimensions of personality a representative set of personality variables must be obtained (Lin & Church, 2004). Thus, researchers aim to find the most important elements of the language used in the targeted culture (Lin & Church, 2004). Church (2000) states that studies that search for indigenous dimensions first, rather than imposing an existing structure from outside the culture, would result in more persuasive evidence of cross-cultural comparability. Such studies would also provide the best opportunity for culture-unique dimensions to be identified (Church, 2000).

However, in the early stages of the development of psychometric testing in Chinese societies, Cheung and Leung (1998) explained the need to adapt an instrument from another culture. The adaptation and validation of well-established instruments is able to satisfy the needs of practitioners efficiently before the availability of indigenous instruments (Cheung & Leung, 1998). The advantage of translating and adapting well-established Western instruments lies in the wealth of evidence gathered to support the conceptual and psychometric properties of these instruments (Cheung & Leung, 1998). The utilisation of a common and equivalent instrument also allows for cross-cultural comparisons on the construct domains measured by the instrument (Cheung & Leung, 1998). However, Cheung, Van de Vijver and Leong (In Press) highlight the fact that even imported Western measures that have been well adapted and that have demonstrated cross-cultural equivalence may still impose Western personality constructs that are assumed to be universal but are not relevant to the local context.

Increased experience related to the use of translated and adapted tests has highlighted the limitations of the imposed-etic approach (Cheung et al., 1996). Consistent cross-cultural differences in the responses to certain construct domains have raised doubts regarding the functional or conceptual equivalence of those domains to the new culture (Cheung & Leung, 1998). This has resulted in researchers applying emic methods, which rely on native languages, cultural informants, or local psychological literatures to identify salient indigenous personality concepts (Church, 2008) and develop an indigenous taxonomy of person descriptions (Cheung et al., In Press). These indigenous constructs enrich the understanding of the cultural meaning of personality patterns and increase the prediction of outcomes in the local contexts (Cheung et al., In Press). Indigenous approaches can be used to identify culturally variable personality dimensions that are missing in the etic measures as well as to identify indicators of universal traits that are culture specific (Cheung, et al., In Press). This suggests that cross-cultural and indigenous studies of personality are complementary and address different aspects of personality assessment (Cheung et al., In Press).

Cheung et al. (1996) highlight the importance of developing an inventory for individuals from a specific culture that includes both major culture-specific personality domains and culture-comparable personality constructs. This personality inventory would provide a reliable and valid assessment instrument for people of that culture. Indigenous instruments would yield different theories regarding the broader area of local reality about personality (Yang & Bond, 1990). The CPAI, an indigenous personality measure, challenges universal theories of personality such as the FFM (Cheung et al., 1996).

A combined etic-emic approach is involved in the development of indigenous personality measures. In this process culturally relevant concepts are collected to identify cross-cultural universals as well as culturally unique dimensions (Cheung & Leung, 1998). Cheung and Leung (1998) point out that the culturally unique dimensions overcome the deficiencies of translated instruments, which often lack functional and conceptual validity. In the combined etic-emic approach to cross-cultural personality studies culturally relevant items are generated to assess dimensions of emic concepts for individual cultures (Cheung & Leung, 1998). These emic dimensions are then examined across cultures by joining the factor structures of the combined item pool from the individual cultures. Dimensions emerging in all cultures represent etic dimensions (shared, universal), whereas dimensions emerging in only one culture represent emic dimensions (non-shared, culture-specific) (Cheung & Leung, 1998; Cheung, et al., In Press).

Indigenous instruments are necessary in order to attain culture specific personality domains and to avoid omitting important emic constructs when translating imported measures. In the SAPI project a combined etic-emic approach (convergence approach) was followed. This involved collecting culturally relevant concepts in order to identify cross-cultural universals as well as culturally unique dimensions. This approach allowed the SAPI project to be sensitive to the South African indigenous cultural context. It also allows for cross-cultural comparison of the meaning of emic or imposed etic traits, extends the interpretation of indigenous traits in a broader cultural context, and accommodates the different language groups.

2.3 THE CONCEPT OF INTEGRITY

2.3.1 Defining Integrity

Integrity has been defined in various ways and is still a subject of debate (Barry & Stephens, 1998). According to Yukl (2006), the most basic definition of integrity “emphasizes honesty and consistency between a person’s values and behavior” (p. 421). Critics (Becker, 1998) argue that this definition is insufficient, since the values must be moral and the behaviour must be ethical. According to these critics, integrity means that: “a person’s behavior is consistent with a set of justifiable moral principles” (Yukl, 2006, p. 421). This definition is hampered by the difficulty of finding consensus about justifiable moral principles (Yukl, 2006). The existence of values and principles that differ across cultures makes this definition problematic (Yukl, 2006). Examples of morally justifiable behaviours could include: “following the same rules and

standards applied to others, being honest and candid when providing information or answering questions, keeping promises and commitments, and acknowledging responsibility for mistakes while also seeking to correct them” (Yukl, 2006, p. 421).

Taylor and Gaita (1981) also claim that there are different ways of characterising a person of integrity. These characterisations include being honest, upright and loyal. Types of behaviour that conform to the socially accepted moral code help to explain the concept of integrity (Taylor & Gaita, 1981). Examples include not telling lies, keeping promises and not cheating in business. A person of integrity should do what they think is right and not always obey accepted codes, regardless of whether this agrees with more conventional views (Taylor & Gaita, 1981). A person of integrity will stick to their personal evolved principles and will be true to the standards they have come to accept (Taylor & Gaita, 1981). These principles are the focus of the person’s honesty and loyalty, and the individual’s own nature will determine his or her behaviour in a social context (Taylor & Gaita, 1981).

According to McFall (1987), integrity is “a complex concept with alliances to conventional standards of morality – especially those of truth telling, honesty, and fairness – as well as to personal ideals that may conflict with such standards” (p. 5). Internal coherence is an important dimension of integrity and is addressed in several forms of coherence (McFall, 1987). In general, personal integrity requires that an individual (1) subscribes to some consistent set of principles or commitments and (2), in the face of temptation or challenge, (3) upholds these principles or commitments (4) for what the individual takes to be the right reasons (McFall, 1987).

McFall (1987) refers to a person of integrity as one who is willing to face the consequences of their convictions, even when this is difficult. Similarly, a single-minded pursuit of approval or wealth is consistent with integrity (McFall, 1987). McFall (1987) also refers to the “Olaf Principle” from the “conscientious object-or” (p. 11) of e. e. Cummings’s poem about Olaf. This principle requires that some commitments be unconditional (McFall, 1987).

Researchers have different views regarding integrity. These differences could result from individuals’ own unique viewpoints and beliefs stemming from their personality, upbringing, background and culture. However, the definitions discussed suggest that integrity is consistently associated with honesty, loyalty, telling the truth, consistency in behaviour, fairness, being morally conscious, taking responsibility and keeping promises.

2.3.2 Models of Integrity Measurement and the SAPI

Some similarities exist between the SAPI structure and personality models found in literature, more specifically the Big Five (FFM), Big Seven Model, the CPAI, and the HEXACO (Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience; Nel, 2008). These personality models are all addressed in this section. However, the main focus is given to the HEXACO model as this model corresponds most strongly to the SAPI's Integrity model.

Recent cross-cultural evidence from the HEXACO model suggests a six-dimensional framework derived from lexical studies of personality structure (Ashton & Lee, 2005). Ashton, Lee, Marcus and De Vries (2007) summarise the identity of the six HEXACO factors. First, the content of the factors labelled Extraversion (which includes sociability, liveliness versus shyness and passivity) and Conscientiousness (organisation, discipline versus laziness, sloppiness, etc.) is similar to that of the FFM factors with the same labels (Ashton et al., 2007). The Openness to Experience factor (which includes inquisitiveness, creativity versus conventionality and unimaginativeness) is broadly similar to the factor with the sample label in the FFM. It is also similar to some variants of the Intellect/Imagination/Unconventionality factor obtained in lexical studies (Ashton et al., 2007). The remaining three HEXACO factors show some substantial departures from the FFM (Ashton et al., 2007). Emotionality (which includes anxiety, sentimentality versus independence and fearlessness) and Agreeableness (for example, patience, gentleness versus stubbornness, ill-temper) represent rotated variants of the Big Five factors Neuroticism (i.e. low Emotional Stability) and Agreeableness. Thus, the anger-related content of Neuroticism is associated with low Agreeableness of the HEXACO (Ashton et al., 2007). Finally, Ashton et al. (2007) point out that Honesty-Humility (for example, sincerity, fairness, modesty versus greed, pretentiousness, slyness, hypocrisy) involves content that is only partly associated with the Big Five factors. The Big Five Agreeableness factor includes content related to sympathy and soft-heartedness, which is treated within the HEXACO framework as a blend of Honesty-Humility, Agreeableness and (to some extent) Emotionality (Ashton et al., 2007). Adjectives representing this content (e.g. sympathetic, soft-hearted, helpful and generous) have been found to shift their locations across these factors in different investigations (Ashton et al., 2007).

Nel (2008) found good correspondence between the Integrity cluster of the SAPI and the HEXACO's Honesty-Humility factor. Honesty-Humility includes almost all of the descriptive

facets that are clustered in the Integrity cluster (Nel, 2008). These facets include sincerity, fairness and modesty (Lee & Ashton, 2004). However, the Honest-Humility factor also includes facets such as greedy, boastful and pompous (Ashton & Lee, 2007), which are not included in the Integrity cluster. The Honesty-Humility factor differs from the Integrity cluster in that Integrity puts emphasis on the moral soundness of a person and correct conduct according to certain values that modify behaviour, whereas Honesty-Humility additionally incorporates the social conduct according to an individual's set impulses or drives (this could include being greedy or boastful) (Nel, 2008).

According to Ashton and Lee (2008), the FFM Agreeableness construct incorporates some traits associated with the Honesty-Humility factor that are absent from classic Big Five Agreeableness factor. These traits are assessed by two of the six 'facet' scales of NEO-PI-R Agreeableness, namely Straightforwardness and Modesty (Ashton & Lee, 2008). These two facets correlate more strongly with the HEXACO Personality Inventory's (HEXACO-PI) Honesty-Humility than with the classic Big Five's Agreeableness (Ashton & Lee, 2005).

The Honesty-Humility dimension of the HEXACO-PI also correlates strongly and negatively with existing personality constructs involving exploitation and entitlement (Lee, Ashton, Morrison, Cordery, & Dunlop, 2008). For example, Lee and Ashton (2005) found that HEXACO-PI Honesty-Humility correlates significantly with variables such as Primary Psychopathy, Machiavellianism and Narcissism. Honesty-Humility also correlates significantly with the Manipulativeness and Integrity scales of the Supernumerary Personality Inventory.

The relatively recent introduction of the HEXACO model of personality structure means that only a few studies have investigated relationships between organisational variables and the HEXACO personality variables (Lee et al., 2008). However, early investigations have yielded interesting findings with regard to the Honesty-Humility factor (Lee et al., 2008). A study by Lee, Ashton and De Vries (2005) compared the HEXACO model and the FFM's ability to predict organisational delinquency and overt integrity test scores. Lee et al. (2005) administered the short form of the HEXACO-PI and three different measures of the FFM to four different samples from three different countries (Australia, Canada and the Netherlands). The HEXACO model outperformed the FFM in predicting organisational delinquency and overt integrity test scores across the different samples and the different measures of the Big Five/FFM (Lee et al., 2005). Subsequent analyses showed that Honesty-Humility was largely

responsible for the predictive superiority of the HEXACO-PI over the FFM measures (Lee et al., 2008).

The Big Seven personality model includes more dimensions than the FFM for the purpose of constructing a more relevant and effective model (Nel, 2008). Initially, the labels for this model were Positive Emotionality (similar to the Big Five's Extraversion), Negative Emotionality (similar to the Big Five's Neuroticism), Dependability (which was later changed to Conscientiousness; Simms, 2007), Agreeableness and Unconventionality (similar to the Big Five's Openness). Two additional constructs were identified which are predominantly self-evaluative, namely Positive Valence (describing and evaluating the self in positive terms) and Negative Valence (describing and evaluating the self in negative terms; Benet-Martínez & Waller, 1997).

Nel (2008) found little correspondence between the SAPI Integrity cluster and the factors of the Big Seven. However, some correspondence was found between Integrity and Negative Valence (Nel, 2008). This could be because Negative Valence involves some facets that are associated with a person not having integrity (specifically, treacherous; Benet-Martínez & Waller, 1997).

According to Nel (2008), there is only an arbitrary relationship between the SAPI's structure and that of the CPAI. The revised version of the CPAI is known as the CPAI-2 (Nel, 2008). The CPAI-2 consists of four main factors labelled Dependability, Interpersonal Relatedness, Social Potency and Accommodation. Clinical factors are also divided into two dimensions: Emotional Problems and Behavioural Problems (Cheung, Cheung, Leung, Ward, & Leong, 2003). The first factor, Dependability, seems to include facets that are widespread throughout the SAPI structure. Dependability consists of responsibility, emotionality, inferiority vs. self-acceptance, practical-mindedness, optimism vs. pessimism, meticulousness, face, internal vs. external locus of control and family orientation. Nel (2008) claims that responsibility seems to correspond with the SAPI's Integrity cluster, which consists of facets such as being trustworthy, honest and morally conscious. The Integrity cluster also includes truthful and responsible as facets, this reinforces the correspondence between Dependability and Integrity (Nel, 2008).

In the CPAI-2 the fourth factor of Accommodation corresponds with more than one cluster of the SAPI structure (Nel, 2008). Accommodation consists of defensiveness, graciousness vs.

meanness, interpersonal tolerance, self vs. social orientation, and veraciousness vs. slickness. Nel (2008) found that veraciousness vs. slickness corresponds with integrity, which encompasses a person's moral soundness. Veraciousness means to be honest and truthful (Nel, 2008). Slickness, on the other hand, has an immoral undertone that could be the misrepresentation of information in order to reap rewards (Nel, 2008). Such an individual would be seen as untrustworthy and pretending and these constructs form part of the Integrity cluster of the SAPI (Nel, 2008).

The first phase of the SAPI project (see chapter one) involved studying personality descriptive terms in the eleven official language groups in South Africa and integrating these terms into personality clusters. Within the SAPI project personality terms were labelled at different levels in a hierarchical manner. In this way 9 clusters, 37 sub-clusters, 190 facets and 550 sub-facets were formed (Nel, 2008).

The data in the first phase of the SAPI project led to the creation of the construct of Integrity based on the finding that moral consciousness and honour are extremely important (Nel, 2008). This construct includes both intrinsic and extrinsic values (Nel, 2008). The intrinsic values influence an individual's demonstration of integrity in relationships (showing loyalty and always telling the truth; Nel, 2008). These values must be embedded in a moral consciousness and sense of responsibility, which in the end leads a person to act out the external values (being loyal and telling the truth) (Nel, 2008). The Integrity cluster consists of two sub-clusters, namely integrity and fairness (Nel, 2008). The broad cluster of Integrity is defined as "the moral consciousness of a human being, characterised by being honest, loyal and dependable" (p. 125). The integrity sub-cluster is defined as: "The moral soundness of a person; the tendency to be trustworthy and to act in an honest, pure, and responsible manner" (Nel, 2008, p. 123). This sub-cluster includes eight facets, namely honesty, integrity, loyalty, morally consciousness, pretending, responsibility, trustworthiness and truthfulness). The fairness sub-cluster is defined as: "The inclination to accept and treat all people equally, rather than discriminate and favour some people over others" and consists of the facets discriminative and fairness (Nel, 2008, p. 123).

The SAPI Integrity cluster does not correspond exactly to any of the FFM clusters (Nel, 2008). However, some facets from the Integrity cluster do correspond with the FFM, for instance responsible and trustworthy (Nel, 2008). Szirmak and De Raad (1994) found Integrity to be the

fifth factor in the FFM in Hungary. The conventional fifth factor of Intellect or Openness was not replicated in their research in Hungary (Szirmak & De Raad, 1994).

A very good fit exists between the Conscientiousness cluster of the SAPI and the Conscientiousness factor of the HEXACO (Nel, 2008). Lexical studies have shown that the irresponsible facet within the Conscientiousness factor of the HEXACO loads onto the Conscientiousness cluster of the SAPI (Ashton, Lee, & Goldberg, 2004). In this context, irresponsible behaviour is viewed as the opposite of being disciplined and careful (Nel, 2008). In the SAPI's structure, the responsible facet features in the Integrity cluster (Nel, 2008). The argument is that a person's moral composition will modify the consequential outcome (more specifically, responsible vs. irresponsible behaviour; Nel, 2008).

With regards to employee selection, Becker (1998) states that some researchers have suggested that integrity tests measure conscientiousness. Within the framework of the Big Five theory of personality, Barrick and Mount (1991) identify some scholars who suggest that Conscientiousness "reflects dependability; that is, being careful, thorough, responsible, organized, and planful" and also "incorporates volitional variables, such as hardworking, achievement-oriented, and persevering" (p. 4). Collins and Schmidt (1993) argue that their integrity test measures conscientiousness "because as a group the most heavily loaded scales measure personal values, behavioural control, sense of duty and responsibility, and risk-taking behavior" (p. 308). Ones et al. (1993) suggest that integrity tests measure general conscientiousness, where "conscientiousness reflects such characteristics as dependability, carefulness, and responsibility" (p. 680).

Therefore, there are a number of correlations between the SAPI structure and other personality measures pertaining to the dimension of integrity (see Table 2.1). These personality measures include the HEXACO model (the Honesty-Humility dimension and responsibility of the Conscientiousness factor), the FFM (Agreeableness correlates with the Honesty-Humility factor of the HEXACO model), the Big Seven (correspondence between Negative Valence and the Integrity cluster of the SAPI), the CPAI (the responsibility facet of the Dependability factor) and the CPAI-2 (the veraciousness vs. slickness factor of Accommodation).

Table 2.1: Similarities between the SAPIs Integrity Cluster and Other Personality Measures

	SAPI INTEGRITY SIMILARITIES
HEXACO	The Honesty-Humility dimension (sincerity, fairness and modesty); and Responsibility of the Conscientiousness factor (the responsible facet features in the Integrity cluster of the SAPI structure).
BIG FIVE	Partial association of the Big Five factors (sympathy and soft-heartedness) with a blend of the HEXACOs Honesty-Humility (sincerity, fairness, modesty versus greed, pretentiousness, slyness, hypocrisy), Agreeableness and Emotionality.
FFM	Agreeableness correlates with the Honesty-Humility factor of the HEXACO model.
BIG SEVEN	Correspondence between Negative Valence and the Integrity cluster of the SAPI.
CPAI	The responsibility facet of the Dependability factor, consisting of facets such as being trustworthy, honest and morally conscious. Also the facets truthful and responsible of the Integrity cluster reinforces the correspondence between Dependability and Integrity.
CPAI-2	The veraciousness vs. slickness factor of Accommodation corresponds with integrity, which encompasses a person's moral soundness.

