EXPLORING THE CONSTRUCT VALIDITY OF THE SOCIAL DESIRABILITY SCALE OF THE SOUTH AFRICAN PERSONALITY INVENTORY

Genevieve Deacon

Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce in the Faculty of Economic and Management Sciences at Stellenbosch University

SUPERVISOR: PROF. DEON MEIRING

December 2016

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signed: ______Genevieve Deacon Date: December 2016

ABSTRACT

Personality assessments are widely used in South Africa for predicting individuals' future job performance (Barrick & Mount, 1996; Hough, Eaton, Dunette, Kamp, & McCloy, 1990; Ones & Viswesvaran, 2001). The manner in which an individual responds to a personality questionnaire is typically assumed to be an accurate and true reflection of their personality that can be compared with other individuals' personality profiles (Van Herk, Poortinga, & Verhallen, 2004; Ziegler, Maccann, & Roberts, 2012). Personality assessments are typically obtained via self-report inventories and they are therefore susceptible to response bias as well as response distortion. It is therefore important to explore the construct of Social Desirability (SD) within the South African Personality Inventory (SAPI) project with specific focus on the factor structure of the Social Desirability scale.

This study developed two substantive research hypotheses on the factor structure underlying the Social Desirability scale of the SAPI. The SD scale of the SAPI comprises six positively keyed items and six negatively keyed items. It was hypothesised that the positively keyed items would share variance, simply because they share a design feature essentially unrelated to the construct of interest. The same argument was made with regards to the negatively keyed items. Two method factors (a positively keyed factor and a negatively keyed factor) in addition to the two substantive social desirability factors of interest were therefore hypothesised. This allowed the development of four operational research hypothesis on the nature of the measurement model underlying the SAPI SD scale.

Confirmatory factor analysis was used to test the four operational hypotheses. Operational hypothesis **1** was discredited by the results. Operational hypotheses **2** to **4** were all to some degree supported but not unambiguously supported. In the case of all three operational hypotheses several statistical null hypotheses formulated with regards to $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ could not be rejected.

The implications of these results are discussed in the final chapter. There were a few limitations and recommendations for further research that were posed in this chapter. The current study suggests that it is probably most prudent to abandon further attempts to describe the psychological mechanism that produced the observed inter-item covariance matrix obtained for the current SAPI SD scale. The current study suggests that it probably is more prudent to rather focus on the clear conceptualisation of the SD construct, the development of a new SD scale for the SAPI and the psychometric evaluation of the reliability, construct validity and measurement bias of the new SD scale. .

Key words: social desirability, personality assessments, construct validity, SAPI

OPSOMMING

Persoonlikheidsmeting word vry algemeen in Suid Africa gebruik om individue se toekomstige werksprestasie te voorspel (Barrick & Mount, 1996; Hough, Eaton, Dunette, Kamp, & McCloy, 1990; Ones & Viswesvaran, 2001). Dit word tipies aanvaar dat die wyse waarop ‘n individu op ‘n persoonlikheidsvraelys reageer ‘n akkurate en ware weerspieëling is van hul persoonlikheid wat vergelyk kan word met die persoonlikheidsprofiele van ander individue (Van Herk, Poortinga, & Verhallen, 2004; Ziegler, Maccann, & Roberts, 2012). Persoonlikheidsmetings word tipies verkry deur middel van selfrapporteervraelyste en sodanige metings is dus vatbaar vir responssydigheid en responsverwringing. Dit is gevolglik belangrik om die konstruk Sosiale Wenslikheid (SW) in die konteks van die Suid-Afrikaanse Persoonlikheidsvraelys (SAPI) projek te ondersoek met spesifieke verwysing na die faktorstruktuur van die Sosiale Wenslikheid skaal.

Twee substantiewe navorsingshipoteses oor die faktorstruktuur onderliggend aan die Sosiale wenslikheidskaal van die SAPI is ontwikkel. Die SW skaal van die SAPI bestaan uit ses positiewe gesleutelde items en ses negatief gesleutelde items. Die hipotese is geformuleer dat die postief gesleutelde items variansie sou beel bloot omdat hul ‘n ontwerpkenmerk deel wat wesenlik onafhanklik is van die konstruk van belang. Dieselfde argument is aangevoer ten opsigte van die negatief gesleutelde items. Naas die twee substantiewe sosiale wenslikheidsfaktore is twee metode-faktore (‘n positief gesleutelde faktor en ‘n negatief gesleutelde faktor) gevolglik veronderstel. Dit het die moontlikheid geskep om vier operasionele hipoteses te ontwikkel oor die aard van die metingsmodel onderliggend aan die SAPI SW skaal.