2.4 MEASUREMENT OF INTEGRITY

Integrity tests first appeared in the late 1940s. Since then over 40 off-the-shelf integrity tests have become available to organisations in the USA (Ones et al., 2003). Integrity tests, also referred to as honesty scales, are prototypical criterion-focused occupational personality scales (Ones & Viswesvaran, 2001). Integrity tests are paper and pencil tests that are specifically developed to assess the integrity, honesty and dependability of applicants and employees, thereby facilitating prediction of job performance and counterproductive behaviours on the job, such as theft, disciplinary problems, absenteeism and future on the job dishonest behaviours (Ones, 1993; Ones et al., 1993). The individual differences perspective on organisationally counterproductive behaviours assumes that individuals who are dishonest, untrustworthy and irresponsible will be poorer employees overall and will engage in more organisationally undesirable behaviours (Ones et al., 1993). Therefore, assessment procedures in the workplace aim to help a particular employer make better selection decisions (Cohen, 1997). As such, “the purpose of an integrity test as part of a pre-employment screen of job applicants is to refuse employment to those individuals who do not meet the minimal test scores deemed suitable by an organization” (Cohen, 1997, p. 2).

Sackett, Burris and Callahan (1989) classify honesty tests into two categories, namely overt integrity tests (also known as 'clear purpose tests') and personality-based integrity tests (also called 'disguised purpose tests'). Overt integrity tests measure an individual's attitude towards theft and ascertain the presence of prior dishonest and illegal behaviour (Sackett et al., 1989). Overt integrity tests typically have two sections (Sackett et al., 1989). The first section deals with attitudes toward theft and contains questions probing an individual's opinions regarding the frequency and extent of theft in society, attitudes toward punishment for theft, likelihood of exposure, knowledge of employee theft, rationalisations about theft and assessments of one's own honesty (Sackett et al., 1989). The second section deals with exposure to theft and other illegal activities, such as the amount of money stolen, drug use or gambling (Sackett & Harris, 1984; Sackett et al., 1989). Examples of overt integrity tests include the Employee Attitude Inventory (EAI; London House Press, 1982), Employee Reliability Inventory (ERI; Borofsky, 1993), Personnel Selection Inventory Honesty Scale (PSI Honesty Scale; London House, Inc., 1975), Phase II Profile (Lousig-Nont, 1987), Reid Report (Reid Psychological Systems, 1951), the SavvyIntegrity test (Cohen, 1997) and the Situational Judgement Test (SJT; Becker, 2005). The SavvyIntegrity test, for example, is a pre-employment integrity test predicting candidates' predisposition towards counterproductive job behaviour (Cohen, 1997). It is designed to measure a candidate's attitudes to a variety of 'honesty-integrity' related behaviours and statements, and/or determine past behaviour in relation to the constructs of theft, bribery, gambling, substance abuse, loyalty and accurate reporting (Cohen, 1997).

De Meijer, Born, Van Zielst and Van der Molen (2007), point out that SJTs have been in use since the 1920s but have become increasingly popular in personnel selection and research over the last ten to fifteen years. SJTs typically consist of hypothetical scenarios concerning a specific problem in the workplace (De Meijer et al., 2007). SJTs can either be paper and pencil tests with written descriptions of situations (Chan & Schmitt, 2002) or it can be video-based tests consisting of multimedia scenarios (Lievens, Buyse, & Sackett, 2005; Weekley & Jones, 1997). Instead of developing an overt or personality-based integrity test, De Meijer et al. (2007) constructed an Integrity-SJT relating to the advantages of SJTs and, more specifically, video-based SJTs. SJTs have high criterion-related validity (Lievens & Sackett, 2006), little adverse impact against ethnic minority applicants and high test realism, leading to more reliable respondent reactions (Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan & Drasgow, 2000). De Meijer et al.'s (2007) SJT was developed for the Dutch police and consists of videos of critical situations in which police-integrity violations are presented. The violations presented include corruption, fraud and theft, accepting dubious gifts and services, misuse of

authority and misuse of information (De Meijer et al., 2007). These police-integrity violations can be viewed as sub-dimensions of police integrity (De Meijer et al., 2007). Fertig's (2009) study concerning the incremental validity of SJTs in a South African retail bank, found that SJT scores are valid predictors of managerial performance.

On the contrary, personality-based integrity tests do not solely predict theft or theft-related behaviours. Instead these tests aim to predict a broad variety of counterproductive work behaviours, including disciplinary problems, violence, excessive absenteeism, excessive tardiness and drug abuse, based on a combination of measurements of personality traits such as reliability, conscientiousness, emotional stability, adjustment, trustworthiness and sociability (Ones, 1993; Sackett & Harris, 1984; Sackett et al., 1989). The Hogan Personnel Selection Series (Hogan, 1981), Personnel Decisions, Inc. Employment Inventory (PDI-EI; Paajanen, 1985), Inwald Personality Inventory (IPI; Inwald, Knatz, & Shusman, 1982), Personal Outlook Inventory (Science Research Associates, 1983) and Personnel Reaction Blank (PRB; Gough, 1954) are all personality-based integrity tests. Caution however needs to be applied when using personality-based integrity test as to predict counterproductive behaviour as it can lead to false positives and false negatives in selection decisions.

South African integrity measures include the Giotto Test (Rust, 1999), Integrity Profile-200 (IP-200; Fick, 2002) and the personality-based South African Integrity Scale (SAIS) developed by Hunter and Engelbrecht (2009). The Giotto Test is a work-based personality questionnaire designed to assess integrity (Rust, 1999). It contains 101 items and generates scores on seven scales based on the Prudentius model of personality. The traits measured are evaluative in nature. "Giotto makes use of an ipsative format in which force choice items are matched for social desirability" (Rust, 1999, p. 755). Neural network programming techniques are used to address the intrinsic nonlinearity of the ipsative framework (Rust, 1999). The IP-200 measuring instrument includes a so-called 'lie factor', thus measuring the person's commitment to answering the instrument in an open, honest and objective way (Fick, 2002). A few carefully worded and placed items are included in the test to facilitate the test taker who wants to create a favourable image that differs from their true image to do so and to be exposed in the process (Fick, 2002). Russouw (2005) used the IP-200 to conduct a study on integrity in organisations. The IP-200 is a valid and reliable instrument that is capable of measuring integrity and its substructures accurately (Fick, 2002).

The preliminary SAIS developed by Hunter and Engelbrecht (2009) consists of five dimensions labelled honesty, trustworthiness, norm abidance, responsibility and punitiveness. The scale is currently being validated against various criteria for counterproductive behaviour using a large multicultural sample from different industries in South Africa (Hunter & Engelbrecht, 2009).

Since the early 1990s interest in and the use of personality measurement and integrity testing for industrial and organisational applications has increased significantly (O'Bannon, Goldinger & Appleby, 1989; Ones et al., 1993; Sackett & Harris, 1984; Sackett et al., 1989). This has resulted in the development of some uncertainties. The similarity of the different types of integrity tests has become questionable and it seems doubtful that these tests all primarily measure a single general construct. According to O'Bannon et al. (1989), different test publishers claim that their integrity tests measure various constructs, including moral reasoning, responsibility, long-term job commitment, consistency, work ethics, dependability, hostility, tendency to violence, depression and energy level. Barrick and Mount (1991) and Digman (1990) suggest that all the various integrity tests actually measure the general construct of Conscientiousness, one of the five dimensions of personality used in the Big Five theory of personality. Conscientiousness is characterised by responsibility, carefulness and dependability and these traits are measured by most integrity tests (Ones et al., 1993). In general the different types of integrity tests are highly correlated and valid for both overall job-performance criteria and counterproductive-behaviours criteria (Ones et al., 1993).

To summarise, integrity tests have been extensively examined with regards to construct validity (Ones, 1993) and criterion-related validity (Ones et al., 1993). The construct validity of both overt and personality-based integrity tests has been measured in a large-scale meta-analytic cumulation and integrity tests have been found to measure the Big Five personality dimensions of (1) Conscientiousness, (2) Agreeableness and (3) Emotional stability (Ones, 1993). Also, substantial incremental validity shows that integrity tests have negligible correlations with cognitive ability (Ones, 1993). Furthermore, both overt and personality-based integrity tests appear to have useful levels of predictive validity when examined in accordance with a variety of criteria (Ones et al., 1993). These tests can help employers to make better employment decisions (Cohen, 1997).

2.5 DEVELOPMENT OF A MEASURING INSTRUMENT

Developing a psychological measure is a complex process with several distinct phases (Foxcroft & Roodt, 2007). The development includes the careful planning of the development of a psychological measure, writing items and administering the initial version of the measure so that the effectiveness of the items can be determined (Foxcroft & Roodt, 2007). Afterwards, the final items are chosen and the measure is administered to a representative group of people in order to establish the measure's validity, reliability and norms (Foxcroft & Roodt, 2007). Finally, the test manual is compiled (Foxcroft & Roodt, 2007).

Murphy and Davidshofer (2005) suggest that test development has three distinct stages. The first stage involves tests construction. In this stage issues such as item writing, item response alternatives, scale construction, response sets and selection of test format are of the primary interest (Murphy & Davidshofer, 2005). In the second stage new tests are standardised for use on particular target populations. This stage also involves the development of norms and considerable research is undertaken to establish estimates of the test's reliability and validity (Murphy & Davidshofer, 2005). In the third and final stage, tests need to be constantly revised to keep them contemporary and current. Revisions should involve the norms available as well as item content (Murphy & Davidshofer, 2005).

Foxcroft and Roodt (2007) propose a seven phase model for developing an instrument. This model was used to guide the development of the experimental integrity instrument used in this study. The seven phase model is discussed briefly in the section below.

The first phase is known as the planning phase and involves specifying the aim of the measure, defining the content and developing a test plan (Foxcroft & Roodt, 2007). Defining the content of the measure involves first defining the construct (content domain) to be tapped by the measure (Foxcroft & Roodt, 2007). Downing and Haladyna (2006) point out that the methods chosen to define test content are critical and depend on the purpose of the test, the consequences of the decisions made on the basis of the test scores and the validity evidence needed to support test score interpretations. Netemeyer, Bearden and Sharma (2003) explain that the rationale behind a construct's domain is that the interpretation of any measure is clearest when the domain is un-confounded. Once the construct has been defined the purpose for which the measure is developed must be considered (Foxcroft & Roodt, 2007). Downing and Haladyna (2006) state that an operational definition of the proposed test is formed by

deciding on a clear, concise, well-delineated purpose for the planned test. This purpose then guides nearly all other validity-related decisions related to test development activities. Developing the test plan (specifications) requires consideration of the format of the test (Foxcroft & Roodt, 2007), also referred to as the test blueprint (Downing & Haladyna, 2006). The test specifications must describe aspects such as (1) the type of testing format to be used (selected response or constructed response/performance); (2) the total number of items to be created or selected for the test, as well as the type or format of test items (such as, multiple choice, three option, single-best answer); and (3) how test scores will be interpreted (for instance norm or criterion referenced; Downing & Haladyna, 2006).

Foxcroft and Roodt (2007) identify two aspects of the test format. These aspects are the test stimulus to which a test taker responds and a mechanism for response (Foxcroft & Roodt, 2007). The items provide the stimulus for the test taker (Foxcroft & Roodt, 2007). Open-ended items are a common item format and place no limitations on the response of the test-taker (Foxcroft & Roodt, 2007). A second common item format is forced-choice items, for example, multiple-choice items (where careful consideration needs to be given to the distracters or alternative options used), and true-false or yes-no items (where the test taker responds to each individual item) (Foxcroft & Roodt, 2007). Netemeyer et al. (2003) also refer to true-false/yes-no items as dichotomous items. A third common item format involves sentence completion items, where a respondent must complete a sentence provided (Foxcroft & Roodt, 2007). Another common item format relates to performance-based items. Such items include tasks where apparatus (for example blocks) needs to be manipulated by the test-taker, a scientific experiment needs to be performed, an essay must be written, an oral presentation must be prepared or a work-related competency must be demonstrated through solving complex realistic problems via innovative interactive computerised activities (Foxcroft & Roodt, 2007). Foxcroft and Roodt (2007) provide various scaling options that can be used to respond to items in a measurement. These scaling options include category scales, Likert-type scales, semantic differential scales, intensity scales, constant-sum scales, paired comparison scales, graphic rating scales, forced choice scales and Guttman scales. Netemeyer et al. (2003) also refer to these types of scales as multichotomous or polytomous scales. These scales have three or more scale points, for example, disagree, agree, strongly disagree and strongly agree. The choice of item and response formats should be congruent with the purposes and intended uses of the test (Downing & Haladyna, 2006).

The second phase involves the *writing of the items* (Foxcroft & Roodt, 2007). Downing and Haladyna (2006) point out that the choice of item format is a major source of validity evidence for the test and, therefore, a clear rationale is required for item format selection. Foxcroft and Roodt (2007), Downing and Haladyna (2006), Murphy and Davidshofer (2005) and Netemeyer et al. (2003) provide some important guidelines for item writing. These authors suggest that after considering the type of measure being developed: (a) the wording must be clear and concise (clumsy wording and long, complex sentences could make it difficult for test takers to understand what is required of them); (b) use vocabulary that is appropriate for the target audience; (c) avoid using negative expressions such as ‘not’ or ‘never’ and double negatives in particular – this often creates uncertainty on the part of the test takers; (d) cover only one central theme in an item; (e) avoid ambiguous items; and (f) the nature of the content covered should be relevant to the purpose of the measure. Murphy and Davidshofer (2005) and Downing and Haladyna (2006) also suggest that test developers should also consider the presence of sexist, racist, ethnic or other offensive language in test items. Subtle linguistic nuances are often overlooked and not perceived as offensive by the developer, this could include reference to only one gender, for example, ‘mailman’ rather than ‘mail carrier’ or ‘fireman’ rather than ‘fire-fighter’ (Murphy & Davidshofer, 2005). Netemeyer et al. (2003) highlight the importance of considering translation validity when writing items. This involves content and face validity, and being over-inclusive of the construct’s domain rather than under-inclusive in generating an item pool (Netemeyer et al., 2003). Care must also be taken to ensure that each content area of the construct has an adequate sampling of items (Netemeyer et al., 2003). For example, if the construct is narrowly defined and the resulting scale is to be short (5-10 items), a pool of 20-30 items may suffice (Netemeyer et al., 2003).

After a pool of items has been developed it should be submitted to a panel of experts for review and evaluation (Foxcroft & Roodt, 2007; Netemeyer et al., 2003). The expert reviewers will judge whether the items sufficiently tap the content domain or dimensions of the construct being assessed. The reviewers will also comment on the cultural, linguistic and gender appropriateness of the items and examine the wording of the items and the nature of the stimulus materials (Foxcroft & Roodt, 2007). Based on the review panel’s recommendations, as well as data gathered from testing the items, certain items in the pool may have to be revised or re-written (Foxcroft & Roodt, 2007).

The third phase involves the assembling and pre-testing of the experimental version of the measure (Foxcroft & Roodt, 2007). This stage would typically involve (a) arranging the items

in a logical way in terms of the construct being measured (for paper-based tests, items need to be grouped or arranged on the appropriate pages in the test booklet); (b) finalising the length, which involves consideration of the time test takers will need to read items; (c) answering the protocols (for paper-based tests, decisions need to be made as to whether items will be completed in the test booklet or whether a separate answer sheet (protocol) needs to be developed); (d) developing clear and unambiguous administration instructions to be used for the experimental try-out of the items; and (e) pre-testing the experimental version of the measure by administering the measure to a large sample (400 to 500) from the target population and gathering information about how test takers responded to the stimulus materials, the ordering or sequencing of the items and the length of the measurement (Foxcroft & Roodt, 2007). Murphy and Davidshofer (2005) and Downing and Haladyna (2006) also refer to this phase as the standardisation of the testing procedure. This standardisation also includes factors such as the physical environment in which testing takes place, the health of test takers and the time of the test administration. Netemeyer et al. (2003) include four issues to consider in pilot testing: (a) the size of the sample, which depends on the number of items in the initial item pool; (b) sample composition, where convenience sampling (for example, college students) may suffice but a sample from a relevant population of interest is preferable; (c) initial item reliability estimates, where items can be assessed initially for internal consistency, means, variances, average inter-item correlation and factor structure, items that are not empirically supported can then either be deleted or adjusted; and (d) the number and type of validity-related scales to include, focussing on where three or four key construct measures should be included for examining initial estimates of validity. In addition, an initial estimate of social desirability bias may be considered if the focal construct has the potential for such bias.

Test administration instructions to test takers need to be clear so that test takers know exactly what they are expected to do (Downing & Haladyna, 2006). Downing and Haladyna (2006) state that in many cases it is desirable to provide information about the test content, item formats, scoring rules and test-taking strategies and practice test materials in advance. Hambleton (2001) identifies types of problems that can be expected in relation to the administration of a test. Test developers and administrators should try to anticipate these problems and take appropriate corrective actions through the preparation of appropriate materials and instructions (Hambleton, 2001). For example, test administrators need to be sensitive to a number of factors related to the stimulus materials, administration procedures, and response modes that can moderate the validity of the inferences drawn from the scores (Hambleton, 2001). Test administrators should be especially wary of 'test-wisness skills' and

make corrections when necessary (Hambleton, 2001). For example, multiple-choice test respondents need to know what to do when they cannot answer a question (skip or guess) and how to pace themselves to finish the test within the time limit (Hambleton, 2001).

Environmental aspects that influence the administration of a test should be made as similar as possible across populations for whom the test is intended. Failure to follow the specific instructions can invalidate comparisons (Hambleton, 2001). In cultures in which untimed or unspeeded tests are the norm, the imposition of time limits can cause problems for respondents. Similarly, it has been suggested that testing in air-conditioned locations should be required when testing in different countries (Hambleton, 2001). The International Test Commission (ITC) recommends that testing conditions should be as ideal as possible. Where conditions are not ideal this information must be noted and taken into account when interpreting scores (Hambleton, 2001).

To minimise the influence of unwanted sources of variation across populations: (a) “test administration instructions should be in the source and target languages”; (b) “the test manual should specify all aspects of the test and its administration that require scrutiny in the application of the test in a new cultural context”; (c) “the administrator should be unobtrusive and the administrator-examinee interaction should be minimized”; and (d) the explicit rules “described in the manual for the test should be followed” (Hambleton, 2001, p. 169).

Murphy and Davidshofer (2005) suggest three scaling methods for item grouping. These scale methods result in psychological scales that have different characteristics and that require different interpretations. Rational scales imply that “some underlying thought, belief, or rationale is used as the basis for selecting items and grouping them together into a scale” (Murphy & Davidshofer, 2005, p. 231). These scales are tied to the validity of the theory on which they are based (Murphy & Davidshofer, 2005). Empirical scales rely on criterion keying for their validity. These scales make use of factor analysis in the item selection process. During factor analysis the correlation among items is examined to determine which factors account for most of the variability in test responses (Murphy & Davidshofer, 2005). Once the items representing each factor have been identified, they are grouped together and given a label (Murphy & Davidshofer, 2005). Rational-empirical scales are the third scaling method used for item grouping. In this scaling method test constructors combine aspects of both the rational and the empirical approaches to scale construction in the test development process (Murphy & Davidshofer, 2005).

The fourth phase of test development involves item analysis (Foxcroft & Roodt, 2007). This phase adds value to item development and the development of the measure in general by examining each item to see whether it serves the purpose for which it was designed (Foxcroft & Roodt, 2007). Item analysis detects flawed items, determines item difficulty, identifies whether items discriminate between good and poor performers, whether items are biased against certain groups and what the shortcomings of an item are (Downing & Haladyna, 2006; Foxcroft & Roodt, 2007). Statistical analyses are conducted to evaluate the characteristics of each item (Foxcroft & Roodt, 2007). The characteristics of items can be analysed using classical test theory (CTT) or item response theory (IRT) statistical methods or both (Foxcroft & Roodt, 2007). The resultant statistics are used to guide the final item selection and the organisation of the items in the measure (Foxcroft & Roodt, 2007).

Revising and standardising the final version of the measure constitutes the fifth phase of test development (Foxcroft & Roodt, 2007). This phase focuses on revising the items and test, and then administering the final version of the test to a large sample for standardisation purposes (Foxcroft & Roodt, 2007). Foxcroft and Roodt (2007) provide criteria for this phase. Firstly, the items must be revised and tested. Items identified as problematic during the item analysis phase need to be considered and a decision must be made for each item regarding whether it should be discarded or revised. Second, items must be selected for the final version. The existing database can be used to check the acceptability of the reliability and validity coefficients of the final measure. Third, administration instructions and scoring procedures must be defined. This step is based on the experience and feedback gained during the pre-testing phase and involves the modification of administration and scoring instructions. Finally, the final version must be administered. The final version is administered to a large, representative sample of individuals for the purposes of establishing the psychometric properties (validity and reliability) and norms.

The sixth phase involves the technical evaluation and the establishment of norms. Two criteria are used for this phase: (1) establishing validity and reliability – the psychometrics of the measure need to be established by computing appropriate types of validity and reliability coefficients; and (2) establishing norms, setting performance standards or cut-scores as appropriate (Foxcroft & Roodt, 2007). Murphy and Davidshofer (2005) provide two steps in establishing norms. The first step involves defining the target population (the composition of the normative group is determined by the intended use of the test) while the second step

involves selecting the sample (once the appropriate target groups have been determined, a representative sample from each of these groups must be obtained).

The seventh and final stage involves publishing and ongoing refinement (Foxcroft & Roodt, 2007). Before a measure can be published, a test manual must be compiled and submitted for classification (Foxcroft & Roodt, 2007). Once a measure has been published, ongoing research is conducted into its efficacy and psychometric properties. This ongoing research should lead to revised editions of the manual being published to provide updated information to assessment practitioners (Foxcroft & Roodt, 2007; Murphy & Davidshofer, 2005).

The current study only made use of the first six phases of Foxcroft and Roodt's (2007) seven phase model for developing an instrument. However, for the purpose(s) of this study only the measure's reliability was established in the sixth phase.

2.5.1 Faking and Social Desirability

The increased use of personality measures in personnel selection has raised concerns about the extent to which applicants can and do distort responses on such measures in an effort to improve their chances of being hired (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Kuncel and Tellegen (2009) define faking as deliberate over-reporting of favourable characteristics. If faking has an impact on personality measurements, the personnel selection process could result in incorrect selection decisions (Ones & Viswesvaran, 1998). These incorrect selection decisions are known as false positive and false negative errors in personnel selection (Ones et al., 1993). Honest applicants would be rejected even if they did not fake (false-positive error), whereas dishonest applicants, who fake responses, would get selected (false-negative error; Ones et al., 1993). However, data from applicants, as well as large-scale meta-analyses, has shown that faking does not destroy predictive validity and therefore has no influence in the prediction for personnel selection on general personality measures (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ones, Schmidt, Viswesvaran, & Lykken, 1996) and integrity tests (Ones et al., 1993).

McFarland (2003) used a method of warning applicants of the inclusion of a social desirability (SD) scale in a personality measure. The presence of a warning was shown to decrease applicant faking. It was found that applicant reactions are not negatively affected by warnings, but warnings change the relationships among personality variables, socially desirable

responding, and organisational justice variables (McFarland, 2003). Therefore, warnings have the added benefit of resulting in less multi-collinearity among personality measures (McFarland, 2003).

Robinson, Shaver, and Wrightsman (1991) define response bias as “a systematic tendency to respond to a range of questionnaire items on some basis other than the specific items content (in other words, what the items were designed to measure)” (p. 17). This bias forms a response style when an individual displays it consistently across time and situations (Robinson et al., 1991). Robinson et al. (1991) point out that socially desirable responding (SDR; for example lying or faking good) is one of the most prominent response biases. SDR reflects tendencies to provide favourable responses with respect to norms and practices (Nederhof, 1985). Kuncel and Tellegen (2009) define SDR as “behaving in a manner that is consistent with what is perceived as desired by salient others” (p. 202). SDR can also be defined as the tendency of individuals to make themselves look good, with respect to cultural norms, when answering researchers’ questions (Mick, 1996). Depending on what the specific questionnaire is measuring, for example personality characteristics, a SD scale may be included (Foxcroft & Roodt, 2007). Robinson et al. (1991) describe the SD as measuring an individual’s tendency to respond favourably when describing the self. SD can also be broadly defined as the tendency to respond to items in a way that is socially acceptable (Berry, Poortinga, Segall, & Dasen, 2002). Kuncel and Tellegen (2009) highlight that the purpose of SD assessments of items is used for item selection and scale development. Any psychological test that measures constructs that are viewed as desirable is susceptible to this kind of response bias (Taylor, 2008).

Various types of SD scales exist. However, according to Paulhus, Harms, Bruce and Lysy (2003), the most well-known are the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984), Marlowe-Crowne Social Desirability scale (MCSD; Crowne & Marlowe, 1960) and the Edwards Social Desirability scale (Edwards, 1957). The BIDR (Paulhus, 1984) defines socially desirable responding as consisting of conscious (impression management) and unconscious (self-deceptive enhancement) responding. Paulhus (1991) explains that the results of unconscious responding are likely to be more stable than those of conscious responding, as the nature of the individual’s self-deception is unlikely to vary over situations.

Paulhus (1984) also highlights the distinction between attribution and denial responses to self-report items. Attribution responses involve claiming socially desirable characteristics for the self, while denial responses involve denying that undesirable characteristics apply to the self

(Paulhus, 1984). The items of the Marlowe-Crowne scale are partitioned into attribution and denial subscales in order to assess these components (Paulhus, 1984). Mixed evidence was found regarding the two components' differential relation to cheating behaviour (Paulhus, 1984). As a whole the case for a separation of attribution and denial components of social desirability is weak (Paulhus, 1984).

A second approach divides SDR by distinguishing between self-deception and impression management (also known as other-deception) as two components that are differentially influenced by social pressure (Paulhus, 1984). Positively and negatively keyed items are balanced across the two dimensions (Paulhus, 1984). Self-deception, where the respondent actually believes his or her positive self-reports, does not vary according to the situation. Impression management, to consciously pretend or hide one's real feelings or intentions, varies according to momentary motives and situational demands (Paulhus, 1984). The distinction makes it difficult to identify the origins of SD, particularly when investigating cross-cultural differences in the patterns of this response style (Taylor, 2008).

The MCSD scale measures both factors of SDR but focuses more on impression management than on self-deception (Crowne & Marlowe, 1960). The scale contains 33 True-False items describing either (a) desirable but uncommon behaviours (for example, admitting mistakes) or (b) undesirable but common behaviours (such as, gossiping) (Crowne & Marlowe, 1960). The behaviours concern everyday events (ordinary personal and interpersonal behaviours) and not psychopathology (Crowne & Marlowe, 1960). Eighteen of the 33 items are keyed in the true direction and the other fifteen are keyed in the false direction (Crowne & Marlowe, 1960). Hence, scores range from 0 to 33, with higher scores representing a higher need for approval (Crowne & Marlowe, 1960). The MCSD scale generally correlates less highly with measures of adjustment, than the Edwards SD scale (Crowne & Marlowe, 1960).

Taylor (2008) identifies possible antecedents of SD as sources of bias across cultures in personality assessment. The three most researched elements that have an impact on the occurrence of SD are cultural norms, personality and context (Taylor, 2008). *Cultural norms* focus on the collectivism-individualism dimension (Taylor, 2008). For example, Bernardi's (2006) study concluded that individuals from countries with high uncertainty avoidance and high collectivism are likely to engage in SDR. These findings are discussed in more detail later in this section. There is currently an ongoing debate regarding SD's status as a personality trait or a response style (Taylor, 2008). Some research findings suggest that SD is a personality trait

(Taylor, 2008). These studies suggest that individuals who are very agreeable are likely to behave in ways that are socially desirable, as would individuals who report higher levels of emotional stability. Similarly, individuals who are more conscientious are likely to be more socially responsible, and display behaviours that are more desirable (Taylor, 2008). Ones, Viswesvaran and Reiss (1996) conducted a meta-analysis of studies of job applicants and found that SD scales were correlated with emotional stability and conscientiousness. In contrast, measures of extraversion and openness to experience were not related to SD scales. Context, the third element that impacts on SD, refers to the circumstances in which individuals find themselves. Much of the research on SDR investigates the impact of this response style on the validity of personality scales in the selection context (Taylor, 2008). Thus, individuals have more motivation to adjust their responses in a socially desirable manner when applying for a job (high stakes) than when completing an assessment for the purpose of self-development (low stakes) (Taylor, 2008). In high stakes settings, SD is often referred to as faking and has an element of deliberate action rather than unintentional positive impression management (Taylor, 2008). Ones et al. (1996) found that SD did not moderate or mediate the relationship between personality and job performance and could not be used to predict success in the workplace.

Kuncel and Tellegen (2009) highlight assumptions made concerning the test taking behaviour of people who intentionally or unintentionally distort their responses. For example, individuals with a desirable response set, as well as individuals intentionally faking, will always approve of 'desirable' items at the highest level of approval permitted by the item format. This is presumably based on the assumption that the strongest approval option is also the most desirable response (Kuncel & Tellegen, 2009). The relation between trait level and SD is critical when SD is used in studies of desirable responding in the development of personality scales. It is also important in applied settings where it is used to evaluate the validity of self-report test results (Kuncel & Tellegen, 2009).

A study by Bernardi (2006) examined the associations among social desirability response bias (SDRB), Hofstede's cultural dimensions and gender in 1 537 students in 12 different countries, including South Africa. Zerbe and Paulhus (1987) highlight two aspects of SDRB. The first aspect involves individuals over-reporting activities that are deemed to be socially or culturally desirable, whereas the second aspect occurs when individuals under-report activities that are deemed to be socially or culturally undesirable (Zerbe & Paulhus, 1987). Bernardi (2006) concluded that SDRB decreased/increased as a country's Individualism/Uncertainty Avoidance increased. The study also found that SDRB was higher for males than for females (Bernardi,

2006). Bernardi (2006) refers to other research that also noted differences between male and female students in ethical sensitivity. The research referred to by Bernardi (2006) found that female students are consistently more concerned about ethical issues than male students.

Robinson et al. (1991) provide four methods for controlling SDR. These methods are labelled the rational, factor analytic, covariate and demand reduction methods. Rational techniques involve control features built into the self-report instrument (Robinson et al., 1991). The purpose of these control features is to prevent the subject from responding in a socially desirable fashion (Robinson et al., 1991). For example, forced choice items that pair items of equal desirability ratings in an ipsative format (Murphy & Davidshofer, 2005; Nederhof, 1985). Factor analytic methods may be applied during test construction if the procedure involves choosing the highest loading items in a factor analysis (Robinson et al., 1991). In covariate methods a measure of SDR, for example the MCSD scale, is administered along with the content measures and no attempt is made to prevent respondents from answering in a desirable fashion (Robinson et al., 1991). Demand reduction (a wide-range form of control) methods include those methods that reduce the situational press for desirable responding (Robinson et al., 1991). Robinson et al. (1991) suggest that the most obvious strategy in demand reduction is to assure respondents of anonymity. Nederhof (1985) and Murphy and Davidshofer (2005) provide an additional method to control or eliminate the effects of SDR on personality tests. These authors recommend the selection of neutral items that are neither strongly positive nor strongly negative on the SD dimension.

To summarise, when developing a measuring instrument it is important to carefully and thoroughly consider all the relevant criteria in order to have a reliable and valid end product. In addition, large-scale meta-analyses have shown that faking on integrity tests does not destroy predictive validity and therefore does not significantly influence the prediction of personnel selection (Ones et al., 1993). Furthermore, social desirability scales are used to monitor respondents' favourable responses and attempts to portray themselves as socially acceptable. The last section of this chapter focused on possible antecedents of SD and looked at techniques to control SDR.

For the purposes of this study a SD scale (more specifically, the MCSD scale) was adapted and used as part of the developmental phase of the experimental integrity instrument. The purpose of the study was to confirm the factor structure of the integrity cluster and therefore no analysis

was carried out on the SD scale results. The broader SAPI project will make use of this data to conduct further analyses.

2.6 SUMMARY

This chapter reviewed cross-cultural personality assessment. The chapter also considered the concept of integrity by focussing on defining integrity, outlining the models of integrity measurement and considering the development of the integrity cluster within the SAPI. The measurement of integrity was also discussed. The chapter then looked at the phases in developing a measuring instrument and concluded with a section on faking and social desirability. The following chapter discusses the research methodology and focuses on the research problems and the empirical study (research design, measuring battery, participants, procedure and statistical analysis).

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

3.1 INTRODUCTION

Chapter three outlines the specific research process that was followed in this study. The chapter includes a discussion of the study's research design, participants, procedures, measuring instruments and the statistical analysis.

3.2 RESEARCH DESIGN

The study made use of a quantitative research design in order to reach a large number of participants. Trochim and Donnelly (2001) define quantitative data as “the numerical representation of some object. A quantitative variable is any variable that is measured using numbers...” (p. 11). Quantitative research can also be defined as “a form of conclusive research involving large representative samples and fairly structured data collection procedures” (Struwig & Stead, 2001).

Struwig and Stead (2001) identify five characteristics of quantitative research:

- (a) Constructs and their measurement. Quantitative research examines constructs (variables) based on hypotheses by means of questionnaires and/or some form of structured observation.
- (b) Causality. Quantitative research involves establishing causal relationships (cause and effect relationships) between constructs by using independent and dependent variables.
- (c) Generalisation. Quantitative research establishes findings that can be legitimately generalised beyond the confines of the research sample.
- (d) Replication. Replication of a study provides a way of determining the extent to which findings are applicable to other contexts, and also serves as a means of examining the biases of the investigator.
- (e) The individual as focus. In quantitative research survey instruments are administered to individuals and the individual's (not the group's) responses are required and collected to form overall measures for the sample.

The SAPI experimental integrity instrument uses an exploratory research approach to develop and clarify ideas and research questions regarding a previously unexplored topic. In this study

an experimental integrity instrument was developed. The study involved a cross-sectional questionnaire design where the integrity of multi-cultural entry-level police applicants of the South African Police Service (SAPS) was assessed. A pilot study was conducted, followed by the administration of the experimental integrity instrument to a larger group of entry-level police applicants.

3.3 PARTICIPANTS AND PROCEDURES

A sample can be described as a smaller collection of units from a population that is used to determine the truth about the population (Field, 2005). Struwig and Stead (2001) identify two groups of sampling methods, namely non-probability sampling techniques and probability sampling techniques. This study focuses only on non-probability sampling techniques. In non-probability sampling “the probability of any particular member of the population being chosen is unknown” (Struwig & Stead, 2001, p. 111). Struwig and Stead (2001) highlight four non-probability sampling techniques: (a) convenience sampling in which the sample is selected purely on the availability, (b) judgement sampling in which the sample is selected on the basis of expert judgement, (c) quota sampling which involves the selection of a sample on their characteristics (for example, age, income, socio-economic status and gender) and (d) snowball sampling which involves a variety of procedures in which initial respondents are selected by probability methods but additional respondents are obtained from information provided by the initial respondents. In this study, convenience sampling was used for both for the pilot and main study.

3.3.1 Pilot Study

The pilot study was conducted with members of the public who applied for administrative clerical positions at the firearm registrar department of SAPS in Pretoria. Table 3.1 provides a description of the demographic composition of the pilot sample.

Table 3.1

Demographics of the Pilot Sample (N = 174)

Item	Category	Frequency	Percentage (%)
Gender	Male	145	83.3
	Female	29	16.7
Race	Black	143	82.2
	White	8	4.6
	Coloured	16	9.2
	Indian	7	4.0
Language	Afrikaans	22	12.6
	Sepedi	52	29.9
	South Sotho	12	6.9
	Zulu	16	9.2
	English	7	4.0
	Tshivenda	9	5.2
	Xitsonga	14	8.0
	Xhosa	6	3.4
	SeTswana	28	16.1
	Swati	1	0.6
	Ndebele	7	4.0
Other	0	0.0	
Education level	Grade 12	68	39.1
	Certificate	67	38.5
	Diploma	35	20.1
	B Degree	1	0.6
	Honours or Equivalent	0	0.0
	Masters	0	0.0
	Doctorate	0	0.0
	Other	3	1.7

The total sample consisted of 174 participants. 83.3% of the participants were male and 16.7% were female. The majority of the participants identified themselves as Black (82.2%). The Indian race group formed the smallest racial group in the sample (4%). The sample consisted of various group languages of which Sepedi (29.9%) and SeTswana (16.1%) speaking participants had the highest representation. Thirty-nine-point-one percent (39.1%) of the sample had a Grade 12 educational level, 38.5% of the sample had some form of certificate course and 20.1% of the participants had a diploma. The mean age of the sample was 25 years.

The purpose of the pilot study was to standardise the testing procedure by experimenting with the pilot measuring instrument. One of the objectives of the pilot study was to develop clear, unambiguous administration instructions for the experimental try-out of the items. An additional objective was to pre-test the experimental version of the measure. A third objective was to gather information about how test takers responded to the stimulus materials, the ordering and sequencing of the items and the length of the measurement.