Bevestigende faktorontleding is gebruik om die vier operasionele hipoteses te toets. Operasionele hipotese **1** is deur die resultate gediskrediteer. Operasionele hipoteses **2** tot **4** het almal tot ‘n mate steun ontvang maar geen hipotese is ondubbelsinnig gesteun nie. In die geval van al drie hierdie operasionele hipoteses kon verskeie statistiese nulhipoteses gestel met betrekking tot $\mathbf{\Lambda}$ en $\mathbf{\Phi}$ nie verwerp word nie.

Die implikasies van hierdie bevindinge word in die finale hoofstuk bespreek. ‘n aantal leemtes en aanbevelings vir verdere navorsing word in hierdie hoofstuk bespreek. Die huidige studie stel voor dat die mees raadsame weg om te volg waarskynlik sou wees om verdere pogings te staak om die sielkundige meganisme te beskryf wat aanleiding gegee het tot die waargenome interitemkovariansiematrys wat vir die huidige SAPI SW skaal verkry is. Die voorstel van die huidige studie is dat dit waarskynlik meer raadsaam sou wees om te fokus op die duidelike konseptualisering van die SW konstruk, die ontwikkeling van ‘n nuwe SW skaal vir die SAPI en die psigometriese evaluasie van die betroubaarheid, konstrukgeldigheid en metingsydigheid van die nuwe SW skaal.

Sleutelwoorde: sosiale wenslikheid, persoonlikheidsassessering, konstrukgeldigheid, SAPI

ACKNOWLEDGEMENTS

I dedicate this thesis to a woman who has with calmness and strength stood by me through the challenges and struggles I have faced. That woman is my mother, Nicole Deacon. I thank her for all her support, encouragement and love. She made me believe in myself, made me believe that I can do anything I set my mind to. She taught me to be strong, to be determined and to persevere. I am forever grateful for her kind words of encouragement over the last few years.

I would like to thank my supervisor, Prof. Deon Meiring, for his support, commitment and continuous feedback each step of the way. I would like to thank the SAPI collaborators for the opportunity and the privilege of being involved with this project and providing me with the SANPAD SAPI bursary to complete my studies. The aim of the SAPI and the impact it will have on the South African psychometric assessment industry, specifically personality assessment, is something with which I am proud to be involved.

A special acknowledgment goes to Prof. Callie Theron; who assisted me in the last stages of my thesis. I value your knowledge and expertise and the exceptional impact that you had on this thesis. Your willingness and enthusiasm to assist is valued and appreciated.

I would like to thank the rest of my family and my friends for believing in me and in my ability to complete this thesis. Thank you for the supportive and encouragement during this time. A special thanks to Megan van Gesselleen for your understanding of my challenges, your endless inspirational words and the undeniable support that you have for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTORY ARGUMENT AND RESEARCH OBJECTIVES

1.1 INTRODUCTION AND JUSTIFICATION FOR THE STUDY

Psychological assessments, which are frequently used in South Africa, are employed to collect information about psychological traits such as personality (Foxcroft & Roodt, 2009). Since the twenty first century, multiculturalism has become the norm in countries across the globe (Foxcroft & Roodt, 2009). The processes of globalisation and migration have contributed to the multicultural nature of populations in many countries, including South Africa (Van de Vivjer & Rothmann, 2004). This has resulted in a number of different ethnic groups constituting the South African population. The impact of these trends on personality assessments in South Africa is yet to be determined.

Most of the personality measures currently in use in South Africa were developed in the United States of America or the United Kingdom and this creates problems for the use of such instruments across the different ethnic groups in South Africa (Foxcroft & Roodt, 2009). This has implications for the cultural appropriateness of personality measures in cross-cultural measurement (Lotter, 2010). The items in the subscales measure the various latent personality dimensions. The concern is that the behavioural denotations that were earmarked to be reflect individuals standing on specific latent personality dimensions serve as behavioural expressions of those latent personality dimensions only in some cultures but not in others. Whether different cultures attach the same connotative meaning to the personality construct and the latent personality dimensions it comprises (i.e. whether it is possible to identify universally generalisable traits Taylor, 2008) is a different question.