The pilot integrity instrument was applied to eleven language groups with approximately 25 participants per group. Participants were given standardised instructions and had 60 minutes to complete the questionnaire. Both the questions and answers were presented in the same booklet. The participants in the pilot study provided positive feedback and indicated that they understood the test instructions and the rating scale used and there were no problems with the items. The main recommendation from the pilot study involved separating the booklet from the answer sheet for use in the experimental integrity instrument. The recommendation was beneficial in that less paper was used, the booklets can be reused and the data capturing was simplified as all the answers were on a single sheet. A difficulty with this recommendation was the negative impact it could have on accuracy. This is because test takers may be subject to errors when capturing the answers for the corresponding items when working with two separate paper sheets.

3.3.2 Experimental Integrity Sample

Based on the recommendations from the pilot study the experimental integrity instrument was developed and administered to police members who were selected into the SAPS in July 2009. The same administration procedures were followed as in the pilot study. Table 3.2 provides a description of the demographic composition of the experimental integrity instrument sample.

Table 3.2

Demographics of the Experimental Integrity Instrument Sample (N = 1023)

Item	Category	Frequency	Percentage
Gender	Male	491	48.0
	Female	532	52.0
Race	Black	997	97.5
	White	7	7.0
	Coloured	15	1.5
	Indian	2	0.2
Language	Afrikaans	16	1.6
	Sepedi	284	27.8
	South Sotho	39	3.8
	Zulu	74	7.2
	English	7	0.7
	Tshivenda	68	6.6
	Xitsonga	152	14.9
	Xhosa	31	3.0
	SeTswana	223	21.8
	Swati	86	8.4
	Ndebele	40	3.9
Other	3	0.3	
Education level	Grade 12	752	73.5
	Certificate	176	17.2
	Diploma	77	7.5
	B Degree	6	0.6
	Honours or Equivalent	0	0.0
	Masters	0	0.0
	Doctorate	0	0.0
Other	7	0.7	

The total sample consisted of 1023 participants. 48% of the participants were male and 52% were female. Black respondents (97.5%) constituted the majority of the sample. The Indian group (0.2%) had the least representation. The sample consisted of various language groups of which Sepedi (27.8%) and SeTswana (21.8%) were the most common. 73.5% of the sample had a Grade 12 educational level, 17.2% had some form of certificate course and 7.5% had a diploma. The mean age of the sample group was 26 years.

3.4 MEASURING INSTRUMENT

The researcher followed several phases in the development of the experimental integrity instrument. These phases broadly followed the phases proposed by Foxcroft and Roodt (2007) discussed in chapter 2. A brief discussion of the phases followed and the operational considerations involved follows.

3.4.1 Phase 1: Aim of the Instrument and Construct Domain

The main aim of the experimental integrity measure was to measure the construct of integrity as conceptualised in the first phase of the SAPI project. The integrity domain consists of ten facets, each with a specific definition (see table 3.3). The items' target behaviours were used as a guide for defining the facets.

Table 3.3: Definitions of the Integrity Cluster

INTEGRITY SUB-CLUSTER	
Facets	Definitions
Honesty	Being genuine, sincere, and straightforward.
Integrity	Being honest and having sound values, rather than doing or saying something bad to others.
Loyal	Being an unconditional, true friend who is committed.
Morally Conscious	Having good morals and values, a sense of what is right or wrong, and being righteous.
Pretending	Hiding who one truly is or how one feels by giving double messages and deceiving others in doing so.
Responsible	Taking up and living up to responsibilities - with respect to tasks and other people. Acknowledging one's mistakes and not blaming others for those.
Trustworthy	Being a reliable person that others can confide in and depend on; keeping one's promises and others' secrets.
Truthful	Telling the truth and not telling lies; wanting to be told the truth; standing up for the truth.
FAIRNESS SUB-CLUSTER	
Facets	Definitions
Discriminative	Being prejudiced towards others with different orientations, backgrounds, and beliefs.
Fair	Treating people in an impartial and ethically responsible manner.

3.4.2 Phase 2: Writing of the Items

Two sets of items were developed. A pool of 117 items was constructed first, and then 15 social desirability items were constructed. A four step procedure was used to develop the item pool. The first step involved considering the original responses per facet. This involved gathering the original Integrity cluster responses of all eleven language groups' data Excel files. The second step was concerned with establishing a definition of the facet. This step involved working through all eleven languages' original responses for a given facet and establishing a definition for that facet. This step also involved extracting content-representative responses. This meant that before a definition could be established responses that were representative of the content of a specific facet had to be extracted. Step three entailed developing item stems. This step was similar to step 2, but items were presented in a more neutral and abstract form. For example, if the original response was 'she'll give an honest opinion' then it was formulated into "giving one's honest opinion". The final item writing took place in step 4. This step included the development of final items by making use of the neutral and abstract items developed in step 3. In this step the following criteria were used to formulate the final items:

- Items had to be short, simple and clear.
- Items were written in the first person, starting with 'I' followed by concrete behaviours, the object and the context.
- Negations were excluded in the first part of an item. A clear statement had to be made by using a negative statement in the second part of an item, but not in the third part.
- Items described a single activity, habit or preference (terms such as like/dislike were avoided).
- Temporal qualifiers were excluded, for example often, always, sometimes.
- Items were formulated in the direction of the construct. More specifically, items that dealt with the negative pole of the construct were used (a) if there were many utterances in the original responses or (b) if a significant item dealt with the negative pole. Double-barrelled items were excluded.
- Items had to refer to concrete behaviours and not beliefs, values or orientations.
- Psychological trait terms were avoided.
- Items excluded the use of idioms and expressions/sayings in order to avoid confusion.
- Items were written with consideration given to their translatability.

3.4.3 Phase 3: Inclusion of other items

In this phase items from other personality instruments (for example, the International Personality Item Pool/IPIP, Goldberg, 1999; and the SAIS, Hunter & Engelbrecht, 2009) were evaluated for inclusion in some of the integrity facets. This was done because some of the integrity facets did not contain enough items. Each facet had to include between ten and twenty items. However, after this phase was completed, five of the nine facets did not meet this criteria, namely Honest (8 items), Responsible (25 items), Truthful (9 items), Discriminative (9 items) and Fair (9 items). Nevertheless, it was reasoned that the amount of items in each of the specific facets were sufficient to be analysed. Table 3.4 details the additional items that were transformed and reformulated under the various integrity facets.

Table 3.4: Additional Personality Instrument Items Transformed

Personality Instrument	Personality Instrument Items	Facet	Integrity Item
South African Integrity Scale (SAIS; Hunter & Engelbrecht, 2009)	I will do anything for my friends	Loyal	20
	It is only human to sometimes break a rule	Morally Conscious	39
	I sometimes do things I know I should not do	Morally Conscious	40
	I have done things that I am not proud of	Morally Conscious	41
International Personality Item Pool (IPIP; Goldberg, 1999)	Am true to my own values	Morally Conscious	43
	Pretend to be concerned for others	Pretending	59
	Try to fool others	Pretending	60
	Does what's expected of him/her	Responsible	88
	Takes good care of his/her belongings	Responsible	89
	Give everyone a chance	Fair	132

3.4.4 Phase 4: Development of Social Desirability items

The Marlowe-Crowne Social Desirability scale (MCSD; Crowne & Marlowe, 1960) is one of the more well-known measures of social desirability (Robinson et al., 1991). It contains 33 items with a true/false response format (Crowne & Marlowe, 1960). The items include behaviours that are desirable but rare, or undesirable but common. The items focus on ordinary

personal and interpersonal behaviour and not psychopathology (Crowne & Marlowe, 1960). Fifteen of the 33 items were selected, modified and distributed across the integrity instrument (see Table 3.5). The MCSD scale was selected for use in this research because its items are similar to those of the integrity instrument in that both focus on ordinary, everyday behaviour.

Table 3.5: MCSD Items Modified

MCSD Items (Original)	MCSD Items (Modified)	MCSD Item	Integrity Item
1. Before voting I thoroughly investigate the qualifications of all the candidates.	1. I think about my options before I make a choice	1	7
2. I never hesitate to go out of my way to help someone in trouble	2. I help others in trouble	2	14
3. It is sometimes hard for me to go on with my work if I am not encouraged	3. I continue with my work if I am motivated	3	21
4. On occasion I have had doubts about my ability to succeed in life	4. I have doubts about my ability to succeed in life	5	28
5. I sometimes feel resentful when I don't get my way	5. I am satisfied when I get my way	6	35
6. I am always careful about my manner of dress	6. I am careful about my way of dressing	7	42
7. I like to gossip at times	7. I gossip	11	49
8. No matter who I'm talking to, I'm always a good listener	8. I am a good listener	13	56
9. I sometimes try to get even, rather than forgive and forget	9. I forgive others for their wrongdoings	19	63
10. When I don't know something I don't at all mind admitting it	10. I admit when I do not know something	20	70
11. At times I have really insisted on having things my own way	11. I do things my way	22	77
12. I would never think of letting someone else be punished for my wrongdoings	12. I let someone else be punished for my wrongdoings	24	84
13. There have been times when I was quite jealous of the good fortune of others	13. I am jealous of others with good fortune	28	91
14. I am sometimes irritated by people who ask favours of me	14. I am irritated by people who ask favours	30	98
15. I have never deliberately said something that hurt someone's feelings	15. I say things that hurt others' feelings	33	105

3.4.5 Phase 5: Assembling of the Experimental Integrity Instrument

The experimental integrity instrument is a paper-based test consisting of a front page, which is followed by directions on how to complete the measuring instrument. The instrument also includes a test item example along with the five-point Likert type response scale (ranging from strongly disagree to strongly agree) for respondents to practice on. The instructions are provided on the third page of the instrument. On this page the 132 items are displayed in two columns per page. The total test booklet is five pages long. For the participants' convenience the response scale is provided at the top of each page. The items are distributed in a non-random fashion in accordance with the chronological order of the sub-clusters presented in Table 3.6 (from the Honest facet to the Fair facet). The selected SD items are distributed across the integrity instrument. Appendix A contains an example of the experimental integrity instrument.

Table 3.6: Distribution of Integrity Items

INTEGRITY CLUSTER	ITEM NUMBERS
Integrity Sub-cluster	
Honest	1-6; 8-9
Loyal	10-13; 15-20
Morally Conscious	22-27; 29-34; 36-41; 43
Pretending	44-48; 50-55; 57-60
Responsible	61-62; 64-69; 71-76; 78-83; 85-89
Trustworthy	90; 92-97; 99-104
Truthful	106-114
Fairness Sub-cluster	
Discriminative	115-123
Fair	124-132

3.5 STATISTICAL ANALYSIS

Mellenbergh (1996) explains that classical psychometrics includes models at the level of the observed test score. Classical test theory (CTT; Lord & Novick, 1968) and its extension, generalisability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972), specify models for the observed test score. These models consider the test score to be a continuous variable (Cronbach et al., 1972). CTT was used to analyse the data from the experimental integrity instrument and different statistical techniques were employed. Schumacker (2005) highlights several benefits obtainable through the application of good instructional objectives and item writing using CTT. First, in comparison to item response theory (IRT) models, analyses can be performed with smaller

representative samples of examinees. This is particularly important when field-testing a measuring instrument (Schumacker, 2005). Second, CTT employs relatively simple mathematical procedures and model parameter estimations are conceptually straightforward (Schumacker, 2005). Third, CTT models are often referred to as ‘weak models’ because their assumptions are easily met by traditional testing procedures (Schumacker, 2005).

Rouse (2007) suggests that researchers attempt to gain a ‘feel’ for the data before undertaking any analysis. Researchers should ascertain whether any values look discrepant and whether there is any variation among the parameters that would make the analysis worthwhile (Rouse, 2007). Therefore, the descriptive statistics of the dataset were analysed first in order to describe and summarise the dataset’s basic characteristics. In this study, the dataset was inspected for unexpected responses, extreme skewness and kurtosis, outliers and missing values. Liu, Parelius and Singh (1999) define skewness as a measure of deviation from symmetry. Kurtosis is defined as the ratio of the fourth central moment to the square of the second central moment, and is interpreted as an inverse of the ‘peakedness’ of a distribution. Kurtosis is also used as a measure of the overall spread relative to the spread in the tails. Second, each item’s correlation with the overall total score and the direction of scoring of the experimental integrity instrument was verified by performing a principal components analysis. Principal components analysis is a commonly available extraction technique and is the approach most often used to identify or extract the number of underlying factors or dimensions (Pallant, 2007). Factor extraction involves “determining the smallest number of factors that can be used to best present the interrelations among the set of variables” (Pallant, 2007. p.181). This involves balancing two conflicting needs: (1) the need to find a simple solution with as few factors as possible, and (2) the need to explain as much of the variance in the original data as possible (Pallant, 2007). Third, a principle components analysis was performed on each of the items within each of the sub-clusters of the experimental integrity instrument. This analysis involved inspecting the loadings of each of these items. Fourth, using the CTT approach responses to the items were subjected to hierarchical factor analysis with the aim of examining the underlying dimensionality of the data. It was hoped that the first-order factors would correspond with the sub-clusters of the integrity facet as conceptualisation in phase 1 of the SAPI project. The reason for this is that factor analysis, when employed properly in the construction of psychometric scales and research instruments, facilitates selection of scale items and aids the development of instruments that measure clear, independent dimensions (Tracy, 1990). Principal components analysis was used as the factor extraction method. The number of factors was decided based on the standardised residuals, the scree-plot, parallel analysis and theoretical

expectation. Factors were rotated according to the oblique Direct Quartimin criterion. Finally, the reliability of scores obtained was calculated by means of Cronbach's coefficient alpha. Coefficient alpha (α) is a commonly used index for reliability (Miller, 1995) that can be used for any test on which scores are produced by summing the scores of two or more test items (Miller, 1995).

3.6 CONCLUSION

Chapter three outlined the research design and methodology. Specifically, descriptions of the participants and procedures were provided and the five phases involved in developing the integrity instrument were reported. A description of the statistical procedure followed in the study was also included. The results of the data analysis are discussed in chapter four.

CHAPTER FOUR

RESEARCH RESULTS

4.1 INTRODUCTION

The purpose of this chapter is to present and discuss the statistical results obtained from the experimental integrity instrument data. The instrument was administered to a sample group of 1 023 police reservists. The chapter provides an overview of the descriptive statistics and the various statistical analyses that were performed by making use of the SPSS Version 16 statistical package (Moore & Inc. SPSS, 2008).

4.2 DATA PREPARATION

4.2.1 Unexpected Responses

Before commencing analysis it is vital that the dataset be checked for errors (Pallant, 2007). It is very easy to make mistakes when entering data and some errors can yield very misleading results, especially when performing factor analysis. In this study the data was first screened for errors and outliers. This means that each variable was checked for scores that were out of range or not within the range of possible scores. Each item in the experimental integrity instrument employed a 5-point response scale and therefore the data file should only contain values between 1 and 5. Other values would most likely indicate that a typing error had occurred during the data capturing process. These errors must be identified and corrected before any other analyses are performed. In this case, all of the values were between 1 and 5.

4.2.2 Missing Values

According to De Bruin (2009) it is very common to find that some respondents did not answer an item. Responses may be missing due to a random process. Missing responses could also be indicative of a systematic process. In the case of random missing data it is possible to use the overall pattern of responses to 'predict' how a person would have responded to the item and then replace the missing value with the predicted value. A similar strategy involves the use of a regression approach to replace missing values with the mean of that particular person's responses to the other items. If there are few missing responses either of these approaches will yield satisfactory results. When missing responses are non-random, or when respondents left

relatively large parts of the questionnaire unanswered, it may be best to eliminate those respondents from the dataset. Normally an arbitrary criterion stating that persons with 5% or more missing values should not be included in further analyses is used.

In the dataset in this study, missing values presented a slight problem and had to be addressed before further evaluation could proceed. We employed the strategy of replacing the missing value with the mean of that particular person's responses to the other items.

4.2.3 Descriptive Statistics in terms of Skewness and Kurtosis

Descriptive statistics identify how well an item corresponds to the content of a scale (Taylor, 2008). The items' means and standard deviations provide essential information. The mean indicates the participants' general selection tendency for each item, while the standard deviation points towards the average deviation of responses from the mean of the item (Taylor, 2008). Pallant (2007) explains that descriptive statistics provide information concerning the distribution of scores on continuous variables (skewness and kurtosis). The skewness variable provides an indication of the symmetry of the distribution. Kurtosis provides information about the peakedness of the distribution. In a perfectly normal distribution both skewness and kurtosis would yield a value of 0.

In this research the data was inspected for extreme skewness and kurtosis. A liberal view was taken and only variables with skewness > 2 and kurtosis > 4 were viewed as problematic. Variables with extreme skewness and kurtosis were rejected and excluded from the dataset for further analyses. Most of the items in the dataset were normally distributed. Only 4 items showed unexpected skewness (values lower than 2) while 18 items showed kurtosis values higher than 4. In Table 4.1 the items with extreme skewness and kurtosis are indicated in bold.

Table 4.1

Descriptive Statistics of the 132 Items (N = 1023)

Item	Mean	SD	Skewness	Kurtosis
INTEGRITY				
<i>Honest</i>				
i1INTEGhon	4.25	0.663	-0.795	1.708
i2INTEGhon	4.12	0.771	-0.842	0.946
i3INTEGhon	4.12	0.763	-0.884	1.225
i4INTEGhon	4.29	0.644	-0.779	1.495
i5INTEGhon	4.14	0.652	-0.887	2.875
i6INTEGhon	4.24	0.696	-0.890	1.725
i7SD	4.46	0.587	-0.846	1.479
i8INTEGhon	4.35	0.617	-0.686	1.148
i9INTEGhon	4.11	0.738	-1.001	2.068
<i>Loyal</i>				
i10INTEGloyal	3.48	1.062	-0.465	-0.390
i11INTEGloyal	4.06	0.720	-0.847	1.740
i12INTEGloyal	3.47	0.948	-0.356	-0.004
i13INTEGloyal	2.57	1.095	0.297	-0.540
i14SD	3.92	0.760	-0.718	1.344
i15INTEGloyal	4.14	0.650	-0.765	2.299
i16INTEGloyal	3.80	0.982	-0.658	-0.126
i17INTEGloyal	3.80	0.920	-0.661	0.227
i18INTEGloyal	4.04	0.879	-0.962	0.791
i19INTEGloyal	4.19	0.667	-0.751	1.469
i20INTEGloyal	2.98	0.957	0.084	-0.261
<i>Morally Conscious</i>				
i21SD	4.11	0.897	-1.289	1.856
i22INTEGmorcon	4.20	0.685	-0.923	2.196
i23INTEGmorcon	4.31	0.647	-0.857	1.834
i24INTEGmorcon	4.62	0.533	-1.191	1.851
i25INTEGmorcon	4.36	0.594	-0.662	1.667
i26INTEGmorcon	4.23	0.672	-0.970	2.749
i27INTEGmorcon	4.12	0.772	-1.362	3.548
i28SD	1.76	1.071	1.635	2.048
i29INTEGmorcon	3.78	1.237	-0.962	-0.068
i30INTEGmorcon	3.99	0.809	-0.928	1.382
i31INTEGmorcon	3.47	1.078	-0.537	-0.339
i32INTEGmorcon	4.21	0.685	-0.660	0.863
i33INTEGmorcon	3.93	0.837	-0.885	1.063
i34INTEGmorcon	3.91	1.115	-1.091	0.563
i35SD	4.19	0.755	-1.141	2.309
i36INTEGmorcon	3.20	1.249	-0.120	-0.996
i37INTEGmorcon	3.78	1.001	-0.936	0.509
i38INTEGmorcon	4.04	1.008	-1.366	1.789
i39INTEGmorcon	3.20	1.168	-0.361	-0.865
i40INTEGmorcon	1.86	0.902	1.115	1.174
i41INTEGmorcon	2.92	1.311	0.003	-1.229
i42SD	4.48	0.576	-0.949	2.697
i43INTEGmorcon	4.35	0.674	-1.546	5.404
<i>Pretending</i>				
i44INTEGpretend	2.39	1.092	0.562	-0.481
i45INTEGpretend	2.79	1.132	0.152	-1.004
i46INTEGpretend	1.43	0.670	1.973	5.631
i47INTEGpretend	1.78	0.885	1.256	1.602
i48INTEGpretend	2.09	0.987	0.764	0.000
i49SD	1.93	1.018	0.830	-0.257
i50INTEGpretend	1.74	0.842	1.339	2.073
i51INTEGpretend	2.28	1.195	0.572	-0.810
i52INTEGpretend	1.94	0.988	0.939	0.171
i53INTEGpretend	2.02	1.009	0.797	-0.146

i54INTEGpretend	1.79	0.860	1.242	1.640
i55INTEGpretend	1.74	0.902	1.565	2.710
i56SD	4.26	0.855	-1.627	3.437
i57INTEGpretend	2.11	1.039	0.885	0.091
i58INTEGpretend	2.68	1.138	0.236	-0.991
i59INTEGpretend	1.73	0.665	0.890	1.996
i60INTEGpretend	1.63	0.722	1.343	2.892
<i>Responsible</i>				
i61INTEGresp	2.03	0.989	0.931	0.606
i62INTEGresp	4.53	0.555	-0.990	2.417
i63SD	4.23	0.767	-1.168	2.426
i64INTEGresp	4.43	0.597	-1.079	3.715
i65INTEGresp	4.45	0.619	-1.144	2.884
i66INTEGresp	4.35	0.757	-1.690	4.603
i67INTEGresp	1.81	0.995	1.506	2.002
i68INTEGresp	1.72	0.886	1.781	3.842
i69INTEGresp	4.48	0.710	-1.756	4.718
i70SD	4.27	0.848	-1.711	3.902
i71INTEGresp	2.57	1.076	0.194	-0.709
i72INTEGresp	3.76	0.998	-0.689	0.142
i73INTEGresp	1.96	1.000	1.077	0.803
i74INTEGresp	4.21	0.862	-1.290	2.096
i75INTEGresp	4.38	0.628	-0.933	2.063
i76INTEGresp	4.50	0.618	-1.481	4.712
i77SD	3.61	1.000	-0.705	0.247
i78INTEGresp	1.39	0.756	2.559	7.450
i79INTEGresp	1.94	1.336	1.171	-0.110
i80INTEGresp	4.14	0.853	-1.698	4.106
i81INTEGresp	4.24	0.607	-0.936	4.066
i82INTEGresp	4.18	0.650	-0.987	3.544
i83INTEGresp	3.98	0.761	-1.223	3.111
i84SD	1.48	0.830	2.304	5.948
i85INTEGresp	4.37	0.766	-1.628	4.002
i86INTEGresp	3.55	1.163	-0.718	-0.370
i87INTEGresp	4.26	0.583	-0.414	1.255
i88INTEGresp	4.32	0.614	-0.632	1.314
i89INTEGresp	4.50	0.558	-0.788	1.406
<i>Trustworthy</i>				
i90INTEGtrustw	4.16	0.762	-1.057	2.249
i91SD	1.58	0.894	1.914	3.698
i92INTEGtrustw	4.22	0.749	-1.210	2.720
i93INTEGtrustw	4.17	0.737	-1.335	3.602
i94INTEGtrustw	4.22	0.789	-1.447	3.498
i95INTEGtrustw	1.58	0.786	1.698	3.461
i96INTEGtrustw	3.49	1.219	-0.682	-0.580
i97INTEGtrustw	1.65	0.832	1.713	3.709
i98SD	2.48	1.117	0.459	-0.502
i99INTEGtrustw	1.80	0.960	1.218	1.117
i100INTEGtrustw	1.67	0.842	1.545	2.860
i101INTEGtrustw	3.93	1.165	-1.161	0.518
i102INTEGtrustw	1.98	1.129	1.027	0.115
i103INTEGtrustw	4.14	0.935	-1.491	2.548
i104INTEGtrustw	4.10	0.908	-1.419	2.531
<i>Truthful</i>				
i105SD	1.86	0.947	1.200	1.250
i106INTEGtruth	4.18	0.729	-1.213	3.368
i107INTEGtruth	4.26	0.654	-1.104	3.669
i108INTEGtruth	4.21	0.674	-0.588	0.704
i109INTEGtruth	4.27	0.615	-0.817	2.972
i110INTEGtruth	4.35	0.705	-1.634	5.369
i111INTEGtruth	1.86	0.964	1.213	1.244
i112INTEGtruth	3.86	1.093	-1.149	0.691
i113INTEGtruth	4.26	0.618	-0.682	2.087

i114INTEGtruth	4.12	0.832	-1.124	1.860
FAIR				
<i>Discriminative</i>				
i115FAIRdiscr	1.94	1.053	1.291	1.163
i116FAIRdiscr	1.47	0.800	2.070	4.503
i117FAIRdiscr	3.10	1.475	-0.244	-1.429
i118FAIRdiscr	1.32	0.605	2.549	9.390
i119FAIRdiscr	1.55	0.771	1.801	4.123
i120FAIRdiscr	1.78	0.896	1.261	1.507
i121FAIRdiscr	1.40	0.601	1.711	4.564
i122FAIRdiscr	1.40	0.591	1.644	4.177
i123FAIRdiscr	1.74	0.899	1.340	1.610
<i>Fair</i>				
i124FAIRfair	4.35	0.679	-1.364	4.167
i125FAIRfair	3.81	0.928	-1.081	1.321
i126FAIRfair	4.13	0.852	-1.513	3.155
i127FAIRfair	2.47	1.077	0.400	-0.584
i128FAIRfair	3.88	0.811	-1.167	2.258
i129FAIRfair	4.30	0.728	-1.271	2.985
i130FAIRfair	4.04	0.925	-1.335	1.987
i131FAIRfair	4.36	0.664	-1.430	4.838
i132FAIRfair	4.43	0.590	-0.744	1.170

4.3 HIERARCHICAL STRUCTURE OF THE DATA

Each SAPI cluster consists of sub-clusters. Each of these sub-clusters consists of facets as conceptualised in the qualitative phase of the SAPI project. In order to model this structure and to obtain evidence to justify scoring on the cluster and sub-cluster levels, hierarchical factor analysis was employed in analysing the dataset.

Factor analysis was performed on the item level of the data as a first step and this yielded a set of first order or primary factors. It was hoped that these primary factors would correspond to the SAPI project sub-clusters. The items were all drawn from the integrity cluster domain and it was therefore expected that the primary factors would be correlated. The primary factors were subjected to a second factor analysis in the hope of finding a higher order/second order factor that corresponds with the Integrity cluster. It was hoped that all the primary factors would be strongly related to the higher order factor. However, it was also of interest to inspect the relations between the individual items and the higher order factor. Ideally, each item should have a substantively meaningful relation with the higher order factor.

4.4 ITEM CORRELATIONS WITH FACETS

The correlations of each item were first inspected in relation to the overall total score for the entire set of items. This was done by means of performing a principal components analysis of the entire item and then specifying that only one component should be retained. In addition, the

sign given to each item's loading indicated whether an item needed to be reverse scored in order to calculate a total score. Table 4.2 provides these results.

Table 4.2

Item Correlations with Facets (N = 1023)

Items	Component Matrix
INTEGRITY	
<i>Honest</i>	
i1INTEGhon	0.78
i2INTEGhon	0.77
i3INTEGhon	0.73
i4INTEGhon	0.77
i5INTEGhon	0.63
i6INTEGhon	0.73
i8INTEGhon	0.50
i9INTEGhon	0.69
<i>Loyal</i>	
i10INTEGloyal	0.56
i11INTEGloyal	0.60
i12INTEGloyal	0.56
i13INTEGloyal	0.51
i15INTEGloyal	0.61
i16INTEGloyal	0.63
i17INTEGloyal	0.66
i18INTEGloyal	0.47
i19INTEGloyal	0.65
i20INTEGloyal	0.59
<i>Morally Conscious</i>	
i22INTEGmorcon	0.48
i23INTEGmorcon	0.64
i24INTEGmorcon	0.58
i25INTEGmorcon	0.75
i26INTEGmorcon	0.70
i27INTEGmorcon	0.59
i29INTEGmorcon	0.27
i30INTEGmorcon	0.60
i31INTEGmorcon	0.26
i32INTEGmorcon	0.65
i33INTEGmorcon	0.49
i34INTEGmorcon	0.34
i36INTEGmorcon	0.24
i37INTEGmorcon	0.33
i38INTEGmorcon	0.43
i39INTEGmorcon	-0.002
i40INTEGmorcon	-0.40
i41INTEGmorcon	-0.11
<i>Pretending</i>	
i44INTEGpretend	0.16
i45INTEGpretend	0.20
i47INTEGpretend	0.62
i48INTEGpretend	0.58
i50INTEGpretend	0.72
i51INTEGpretend	0.71
i52INTEGpretend	0.73
i53INTEGpretend	0.74
i54INTEGpretend	0.79
i55INTEGpretend	0.53
i57INTEGpretend	0.50
i58INTEGpretend	0.30

i59INTEGpretend	0.63
i60INTEGpretend	0.66
<i>Responsible</i>	
i61INTEGresp	-0.43
i62INTEGresp	0.61
i64INTEGresp	0.60
i65INTEGresp	0.68
i67INTEGresp	-0.36
i68INTEGresp	-0.42
i71INTEGresp	-0.18
i72INTEGresp	0.47
i73INTEGresp	-0.50
i74INTEGresp	0.51
i75INTEGresp	0.59
i79INTEGresp	-0.26
i82INTEGresp	0.61
i83INTEGresp	0.46
i86INTEGresp	0.07
i87INTEGresp	0.62
i88INTEGresp	0.68
i89INTEGresp	0.70
<i>Trustworthy</i>	
i90INTEGtrustw	0.59
i92INTEGtrustw	0.64
i93INTEGtrustw	0.52
i94INTEGtrustw	0.64
i95INTEGtrustw	-0.56
i96INTEGtrustw	0.34
i97INTEGtrustw	-0.51
i99INTEGtrustw	-0.52
i100INTEGtrustw	-0.62
i101INTEGtrustw	0.47
i102INTEGtrustw	-0.52
i103INTEGtrustw	0.54
i104INTEGtrustw	0.57
<i>Truthful</i>	
i106INTEGtruth	0.70
i107INTEGtruth	0.50
i108INTEGtruth	0.80
i109INTEGtruth	0.71
i111INTEGtruth	-0.53
i112INTEGtruth	0.40
i113INTEGtruth	0.80
i114INTEGtruth	0.64
FAIR	
<i>Discriminative</i>	
i115FAIRdiscr	0.46
i117FAIRdiscr	0.35
i120FAIRdiscr	0.77
i123FAIRdiscr	0.77
<i>Fair</i>	
i125FAIRfair	0.46
i126FAIRfair	0.65
i127FAIRfair	-0.04
i128FAIRfair	0.49
i129FAIRfair	0.73
i130FAIRfair	0.54
i132FAIRfair	0.73

The following items needed to be reverse scored: 40, 61, 67, 68, 95, 97, 99, 100, 102 and 111. Items with very low loadings on the first component were set aside as they correspond to the

other items. The arbitrary guideline used involved rejecting and setting aside those items with a loading < 0.30 . The analysis showed that almost all of the inter-item correlations were considered acceptable and only a few items did not meet the set criteria. Some items from the Morally Conscious (29, 31, 36, 39 and 41), Pretending (44 and 45), Responsible (71, 79 and 86) and Fair (127) facets had scores lower than 0.3 and were therefore removed from the dataset.

4.5 FIRST-ORDER FACTOR ANALYSIS

The goal of a factor analysis is to identify all the major (psychologically meaningful) factors that account for the covariances/correlations of the items (De Bruin, 2009). Various criteria and rules of thumb exist for deciding the number of factors to retain. These criteria include the scree-plot, eigenvalues-greater-than-one-criterion, parallel analysis (for the overhead cluster of Integrity), minimum average partial test and root mean squared residual (De Bruin, 2009). These criteria and rules of thumb tend to work best with scale level data rather than item level data (De Bruin, 2009). In fact, with large numbers of items the eigenvalues > 1 and the scree-plot criteria are unlikely to yield useful information (De Bruin, 2009). Comrey (1988) propose that when analysing items it is best to retain all factors with at least three or four factor pattern coefficients > 0.30 (after rotation) and to reject all factors that do not satisfy this criterion. The rejected factors are too weak to be substantively meaningful and cannot be realistically scored as scales or sub-scales (De Bruin, 2009). Ultimately, psychological meaningfulness rather than mathematical and statistical criteria should carry the most weight in deciding on the number of factors to be retained (De Bruin, 2009).

Eigenvalues of > 1 , the scree-plots were used as criteria to decide on the number of factors to retain for each facet. Parallel analysis was added to these criteria for the whole of the Integrity cluster. Parallel analysis is based on the rationale that factors worth retaining should account for more variance than can be attributed to chance alone (Horn, 1965). The procedure requires that the eigenvalues of the reduced correlation matrix (with communalities in the main diagonal) and the eigenvalues of parallel random data be jointly plotted against the roots (De Bruin, 2006). Only factors with actual eigenvalues greater than the eigenvalues of the parallel random dataset should be retained (Hayton, Allen, & Scarpello, 2004).

4.5.1 The Factor Analysis Results of the Honest Facet

Table 4.3 and Figure 4.1 provide the results of completed analyses of the Honest facet.

Table 4.3

Eigenvalues for the Honest Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.973	49.661	49.661	3.973	49.661	49.661
2	0.821	10.268	59.928			
3	0.722	9.021	68.949			
4	0.617	7.717	76.666			
5	0.551	6.891	83.557			
6	0.489	6.113	89.670			
7	0.455	5.692	95.362			
8	0.371	4.638	100.000			

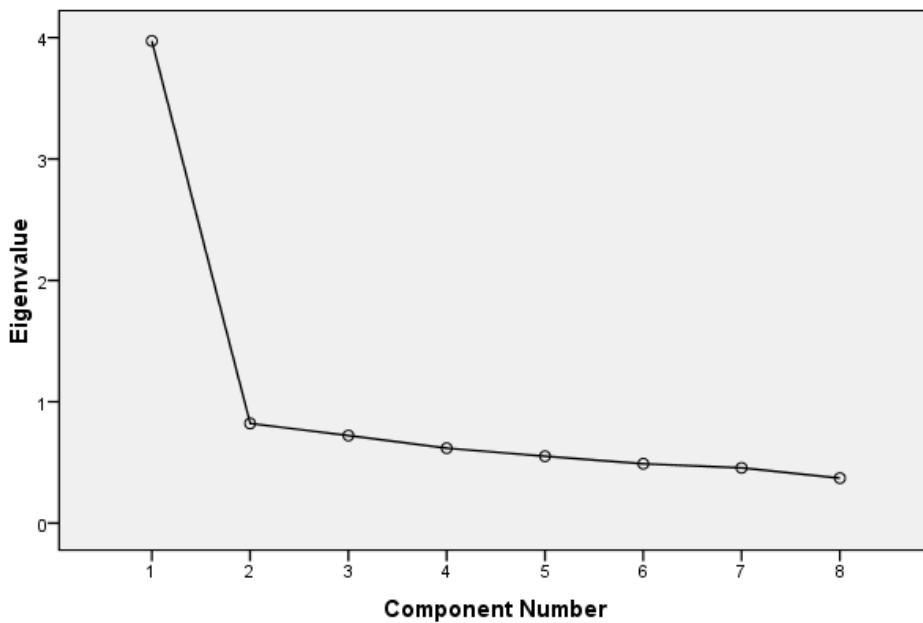


Figure 4.1. Scree-plot for the Honest facet

The results indicated in the eigenvalues and the scree-plot confirmed that only one factor was extracted from the Honest facet in the first-order factor analysis.

4.5.2 The Factor Analysis Results of the Loyal Facet

Table 4.4 and Figure 4.2 provide the results of completed analyses of the Loyal facet.