The development of personality assessments in South Africa began in a time characterised by unequal distribution of resources due to a political environment called Apartheid (put in place by the National Party from 1948 until 1994), which was based on previous discriminatory colonial laws. This presented a possible and very probable, situation in which respondents to a personality assessment would utilise a different frame of reference when answering questions impacted by cultural, environmental, socioeconomic, and educational factors resulting from Apartheid (Meiring, 2007). That which serves as behavioural denotations of a given latent personality dimension in one group might not be expressions of that personality trait in another group. The transition to democracy in 1994 changed the manner in which personality tests were viewed in South Africa (Meiring, 2007). Previously the use of separate tests for different groups was regarded as acceptable and the development of dedicated tests for different groups was actively pursued by the Human Sciences Research Council (HSRC) during the Apartheid years (Owen & Taljaard, 1988). The need for personality assessments that are applicable to all South Africans and that showed no construct bias[1] across cultural and ethnic groups was important as these assessments were being used to predict job-performance during the selection phase of employment for all South Africans. The use of personality instruments in personnel selection and decision-making has increased as these instruments have been shown to successfully predict job performance and future behaviour (Barrick & Mount, 1996; Hough et al., 1990; Ones & Viswesvaran, 2001). In a meta-analysis study conducted by Meiring, Rothmann, Van der Vijver, and Barrick (2005) on personality and the prediction of job performance it was found that the Five Factor Model (FFM) is a valid predictor of job performance, especially with individuals that have a qualification of Grade 12 or higher. Emotional Stability ($\rho = 0.19$) showed the highest validity, and was followed by Extraversion ($\rho = 0.17$) and Conscientiousness ($\rho = 0.12$)).

Evidence that personality allows for the valid prediction of job performance is valuable in the work environment, as it highlights the need for valid assessments as a necessary condition for equity and the efficient management of selection and personnel development (Plug, 1996). This need has received attention in the new democratic South Africa and legislation has been adopted that forbids any discriminatory practices based on assessment results in the workplace. These pieces of legislation (see Mauer, 2000, and Meiring, 2007) include the Constitution of the Republic of South Africa Act no. 108 of 1996 (hereafter referred to as the Constitution), the Labour Relations Act no. 66 of 1995 (LRA no. 66 of 1995; hereafter referred to as the LRA) and

[1] It would at the same time be desirable that personality measures should also not display any item bias across cultural and ethnic groups in the interest of good workmanship. The absence of item bias is, however, not a necessary condition to avoid predictive bias when using an personality measures in a valid actuarial selection procedure (Theron, 2009). If consistent uniform or non-uniform item bias would exist in the items of any given subscale it would systematically negatively bias the observed subscale scores for a specific group and thereby influence the regression of the criterion on the subscale score for that particular group. The difference in the regression of the criterion on the predictor can, however, be accommodated in the actuarial prediction model.

the Employment Equity Amendment Act 47 of 2013 (hereafter referred to as the EEA no. 47 of 2013).

These legislations require personality assessments to be culturally appropriate (Van de Vijver & Rothmann, 2004). These legislations also ensure that the development of new tests ensure that there is no longer a discrepancy between the different cultural groups, thereby creating a situation in the testing arena where all test-takers have an equal likelihood of obtaining scores on a test that are a true representation of their skills, abilities and personality. This study focused on personality assessments because they are widely used in South Africa (Van de Vijver & Rothmann, 2004). According to Nel (2008), the main problem currently facing the use of personality assessments in South Africa is the non-psychometrically appropriate applications of these assessments. Not all South African cultural groups were represented when imported personality assessments were standardised (Laher & Cockroft, 2011). In order to overcome this challenge assessment practitioners need to ensure that culturally appropriate tests are developed. These tests will then meet the psychometric standards as well as ensure that various versions of tests are available in the participants first language (Laher & Cockroft, 2011). The second point is particularly important given the multi-lingual nature of the South African society (Cheung, 2004; Foxcroft, Paterson, Le Roux, & Herbst, 2004).