Table 4.4

Eigenvalues for the Loyal Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.437	34.371	34.371	3.437	34.371	34.371
2	1.265	12.647	47.019			
3	1.016	10.157	57.176			
4	0.799	7.990	65.166			
5	0.755	7.554	72.720			
6	0.690	6.900	79.620			
7	0.562	5.619	85.239			
8	0.543	5.432	90.671			
9	0.501	5.007	95.678			
10	0.432	4.322	100.000			

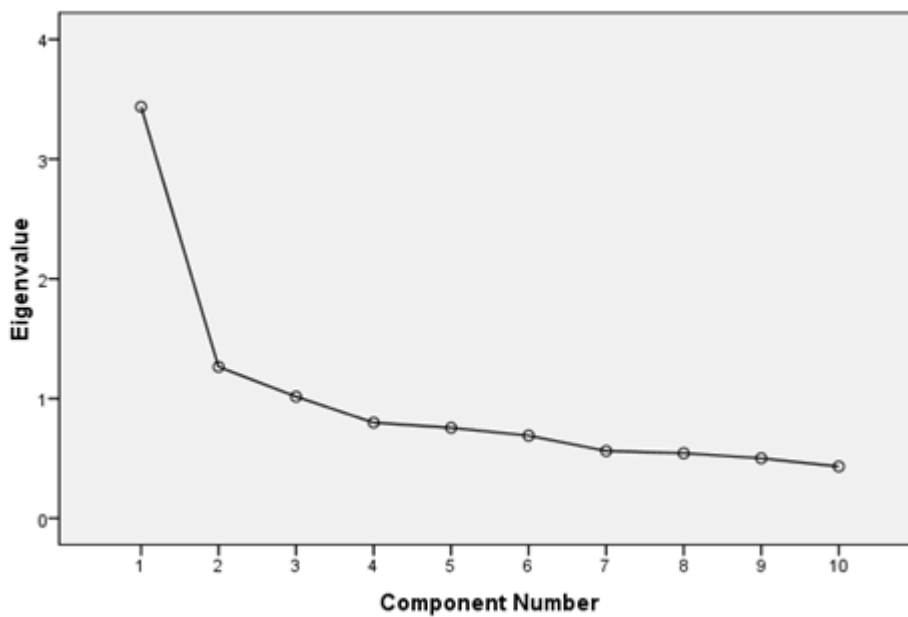


Figure 4.2. Scree-plot for the Loyal facet

The results from the eigenvalues and scree-plot indicated that one factor pattern was extracted from the Loyal facet.

4.5.3 The Factor Analysis Results of the Morally Conscious Facet

Table 4.5 and Figure 4.3 provide the results of completed analyses of the Morally Conscious facet.

Table 4.5

Eigenvalues for the Morally Conscious Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.167	23.152	23.152	4.167	23.152	23.152
2	2.324	12.913	36.066			
3	1.348	7.491	43.557			
4	1.047	5.814	49.371			
5	.969	5.383	54.753			
6	.887	4.929	59.682			
7	.818	4.542	64.224			
8	.772	4.287	68.511			
9	.739	4.103	72.614			
10	.691	3.838	76.452			
11	.653	3.627	80.079			
12	.621	3.452	83.531			
13	.585	3.252	86.783			
14	.564	3.131	89.914			
15	.551	3.062	92.976			
16	.518	2.878	95.855			
17	.387	2.152	98.007			
18	.359	1.993	100.000			

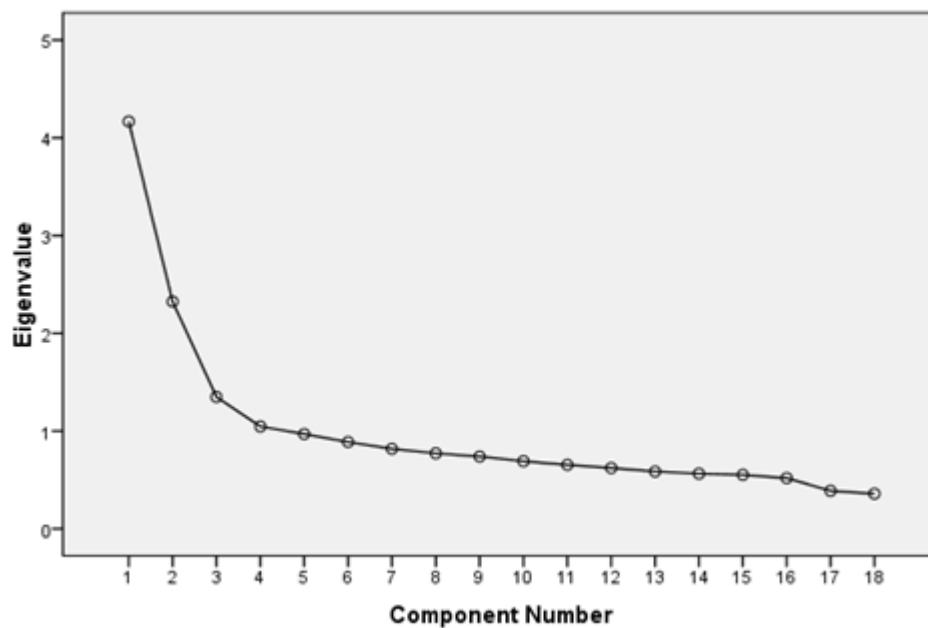


Figure 4.3. Scree-plot for the Morally Conscious facet

According to the results indicated in the eigenvalues and scree-plot, two factors emerged from the Morally Conscious after applying the principal component analysis using the Direct Oblimin rotation method (see Table 4.6 with the values in bold). The second factor of Morally Conscious was labelled Avoidance.

Table 4.6

Component Loadings of Items Comprising the Morally Conscious Facet, Excluding Items 28 and 35

	Pattern Matrix ^a	
	Component	
	Morally Conscious	Avoidance
i22INTEGmorcon	0.470	0.066
i23INTEGmorcon	0.685	-0.086
i24INTEGmorcon	0.566	0.091
i25INTEGmorcon	0.784	-0.035
i26INTEGmorcon	0.744	-0.072
i27INTEGmorcon	0.622	-0.044
i29INTEGmorcon	0.025	0.776
i30INTEGmorcon	0.598	0.039
i31INTEGmorcon	0.041	0.702
i32INTEGmorcon	0.676	-0.046
i33INTEGmorcon	0.474	0.093
i34INTEGmorcon	0.094	0.786
i36INTEGmorcon	0.058	0.583
i37INTEGmorcon	0.246	0.280
i38INTEGmorcon	0.331	0.348
i39INTEGmorcon	-0.080	0.248
i40INTEGmorcon	-0.435	0.074
i41INTEGmorcon	-0.200	0.277

The items (29, 31, 34, 36, 38) that scored < 0.3 when only one factor was extracted now loaded significantly higher on the second factor when two factors were extracted. However, the scores of items 39 and 41 still remained < 0.3 and these two items were removed from further analyses.

4.5.4 The Factor Analysis Results of the Pretending Facet

Table 4.7 and Figure 4.4 provide the results of completed analyses of the Pretending facet.

Table 4.7

Eigenvalues for the Pretending Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.966	35.474	35.474	4.966	35.474	35.474
2	1.285	9.179	44.653			
3	1.094	7.812	52.466			
4	.961	6.862	59.327			
5	.898	6.413	65.740			
6	.765	5.467	71.207			
7	.709	5.063	76.271			
8	.701	5.005	81.276			
9	.636	4.540	85.816			
10	.506	3.613	89.430			
11	.432	3.084	92.514			
12	.413	2.950	95.464			
13	.360	2.572	98.035			
14	.275	1.965	100.000			

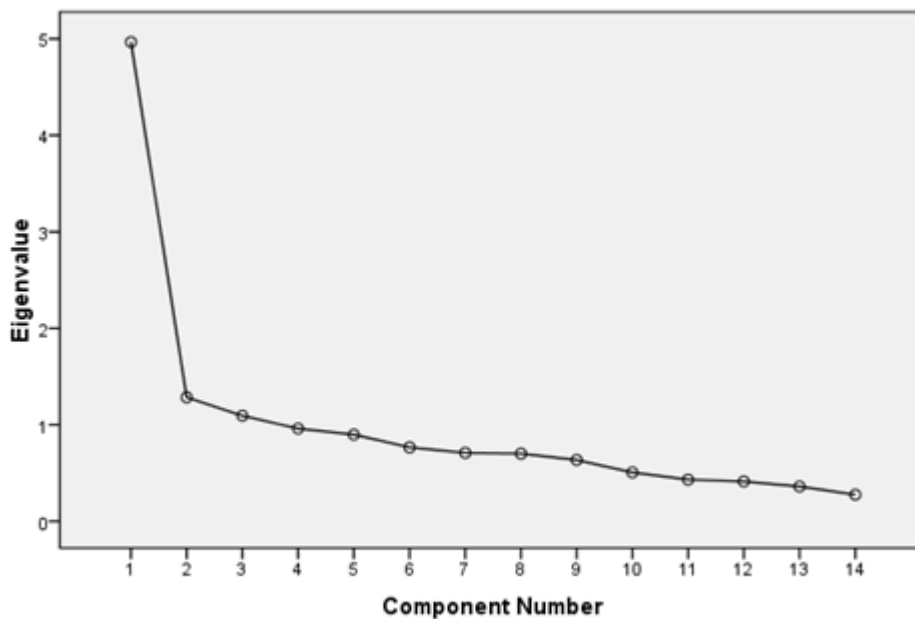


Figure 4.4. Scree-plot for the Pretending facet

The Pretending facet's results indicated that one factor was extracted.

4.5.5 The Factor Analysis Results of the Responsible Facet

Table 4.8 and Figure 4.5 provide the results of completed analyses of the Responsible facet.

Table 4.8

Eigenvalues for the Responsible Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.760	26.447	26.447	4.760	26.447	26.447
2	1.482	8.235	34.682			
3	1.362	7.566	42.247			
4	1.192	6.624	48.872			
5	1.030	5.723	54.595			
6	.896	4.979	59.574			
7	.852	4.733	64.307			
8	.834	4.634	68.941			
9	.776	4.310	73.250			
10	.685	3.804	77.055			
11	.674	3.743	80.797			
12	.631	3.506	84.304			
13	.585	3.250	87.553			
14	.528	2.936	90.489			
15	.465	2.582	93.072			
16	.444	2.469	95.541			
17	.407	2.259	97.799			
18	.396	2.201	100.000			

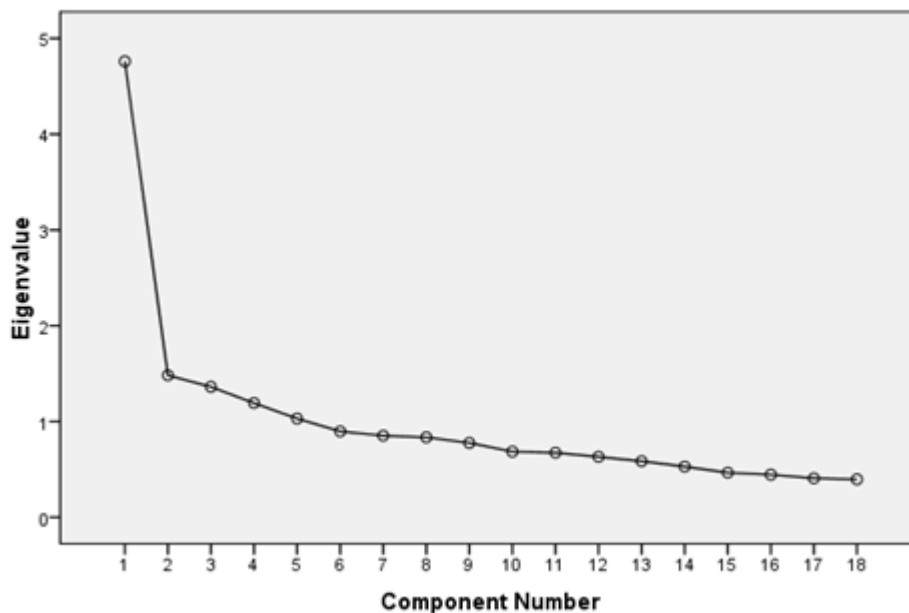


Figure 4.5. Scree-plot for the Responsible facet

The results for the Responsible facet indicated extraction of only one factor.

4.5.6 The Factor Analysis Results of the Trustworthy Facet

Table 4.9 and Figure 4.6 provide the results of completed analyses of the Trustworthy facet.

Table 4.9

Eigenvalues for the Trustworthy Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.882	29.861	29.861	3.882	29.861	29.861
2	1.724	13.263	43.124			
3	1.294	9.950	53.074			
4	1.007	7.749	60.823			
5	.863	6.637	67.460			
6	.650	5.004	72.464			
7	.646	4.971	77.434			
8	.614	4.724	82.159			
9	.542	4.168	86.327			
10	.512	3.936	90.262			
11	.491	3.777	94.039			
12	.461	3.548	97.587			
13	.314	2.413	100.000			

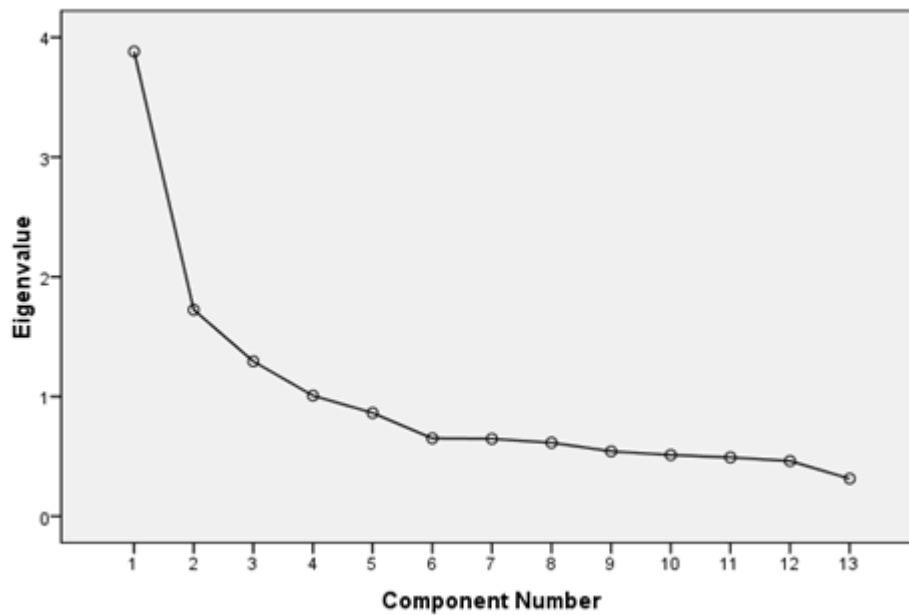


Figure 4.6. Scree-plot for the Trustworthy facet

The Trustworthy facet yielded one factor.

4.5.7 The Factor Analysis Results of the Truthful Facet

Table 4.10 and Figure 4.7 provide the results of completed analyses of the Truthful facet.

Table 4.10

Eigenvalues for the Truthful Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.378	42.230	42.230	3.378	42.230	42.230
2	1.087	13.584	55.814			
3	.848	10.601	66.415			
4	.707	8.837	75.253			
5	.654	8.171	83.424			
6	.560	7.003	90.427			
7	.408	5.099	95.526			
8	.358	4.474	100.000			

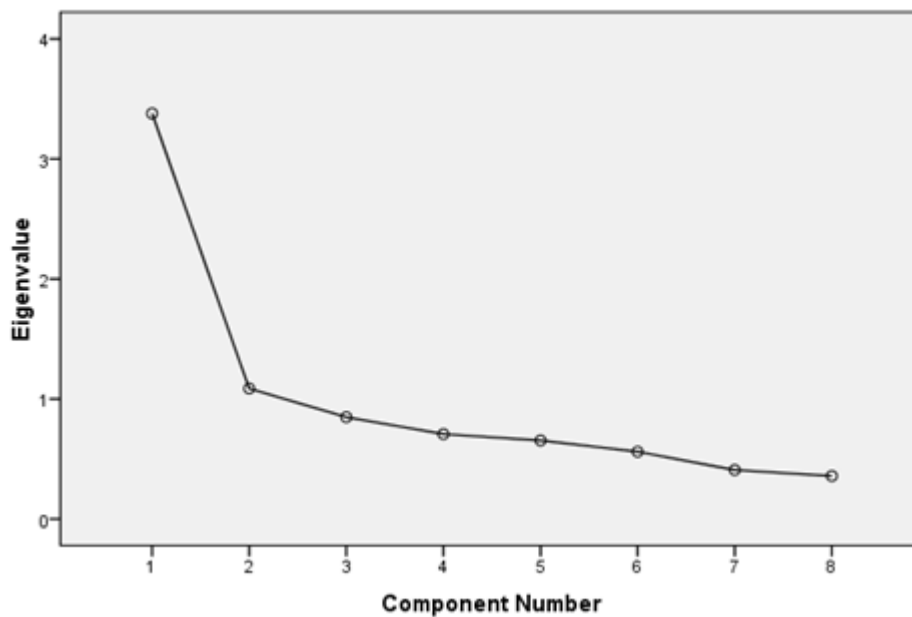


Figure 4.7. Scree-plot for the Truthful facet

The results indicated that one factor was extracted from this facet.

4.5.8 The Factor Analysis Results of the Discriminative Facet

Table 4.11 and Figure 4.8 provide the results of completed analyses of the Discriminative facet.

Table 4.11

Eigenvalues for the Discriminative Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative%
1	1.533	38.322	38.332	1.533	38.322	38.322
2	0.981	24.527	62.849			
3	0.895	22.374	85.223			
4	0.591	14.777	100.000			

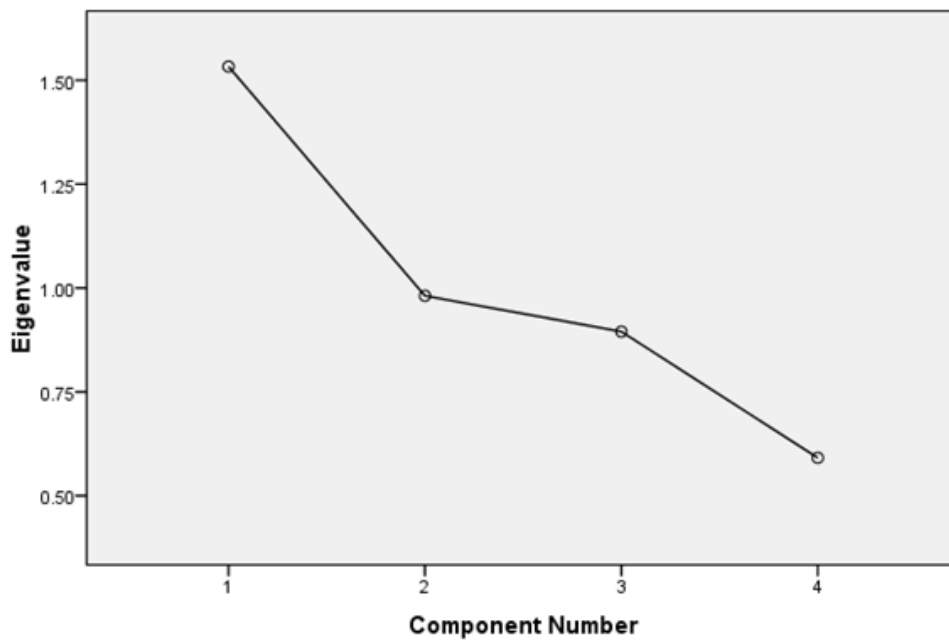


Figure 4.8. Scree-plot for the Discriminative facet

The results for the Discriminative facet indicated that no significant factors could be extracted.

4.5.9 The Factor Analysis Results of the Fair Facet

Table 4.12 and Figure 4.9 provide the results of completed analyses of the Fair facet.