The South African Personality Inventory project set out to address the concerns regarding personality assessment in South Africa and develop a psychometric measure of personality that complies with the legislative stipulations. The SAPI[2] project addressed the issue of cultural differences in personality assessment by developing an entirely new indigenous measurement instrument specifically for South Africa. The SAPI project was launched in 2005 and is making progress in developing a personality measurement instrument that is valid and reliable across all the cultures presented by the 11 official languages in South Africa. The final aim of the project is to produce a unified personality inventory that takes into consideration both universal (etic) and culture-specific (emic) personality factors across the eleven official language groups in South

[2]The SAPI, an acronym for the South African Personality Inventory, is a project that aims to develop an indigenous personality measure for all 11 official languages in South Africa. Participants are Byron Adams (University of Johannesburg and University of Tilburg, the Netherlands), Carin Hill (University of Johannesburg), Deon Meiring (University of Pretoria), Jan Alewyn Nel (North-West University), VelichkoValchev (Tilburg University, the Netherlands), and Fons van de Vijver (North-West University, Tilburg University, the Netherlands, and University of Queensland, Australia).

Africa (Meiring, 2007). The project includes two phases. In the first phase the indigenous perceptions of personality were explored with the intention of deriving authentic and relevant personality factors across the language groups by using information that was collected from semi-structured interviews (Nel, 2008). In the second phase the descriptive terms collected in the first phase were converted into a validated quantitative inventory measuring the personality factors prevalent in the South African context (Meiring, 2012).Determining the validity and reliability of this instrument as well as the factor structure of personality within South Africa were also explored in phase two.

A respondent's score on a personality assessment can be compared to other respondents' scores if it is assumed an accurate and true reflection of personality (Van Herk et al., 2004; Ziegler, Maccann & Roberts, 2012). The SAPI relies on self-report and respondents could therefore make use of self-enhancement strategies including responding in a manner that misrepresents their behaviour or attributes (Barger, 2002; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). This is known as response bias and is formally defined as the systematic tendency to distort responses to an assessment that results in observed scores on the test being unrelated to the true score of the individual (Fischer, 2004; Van de Vivjer, & Leung, 1997). An individual's response style is defined as the situation where the respondent consistently displays bias across situations or items. A variety of response styles have been identified, including acquiescence (tendency to agree to items), extremity (tendency to choose the extreme ends of the response scale irrespective of item content) and social desirability (tendency to present oneself favourably) (Odendaal, 2013).

This study focused specifically on the social desirability (SD) response style. Personality assessments often include a scale that measures social desirability (also referred to as impression management) and Industrial Psychologists use these scales to identify candidates who are adjusting their personality scores as well as to flag potentially invalid test profiles (Odendaal, 2013).This is also the case with the SAPI.

Social desirability responding occurs due to cultural norms about the desirability of particular values, attitudes, opinions, interests, behaviours or traits (Steenkamp, De Jong, & Baumgartner, 2010). Various terms have been used to describe the construct of social desirability including impression management (Ferrando, 2008; Paulhus, 1984), faking (Barrick & Mount, 1996;

Hough, 1998; Hough et al., 1990; Ones et al., 1996; Rosse, Stecher, Miller, & Levin, 1998), self-deception (Paulhus, 1984) and self-enhancement (Heine & Lehman, 1997). Although these terms are often used interchangeably there are clear differences in meaning and application (Li & Bagger, 2006; Ones et al., 1996).

The exploration of the construct of social desirability is an important aspect of the SAPI project. The social desirability scale of the SAPI should provide a reliable, construct valid and unbiased measure of the social desirability construct as it is constitutively defined by the SAPI. This should then successfully serve the function of flagging candidates that might not have responded authentically to test items. If respondents do not respond authentically they rather consciously or unconsciously attempted to portray themselves in a more socially acceptable manner. This study will permit further research on what exactly determines the degree to which test respondents respond in a socially desirable manner. It is important that the social desirability scale included in the SAPI is a construct valid, unbiased measure of the social desirability construct as the SAPI constitutively defines it. It also is important that it adheres to current legislation in its use so as to ensure that is use is beneficial to the field of psychological assessment in South Africa.