Table 4.12

Eigenvalues for the Fair Facet (N = 1023)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative%
1	2.209	31.562	31.562	2.209	31.562	31.562
2	1.086	15.521	47.083			
3	.946	13.519	60.602			
4	.836	11.942	72.544			
5	.749	10.705	83.249			
6	.674	9.634	92.883			
7	.498	7.117	100.000			

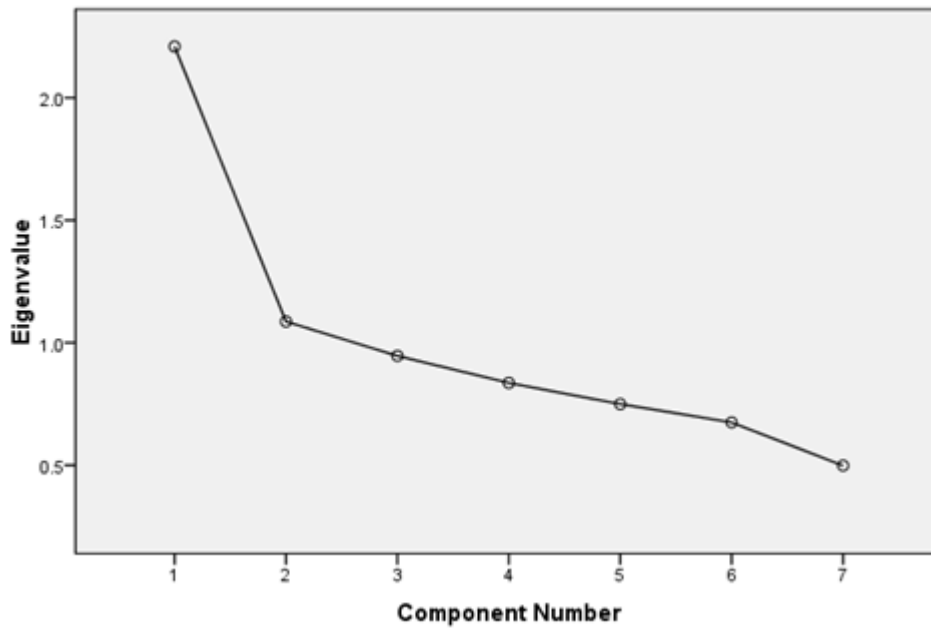


Figure 4.9. Scree-plot for the Fair facet

One factor was extracted from the Fair facet.

4.5.10 The Factor Analysis Results of the Integrity Cluster

Table 4.13 and Figure 4.10 provide the results of completed analyses of the Integrity cluster.

Table 4.13

Eigenvalues for the Integrity Cluster (N = 1023)

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.678	51.982	51.982	4.197	46.632	46.632
2	1.006	11.175	63.157			
3	0.674	7.490	70.647			
4	0.651	7.237	77.884			
5	0.517	5.744	83.629			
6	0.459	5.102	88.730			
7	0.395	4.393	93.123			
8	0.330	3.663	96.787			
9	0.289	3.213	100.000			

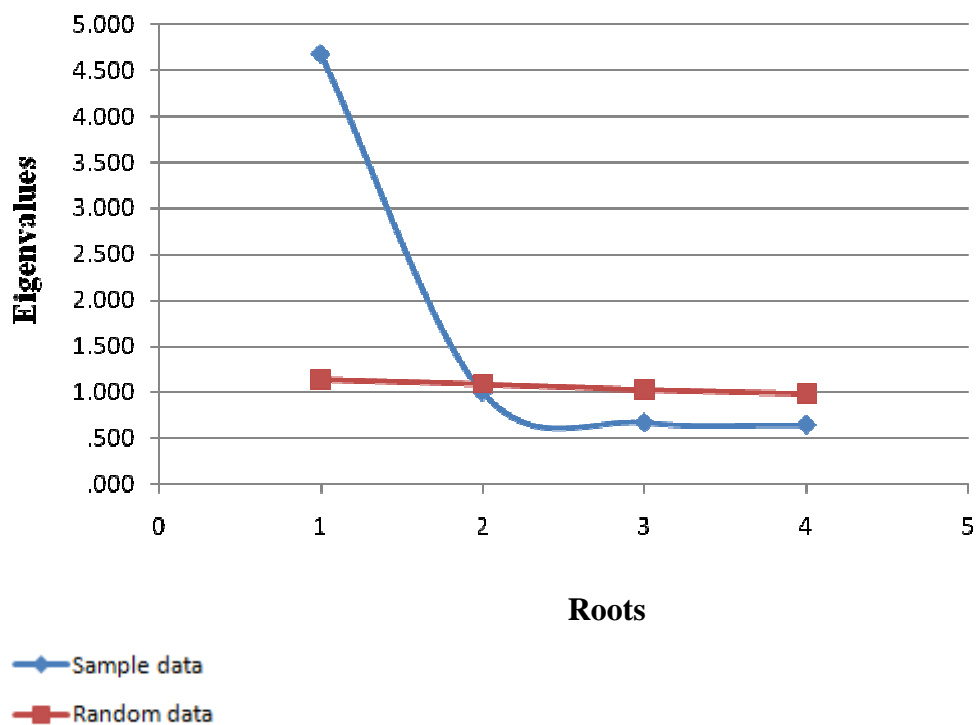


Figure 4.10. Scree-plot and parallel analysis plot for the Integrity cluster

4.6 SECOND-ORDER FACTOR ANALYSIS

The purpose of a second-order factor analysis is to obtain the higher order or second-order factor. This analysis was performed on the factor correlation matrix obtained through the first-order factor analysis. However, this procedure was deemed to be extraneous as all eight facets measured the Integrity cluster.

4.7 RELIABILITY

The internal Cronbach's coefficient alphas were calculated for each facet. The calculation of a reliability coefficient proceeds on the assumption of unidimensionality within all nine facets. The results are reported in Table 4.15.

Table 4.14

Descriptive Statistics, Skewness, Kurtosis, and Cronbach Alpha Coefficients (N = 1023)

Facet	Mean	SD	Skewness	Kurtosis	Cronbach's Alpha
Honest	33.6	3.8	-0.26	0.21	0.85
Loyal	36.5	5.1	0.06	-0.02	0.78
MorallyConscious	53.8	5.3	0.03	-0.10	0.77
Pretending	23.5	7.0	0.53	0.38	0.85
Responsible	59.4	5.5	-0.02	-0.15	0.80
Trustworthy	53.0	6.3	-0.25	-0.03	0.78
Truthful	33.2	3.8	0.02	-0.44	0.76
Discriminative	8.5	2.6	0.10	0.01	0.38
Fair	24.5	2.8	-0.02	0.11	0.62

Table 4.15 indicates that, with the exception of the Discriminative facet, the subscales of the experimental Integrity cluster reflect normal distributions. The Cronbach alpha coefficients show acceptable levels of reliability (higher than the $\alpha > 0.70$ guideline; Nunnally & Bernstein, 1994) with the exception of the Discriminative facet which has a low reliability coefficient (0.38) and the Fair facet which has an average reliability coefficient of 0.62 with only 6 items contributing to the reliability loadings.

The Discriminative facet was removed from the reliability calculations of the Integrity cluster to determine whether the results would be more satisfactory. The results indicated that the overall reliability improved from 0.64 to 0.69.

Although the overall picture of the reliability levels was satisfactory for the preliminary test development stages, a reliability score of 0.95 should be considered in future, especially when the test is used for high stakes selection decisions (Nunnally & Bernstein, 1994).

4.8 CONCLUSION

The purpose of this chapter was to report the results obtained in this study. Items with extreme skewness and kurtosis were rejected and excluded from the dataset for further analyses. Item correlations with facets revealed items that needed reverse scoring. These correlations also

revealed some items with scores < 0.3 and these items were removed from the dataset. The first-order factor analysis specified the number of factors to be extracted for each facet. Seven of the nine facets in the Integrity cluster revealed unidimensionality. However, two factors emerged for the Morally Conscious facet, and no significant factor was extracted for the Discriminative facet. The internal Cronbach's coefficient alphas for most of the facets reflected normal distributions. The exceptions were the Discriminative facet which had a low reliability coefficient and the Fair facet which had an average reliability coefficient.

The following chapter discusses the general conclusions drawn from the research. Recommendations for future research are also included.

CHAPTER FIVE

DISCUSSION AND RESULTS

5.1 INTRODUCTION

This chapter provides a summary of the findings from this study (presented in chapter 4). This is done by linking these results with the original objectives and with the theory and research used to support the arguments in this study. The final section of this chapter addresses the practical implications of the research and provides recommendations for future research.

In order to understand universal and culturally variable personality dimensions a combined emic-etic approach must be used to develop indigenous personality measures (Cheung et al., In press). This approach bridges the divide between mainstream and indigenous psychology and provides a comprehensive framework from which to develop personality measures (Cheung et al., In Press). The SAPI and the CPAI are combined etic-emic measures that include both universal and indigenous constructs by adopting the convergence approach to cross-cultural psychology (Cheung et al., In Press).

The SAPI project is the first South African study that aims to develop a comprehensive questionnaire to assess personality among all South-African language groups. The comprehensiveness of the measure means that it aims to measure all major aspects of personality that are relevant in the South-African context. Therefore, the project started with everyday conceptualisations of personality as found in South-African language groups, rather than with well-known conceptualisations of personality such as the Big Five.

A similar Chinese study conducted by Cheung and colleagues (CPAI; Cheung et al., 1996) found that Chinese personality contained more factors (mainly interpersonal) than typically found in Western studies. In the development of the CPAI and its revised version (Cheung et al., 1996; Cheung, Fan, & To, 2008) universal and indigenous personality traits considered important in Chinese culture were generated in a bottom-up approach. This approach aimed to develop a set of normal personality and clinical scales for comprehensive personality assessment (Cheung et al., In press). Instead of adopting a lexical approach involving identifying personality adjectives from the dictionary, the researchers explored multiple sources for folk description of personality. The sources used included contemporary Chinese novels, Chinese proverbs and the psychological research literature (Cheung et al., In press).

The SAPI project aims to develop an assessment for eleven ethnic and linguistic groups, and not just for a single ethnic and linguistic group as with the CPAI (Cheung et al., In press).

When initially compared to the FFM it seems that Neuroticism, Extraversion, Conscientiousness and Openness are well represented in the proposed SAPI. Agreeableness is not well represented and seems to be somewhat subsumed in soft-heartedness. Interesting, a study of Chinese culture connection also found human heartedness to be an important factor in a large cross-cultural value study (Cheung et al., 1996). Integrity and Facilitating were unique to the South African study, while Relationship Harmony also emerged in the Chinese study (Cheung et al., 1996). Soft-Heartedness (amiability, empathy and active support) appears to have a close link with the concept of Ubuntu (Nel, 2008). It was concluded that social and relational aspects of personality were more salient in the eleven language groups in South Africa than in most Western societies. This finding is in line with studies of Chinese personality (Cheung et al., In press).

Ones (1993) reports that Conscientiousness was found to overlap with the integrity construct. Conscientiousness reflects characteristics such as dependability, carefulness and responsibility (Ones et al., 1993). The integrity-testing literature appears to view and measure the construct of Conscientiousness from a negative perspective (for example, irresponsibility, carelessness and violation of rules; Ones et al., 1993). A facet labelled Responsible appears in the SAPI Integrity cluster. The facet is defined as: "Taking up and living up to responsibilities – with respect to tasks and other people. Acknowledging one's mistakes and not blaming others for those." Similarities exist between the SAPI Responsible facet and the Hard-working, Dedicated, Disciplined and Reckless facets of the Conscientiousness cluster in the SAPI project (Janse van Rensburg, 2010). It is believed that the Responsible facet of the SAPI Integrity cluster may partly correlate with the Conscientiousness cluster.

Correlations with other identified South African integrity measures, such as the Giotto test (Rust, 1999), Integrity Profile-200 (IP-200; Fick, 2002) and the South African Integrity Scale (SAIS; Hunter & Engelbrecht, 2009), would help to investigate the Integrity cluster's construct validity. Construct validity involves quantitative, statistical analysis procedure and "is the extent to which it [an instrument] measures the theoretical construct or trait it is supposed to measure" (Foxcroft & Roodt, 2007, p. 57). Ones (1993) refers to the construct validity of a measurement as "the correlation between the true scores on the measure and the construct that it is intended to measure" (p. 11). Foxcroft and Roodt (2007) and Ones (1993) list a number of

different statistical methods that determine whether a measure actually measures what it is supposed to measure. These statistical measures include correlation with other tests, factorial validity, convergent and discriminant validity, incremental validity and differential validity.

Integrity tests are also designed to measure predictive validity. Integrity tests are specifically developed to assess the integrity, honesty and dependability of applicants, thereby facilitating the prediction of job performance and counterproductive behaviours on the job including theft, disciplinary problems, absenteeism and future on the job dishonest behaviour (Ones et al., 2003). Foxcroft and Roodt (2007) define predictive validity as “the accuracy with which a measure can predict the future behaviour or category status of an individual” (p. 59). Predictive validity is also referred to as the “relationship between the scores of a test (the predictor) [and] the scores of another test (the criterion) measured some time afterwards” (Struwig & Stead, 2001, p. 140). Ones et al. (1993) conducted a comprehensive meta-analysis to investigate whether integrity test validities are generalisable and to estimate differences in validity due to potential moderating influences. The results indicated that integrity test validities are significant for predicting job performance and counterproductive behaviours on the job (Ones et al., 1993). The study also found that integrity tests predict the broad criterion of organisationally disruptive behaviours better than they predict the narrower criterion of employee theft. These results are taken from predictive validity studies conducted on applicants and using external criterion measures (i.e. excluding self-reports; Ones, 1993). Ones et al. (1993) also found that integrity test validities are positive across situations and settings, despite the influence of moderators.

An inspection of the SAPI Integrity instrument shows that all the facets are, to some extent, indicative of counterproductive work behaviour. For example, theft and absenteeism form part of the Pretending and Responsible facets, while disciplinary problems form part of the Responsible facet and future on-the-job dishonest behaviour is incorporated in all the facets.

To summarise, the SAPI is a combined emic-etic measures that includes both universal and indigenous constructs through the adoption of the convergence approach in cross-cultural psychology (Cheung et al., In press). The SAPI is comparable to a similar study conducted in China that resulted in the development of the CPAI. The SAPI structure compares to that of the FFM in various ways. This section also discussed congruencies between integrity tests and general conscientiousness and specifically highlighted congruencies between the SAPI

Integrity dimension and the SAPI Conscientiousness dimension. Construct and predictive validity, and their future impact on the SAPI Integrity instrument, were discussed.

The SAPI experimental integrity instrument was an exploratory study with the aim of developing and clarifying ideas and research questions for a previously unexplored territory. Items were generated from the various sub-facets scales of the integrity instrument once the sub-facet scales had been defined. A pilot study was conducted to standardise the testing procedure of the integrity instrument. Finally, a first draft instrument was developed and applied. The application of this instrument within multi-cultural groups still needs to be determined further as some culture groups (Whites (7%), Coloureds (1.5%) and Indians (0.2%) were under represented compared to the Black group (97.5). The main aim of the experimental integrity instrument was to measure the construct of integrity as conceptualised in the first phase of the SAPI project. The ten facets of the integrity domain were defined by using target behaviours as a guide when constructing the items. Following analyses only eight facets were retained.

5.2 SUMMARY OF FINDINGS

A measuring instrument for an integrity personality construct was developed by formulating items and confirming their correlation with the construct of Integrity and each of the nine facets that constitute that construct.

Firstly, following the preparation of the data for analyses, the correlations between each item and the overall total test score were inspected. Items that failed to relate to the broad cluster of Integrity (< 0.30) were identified and removed from further analyses. Close inspection of the item content suggested possible reasons for the specific items' failures. For example, one of the weak items indicated in the Morally Conscious facet was item 31 (0.26; 'I avoid people who are morally unjust'). A possible explanation for the weak correlation may be that this item does not reflect the meaning of the Morally Conscious facet ('Having good morals and values, a sense of what is right or wrong, and being righteous'). It is also possible that the various cultural language groups had difficulty understanding the meaning of the term 'unjust'. Item 36 ('I end a friendship if he/she does bad things') in the Morally Conscious facet also had weak loadings (0.24). It is possible that this item is actually related to the facet of Loyal, defined as: 'Being an unconditional, true friend who is committed'.

The second factor, Avoidance, which emerged from the Morally Conscious facet will be considered in the next phase of the SAPI project. This factor may be added as a sub-facet within the Morally Conscious facet or may be considered as a facet on its own within the Integrity sub-cluster. It may also be moved to another cluster within the SAPI project.

The Discriminative facet was problematic concerning its 'fit' in the Integrity cluster. Only four items remained in this facet following the removal of items that showed skewness and kurtosis. It is suggested that this facet be permanently removed from the Integrity cluster. Item 127 of the Fair facet ('I am biased against certain people') had a very low score of -0.04. This score appears to result from participants failing to understand the meaning of the term 'biased'. The items that were flagged and removed during analyses should either be revised or moved to a different cluster.

First-order factor analyses assisted in the decision regarding the number of factors to retain for each facet. This decision was guided by the Eigenvalues > 1 , scree-plots and parallel analysis techniques and criteria. The overall results yielded one factor per facet, which indicated that the items were measuring the relevant facet. The results for the Discriminative facet indicated that no significant factor could be extracted. This facet was therefore permanently discarded from the Integrity cluster. After removing the Discriminative facet from the cluster factor analysis, as well as from the reliability analysis the eigenvalues, scree-plot and reliability results were more satisfactory for Integrity.

With the exception of the Discriminative facet ($\alpha = 0.38$) and the Fair facet ($\alpha = 0.62$) the reliability of the facets was acceptable. For the Fair facet only 6 items contributed to the reliability loadings. If an integrity test is to be accepted as reliable it must be consistent in its measurement of what it is supposed to measure (Fick, 2002). Ones and Viswesvaran (1998) state that one of two meta-analytic estimates of the reliability of integrity tests is the internal consistency of items as given by coefficient alpha. High coefficient alphas indicate that test items are inter-related and share a common core construct (Ones & Viswesvaran, 1998). The second meta-analytic estimate of the reliability of integrity tests is the stability of the integrity trait over time as estimated by test-retest reliability (Ones & Viswesvaran, 1998). This estimate of reliability was not addressed in this study, since the experimental integrity instrument is still in its developmental phases. Test-retest reliability is an important consideration for future research.

The concept of integrity remains a difficult concept to grasp. However, the results of this study confirmed that in the South African context integrity involves the character traits of honesty, loyalty, morally conscious, pretending, responsible, trustworthy, truthful and fair.

The development of the SAPI integrity instrument complied with the relevant phases of assessment development outlined by Foxcroft and Roodt (2007). These phases included careful planning, the writing of items, and the administration of the initial version of the instrument to a pilot sample ($N = 174$) in order to determine the effectiveness of the items. In the final phases the items were finalised and the instrument was administered to a representative group of people ($N = 1023$) in order to establish the measure's reliability.

5.3 FUTURE RECOMMENDATIONS

Several areas of future research will contribute to the SAPI project's next phase with regards to the integrity instrument. Firstly, the measuring instrument should be tested for bias and equivalence. Bias and equivalence are of pivotal importance when applying psychological measures in a multicultural society such as South Africa and are key concepts in cross-cultural assessment (Van de Vijver & Tanzer, 2004). Van de Vijver and Leung (1997) hold that bias threatens the equivalence of measurement outcomes across cultures. It is only when instruments are unbiased that measurement outcomes are equivalent and have the same meaning within and across cultures. Equivalence of measures (or lack of bias) is a prerequisite for valid comparisons across cultural populations (Van de Vijver & Tanzer, 2004).

Construct equivalence relates to an instrument's ability to measure the same underlying concept in all groups (Van de Vijver & Leung, 1997), regardless of whether or not the measurement of the construct is based on identical instruments across all cultures. Historically, personality studies showed a strong interest in universality and many researchers employed exploratory factor analysis followed by tests of the similarity of factors obtained in different groups to evaluate the equivalence (Cheung et al., In press). Examples of these studies include the FFM (see McCrae & Allik, 2002; McCrae & Costa, 1987) and Eysenck's Giant Three (Barrett, Petrides, Eysenck, & Eysenck, 1998).

However, equivalence studies can be used to identify both etic (shared, universal) and emic (non-shared, culture-specific) personality constructs. In addition, such studies can identify indigenous components at the item level (Cheung et al., In press). Studies concerning item bias

(or differential item functioning, DIF) identify emic items. In some cases the source of the bias will be technical (such as translation problems), but in other cases the biased item has a specific meaning that is culture specific (Poortinga & Van der Flier, 1988).

Secondly, the data can be further analysed using Item Response Theory (IRT). IRT modelling aims to detect response styles (Taylor, 2008). IRT is an elegant and powerful model of test performance that obviates virtually all of the shortcomings of CTT (Reise & Henson, 2003). Under the IRT approach data can be subjected to a Rasch rating scale analysis (the Rasch rating scale model is one of a family of item response models), with the aim of identifying items that elicit unexpected responses conditional on the latent trait (Taylor, 2008). This analysis will also yield comparisons of item location parameters across cultures, which will allow for the detection of item bias (Taylor, 2008).

Thirdly, Ones and Viswesvaran (1998) consider the stability of the integrity trait over time as estimated by test-retest reliability to be an important meta-analytic estimate of the reliability of integrity tests. High test-retest coefficients indicate that the measurement of the underlying trait is not influenced by temporary changes in a person's state at the time of testing (Ones & Viswesvaran, 1998). It is thus recommended that further studies make use of this instrument in order to determine test-retest reliability.

Fourthly, research can be conducted concerning whether this integrity instrument can predict future training and job performance as well as counterproductive on the job behaviours for all South African language groups.

Fifthly, this integrity instrument should be translated into the various South African languages and administered to first language speakers from each language group. The development of different language versions can also lead to a comparison of these versions to test for equivalence of the facets.

5.4 PRACTICAL IMPLICATIONS

It is obvious that a need exists for the development of indigenous personality measuring instruments that can be administered to all South African cultural and language groups. Indigenous instruments are necessary in order to attain cultural specific personality domains and to avoid omitting important emic constructs when translating imported measures. Cheung

et al. (1996) claim that the development of an inventory that includes major culture-specific personality domains in addition to culture-comparable personality constructs provides a reliable and valid assessment instrument for people of that culture. In the SAPI project a combined etic-emic approach is being used to develop an indigenous personality measure. The approach involves the collection of culturally relevant concepts in order to identify cross-cultural universals as well as culturally unique dimensions. Therefore, the project makes use of a convergence approach (Van de Vijver & Leung, 1997). It is hoped that the combined emic-etic approach of the SAPI project will provide a measure that is sensitive to the indigenous cultural context of South Africa, allows cross-cultural comparison of the meaning of emic or imposed etic traits, extends the interpretation of indigenous traits in a broader cultural context and accommodates the different language groups.

5.5 CONCLUDING REMARKS

The SAPI project is addressing the major problem of a lack of locally developed personality tests within South Africa. This project will not only meet the requirements laid out in the Employment Equity Act but will also be used to assess all cultural and language groups in South Africa.

The experimental integrity instrument yielded promising reliability results and is likely to make a significant contribution to the larger quantitative phase of the SAPI project.

REFERENCES

- Abrahams, F. (1996). *The cross-cultural comparability of the Sixteen Personality Factor Inventory (16PF)*. Unpublished doctoral thesis, University of South Africa, South Africa.
- Abrahams, F. (2002). Fair usage of the 16PF (SA 92) in South Africa: A response to C. H. Prinsloo & I. Ebersohn. *South African Journal of Psychology*, *32*, 58–61.
- Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*, 1321–1353.
- Ashton, M. C., & Lee, L. K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*, 150–166.
- Ashton, M. C., & Lee, K. (2008). The prediction of honesty-humility-related criteria by the HEXACO and five-factor models of personality. *Journal of Research in Personality*, *42*, 1216–1228.
- Ashton, M. C., Lee, L. K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, *87*, 707–721.
- Ashton, M. C., Lee, K., Marcus, B., & De Vries, R. E. (2007). German lexical personality factors: Relations with the HEXACO model. *European Journal of Personality*, *21*, 23–43.
- Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck personality questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences*, *25*, 805–819.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.
- Barry, B., & Stephens, C. U. (1998). Objections to an objectivist approach to integrity. *Academy of Management Review*, *23*, 162–169.
- Becker, T. E. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management Review*, *23*, 154–161.
- Becker, T. E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, *13*, 225–232.
- Benet-Martínez, V., & Waller, N. G. (1997). Further evidence for the cross-cultural generality of the big seven factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality*, *65*, 567–598.
- Bernardi, R. A. (2006). Associations between Hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics*, *65*, 43–53.

- Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119–128.
- Berry, J. W., Poortinga, Y. H., Pandey, J., Dasen, P. R., Saraswathi, T. S., Segall, H. S., & Kagitcibasi, C. (1996). *Handbook of cross-cultural psychology* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology. Research and applications* (2nd ed.). Cambridge, MA: Cambridge University Press.
- Biesheuvel, S. (1943). *African intelligence*. Johannesburg: South African Institute of Race Relations.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335.
- Borofsky, G. L. (1993). Employee reliability inventory screening system. Retrieved July 8, 2009, from <http://www.eri.com/docs/pfactor.pdf>
- Butcher, J. N. (2006) Assessment in clinical psychology: A perspective on the past, present challenges, and future prospects. *Clinical Psychology Science and Practice*, 13, 205–225.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perception. *Journal of Applied Psychology*, 82, 143–159.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233–254.
- Cheung, F. M., Cheung, S. F., Leung, K., Ward, C., & Leong, F. (2003). The English version of the Chinese personality assessment inventory. *Journal of Cross-Cultural Psychology*, 34, 433–452.
- Cheung, F. M., Fan, W. Q., & To, C. W. W. (2008). The CPAI as a culturally relevant personality measure in applied settings. *Social and Personality Psychology Compass*, 2, 74–89.
- Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology*, 29, 233–248.
- Cheung, F. M., Leung, K., Fan, R. M., Song, W. Z., Zhang, J. X., & Zhang, J. P. (1996). Development of the Chinese personality assessment inventory. *Journal of Cross-Cultural Psychology*, 27, 181–199.
- Cheung, F. M., Van de Vijver, F. J. R., & Leong, F. T. L. (In Press). Indigenous Approaches to Personality Measurement: Review and Recommendations.

- Church, A. T. (2000). Culture and personality: Toward an integrated cultural trait psychology. *Journal of Personality, 68*, 651–703.
- Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality, 69*, 979–1006.
- Church, A. T. (2008). Current controversies in the study of personality across cultures. *Social and Personality Psychology Compass, 2*, 1930–1051.
- Claassen, N. C. W. (1997). Culture differences, politics and test bias in South Africa. *European Review of Applied Psychology, 47*, 297–307.
- Cohen, E (1997). Pre-employment integrity testing: The HRVision SavvyIntegrity solution. Retrieved June 18, 2009, from <http://www.hrvision.com/savvyintegrity.htm>
- Collins, J. M., & Schmidt, F. L. (1993). Personality, integrity, and white collar crime: A construct validity study. *Personnel Psychology, 46*, 295–311.
- Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*, 754–761.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354.
- De Bruin, G. P. (2006). The dimensionality of the general work stress scale: A hierarchical exploratory factor analysis. *SA Journal of Industrial Psychology, 32*, 68–75.
- De Bruin, G. P. (2009). *SAPI data analysis manual*. University of Johannesburg.
- De Meijer, L. A. L., Born, M., Van Zielst, J., & Van der Molen, H. T. (2007). *The Construct-driven development of a video-based situational judgment test measuring integrity: A study in a multi-ethnic setting*. Submitted for publication.
- De Raad, B., Barelds, D. P. H., Levert, E., Mlačić, B., Hřebíčková, M., Szarota, P. et al. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology, 98*, 160–173.
- De Raad, B., & Peabody, D. (2005). Cross-culturally recurrent personality factors: analyses of three factors. *European Journal of Personality, 19*, 451–474.
- De Raad, B., & Perugini, M. (2002). *Big five factor assessment: Introduction*. Seattle, WA: Hogrefe and Huber.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

- Downing, S. M., & Haladyna, T. M. (2006). *Handbook of tests development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden 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 dissertation, University of Stellenbosch, Stellenbosch, South Africa.
- Fick, L. (2002). *Manual for the integrity profile-200*. User manual for the measuring instrument Integrity Profile-200. Unpublished manual.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: SAGE Publications.
- Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment, 13*, 229–235.
- Foxcroft, C.D. (2004). Planning a psychological test in the multicultural South African context. *SA Journal of Industrial Psychology, 30*, 8–15.
- Foxcroft, C., & Roodt, G. (2007). *Introduction to psychological assessment* (3rd ed.). Maitland, Cape Town: Oxford University Press.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe, Vol. 7* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Gough, H. G. (1954). *Personnel reaction blank*. Palo Alto, CA: Consulting Psychologists Press.
- Government Gazette, Republic of South Africa, Vol. 400, no. 19370. Cape Town, 19 October 1998.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164–172.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*, 191–205.
- Hogan, R. (1981). *Hogan personality inventory*. Minneapolis, MN: National Computer Systems.
- Horn, J. L. (1965). A rationale and test for the number of factors. *Psychometrika, 32*, 179–185.
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial–organizational psychology: Reflections, progress, and prospects. *Industrial and Organisational Psychology, 1*, 272–290.

- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A., (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581–595.
- Hunter, W. F. J. R., & Engelbrecht, A. S. (2009). *The development of a South African integrity scale*. Research Report, University of Stellenbosch, Stellenbosch, South Africa.
- Inwald, R., Knatz, H., & Shusman, E. (1982). *Inwald personality inventory manual*. Kew Gardens: Hilson Research.
- Janse van Rensburg, C. (2010). *Development of an indigenous conscientiousness measuring instrument for usage in a multi-cultural South Africa*. Unpublished master's dissertation, University of Stellenbosch, Stellenbosch, South Africa.
- Kuncel, N. R., & Tellegen, A. (2009). A conceptual and empirical re-examination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology, 62*, 201–228.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioural Research, 39*, 329–358.
- Lee, K., & Ashton, M. C. (2005). Psychopathy, machiavellianism, and narcissism in the five-factor model and the HEXACO model of personality structure. *Personality and Individual Differences, 38*, 1571–1582.
- Lee, K., Ashton, M. C., & De Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Human Performance, 18*, 179–197.
- Lee, K., Ashton, M. C., Morrison, D. L., Cordery, J., & Dunlop, P. D. (2008). Predicting integrity with the HEXACO personality model: Use of self- and observer reports. *Journal of Occupational and Organizational Psychology, 81*, 147–167.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admission: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442–452.
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*, 1181–1188.
- Lin, E. J. L., & Church, A. T. (2004). Are indigenous Chinese personality dimensions culture-specific? *Journal of Cross-cultural Psychology, 35*, 586–605.
- Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *The Annals of Statistics, 27*, 783–858.

- London House, Inc. (1975). *Personnel selection inventory manual* (12 versions). Park Ridge, IL: Author.
- London House Press (1982). *The employee attitude inventory*. Park Ridge, IL: Author.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- Lousig-Nont, G. M. (1987). *Phase II profile*. Las Vegas, NV: Lousig-Nont & Associates.
- McCrae, R. R., & Allik, J. (2002). *The five-factor model of personality across cultures*. New York: Kluwer Academic/Plenum Publishers.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81–90.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
- McFall, L. (1987). Integrity. *Ethics*, 98, 5–20.
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment*, 11, 265–276.
- Meiring, D. (2007). Bias and equivalence of psychological measures in South Africa. The Netherlands: Labyrinth Publication.
- Meiring, D., Van de Vijver, F. J. R., Rothmann, S., & Barrick, M. (2005). Construct, item and method bias of cognitive and personality tests in South Africa. *South African Journal of Industrial Psychology*, 31, 1–8.
- Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods*, 1, 293–299.
- Mick, D. G. (1996). Are studies of dark side variables confounded by socially desirable responding? The case of materialism. *Journal of Consumer Research*, 23, 106–119.
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2, 255–273.
- Moore, D. S., & Inc. SPSS (2008). *Statsportal for the basic practice of statistics & SPSS CD-ROM* Version 16. USA: W. H. Freeman & Company.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Upper Saddle River, NJ: Pearson Education International.
- Nederhof, A. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15, 263–280.

- Nel, J. A. (2008). *Uncovering personality dimensions in eleven different language groups in South Africa: an exploratory study*. Unpublished doctoral thesis, North-West University, Potchefstroom, South Africa.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McCraw-Hill.
- O'Bannon, R. M., Goldinger, L. A., & Appleby, G. S. (1989). *Honesty and integrity testing*. Atlanta, GA: Applied Information Resources.
- Ones, D. S. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Ones, D. S., Schmidt, F. L., Viswesvaran, C., & Lykken, D. T. (1996). Controversies over integrity testing: Two Viewpoints. *Journal of Business and Psychology, 10*, 487–501.
- Ones, D. S., & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83*, 35–42.
- Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International journal of selection and assessment, 9*, 31–39.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: A meta-analysis of integrity tests. *European Journal of Personality, 17*, 19–38.
- Paajanen, G. (1985). *PDI employment inventory*. Minneapolis, MN: Personnel Decisions International.
- Pallant, J. (2007). *SPSS survival manual* (3rd ed.). Two Penn Plaza, NY: McGraw-Hill House.
- Paulhus, D. L. (1984). Personality processes and individual differences: Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890–904.
- Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-cultural Psychology, 29*, 150–170.
- Poortinga, Y. H., & Van der Flier, H. (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 166–183). Cambridge: Cambridge University Press.
- Reid Psychological Systems (1951). *Reid report*. Chicago: Author.
- Reise, S. P., & Henson, J. M. (2003). A Discussion of Modern versus Traditional Psychometrics as Applied to Personality Assessment Scales. *Journal of Personality Assessment, 81*, 93–103.
- Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology, 85*, 880–887.
- Robins, R. W., Fraley, R. C., & Krueger, R. F. (2007). *Handbook of research methods in personality psychology*. New York: Guilford Press.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes*. San Diego, CA: Academic Press.
- Rouse, S. V. (2007). Using reliability generalization methods to explore measurement error: An illustration using the MMPI-2 PSY-5 scales. *Journal of Personality Assessment, 88*, 264–275.
- Russouw, E. (2005). *Institute a culture of integrity in organisations*. Unpublished master's dissertation, University of South Africa, South Africa.
- Rust, J. (1999). The validity of the Giotto integrity test. *Personality and Individual Differences, 27*, 755–768.
- Sackett, P. R., Burris, L. R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. *Personnel Psychology, 42*, 491–529.
- Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique. *Personnel Psychology, 37*, 221–246.
- Schmidt, M. J., Kihm, J. A., & Robie, C. (2000). Development of a global measure of personality. *Personnel Psychology, 53*, 153–193.
- Schnabel, S., Banse, R., & Asendorpf, J. B. (2006). Assessment of implicit personality self-concept using the implicit association test (IAT): Concurrent assessment of anxiousness and anger. *British Journal of Social Psychology, 45*, 373–396.

- Science Research Associates (1983). *Personal outlook inventory*. Park Ridge, IL: Author.
- Schumacker, R. E. (2005). *Classical Test Analysis*. Unpublished doctoral thesis, Applied Measurement Associates.
- Simms, L. J. (2007). The Big Seven Model of personality and its relevance to personality psychology. *Journal of Personality, 75*, 65–94.
- Struwig, F. W., & Stead, G. B. (2001). *Planning, designing and reporting research*. Cape Town: Pearson Education South Africa.
- Sue, S. (1983). Ethnic minority issues in psychology: A re-examination. *American Psychologist, 38*, 583–592.
- Szirmak, Z., & De Raad, B. (1994). Taxonomy and structure of Hungarian personality traits. *European Journal of Personality, 8*, 95–117.
- Taylor, G., & Gaita, R. (1981). Integrity. *Proceedings of the Aristotelian Society, Supplementary Volumes, 55*, 143–159+161–176.
- Taylor, N. (2008). Construct, item, and response bias across cultures in personality measurement. Unpublished doctoral thesis, University of Johannesburg, South Africa.
- Taylor, N., & De Bruin, G. P. (2005). *Basic Traits Inventory: Technical manual*. Johannesburg: Jopie van Rooyen & Partners SA.
- Tracy, L. (1990). Treating factor interpretation as hypothesis. *Social Behaviour and Personality, 18*, 309–326.
- Trochim, W., & Donnelly, J. P. (2001). *The research methods knowledge base* (3rd ed.). Cincinnati, Ohio: Atomic Dog Publishing.
- Van de Vijver, F. J. R. (2002). Cross-cultural assessment: Value for money? *Applied Psychology: An International Review, 51*, 545–566.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89–99.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54*, 119–135.
- Weekley, J. A., & Jones, C. (1997). Video-based situational judgment testing. *Personnel Psychology, 50*, 25–49.
- Yang, K. S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous of imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58*, 1087–1095.
- Yukl, G. (2006). *Leadership in organizations* (6th ed.). Upper Saddle River, NJ: Pearson.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Journal*, 12, 250–264.