1.2 RESEARCH OBJECTIVES OF THIS STUDY

The objective of the research study is to investigate the reliability and construct validity of the Social Desirability (SD) scale for the SAPI. The SD scale of the SAPI is based on a specific conceptualisation of social desirability. Specific social desirability dimensions are distinguished in terms of this interpretation. The architecture of the SD-scale reflects a specific design intention. The SD-scale of the SAPI was developed with the intention that specific items should reflect test respondents standing on specific latent social desirability dimensions. Specific items were therefore selected for each subscale because they are believed to reflect in a relatively uncontaminated manner a specific latent social desirability dimension and only that dimension. The scoring key of the SD-scale of the SAPI reflects the expectation that all items comprising a specific subscale should show high loadings on a single dominant factor. This implies that the items can be used to obtain an observed score for that specific latent social desirability dimension, and that dimension only. When computing a subscale score for a specific social desirability dimension, only the items comprising that specific subscale are combined. This does not imply that the latent social desirability dimension are not allowed to correlate and therefore to a certain degree share variance. A very specific measurement model is thereby implied in which each specific latent social desirability dimension reflects itself primarily in the specific items written for the specific subscale.

The objective of this research study is therefore to evaluate the fit of the first-order measurement model of the SD-scale of the SAPI, as implied by the architecture of the scale and the constitutive definition underlying the social desirability construct, on a relatively large sample of South African citizens.

1.3 STUDY OUTLINE

Chapter 1 gave an outline of the justification for the study with reference to the rationale. It further stipulated the specific objectives that will be pursued. Chapter 2 reviews existing literature concerning psychological assessment in South Africa, provides an introduction to cross-cultural personality assessment as well as an overview of the South African Personality Inventory (SAPI) project to date. The main focus of the literature review is the discussion on social desirability in the context of the SAPI. The chapter also contains a review of the current social desirability literature. The construct is defined and the various factor models are described. The link between social desirability and personality is discussed. The various ways in which social desirability has been measured in personality assessments is also discussed.

Chapter 3 discusses the research design and methodology used in the study. It unpacks how the data will be analysed using item analysis and factor analytic techniques.

Chapter 4 presents and discusses the results of the statistical analysis performed in this study. These results are presented in four broad categories, namely item analysis of the items of each subscale, exploratory factor analysis of each of the subscales of the SD scale, and confirmatory analysis of the factor structure implied by the constitutive definition of the social desirability construct and the architecture of the SD scale of the SAPI. Chapter 5 provides an integrated discussion of the results for the study linking the findings to the literature that was presented in Chapter 2. This final chapter also discussed the limitations and recommendations.

1.4 SUMMARY

Psychological assessment in South Africa is governed by various laws relating to the use and administration of assessments. Most of the personality measurement instruments currently used in South Africa have been imported from Western countries. South Africa does not have personality tests that have been developed, neither made available or normed for all South African cultural groups. The SAPI project aims to address this gap by identifying the structure of South African personality and developing an instrument to measure this personality structure. The SAPI project has recently conceptualised a six-factor model. This study focused on the factor structure of the Social Desirability scale in order to determine whether the SD scale of the SAPI provides a construct valid measure of the social desirability construct as conceptualised by the SAPI.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

The goal of testing and measurement is to reflect some facet of the world with the expectation that what these tests reflect can be generalised to the broader society in which they are administered (Claassen, 1995). Psychological testing plays a fundamental role in making certain professional decisions, such as selection and promotion decisions. It is therefore imperative that such testing is always informed by the results of basic research in the field of psychology (Huysamen, 1979; Murphy & Davidshofer, 2005). Researchers in the field of psychology are constantly trying to understand the complexity of human behaviour. They do so with the purpose of identifying those specific attributes underpinning a particular behaviour that are predictive of a desired outcome. Inferences can then be made during selection decision-making regarding the extent to which this behaviour will be displayed in practice (Murphy & Davidshofer, 2005). According to Foxcroft and Roodt (2005) psychological testing, such as personality assessments, measures attributes manifested in the behaviour of individuals.

This study aims to determine whether the construct of social desirability, as conceptualised by the SAPI, is measured in a construct valid manner amongst South Africans through the behavioural denotations included in the SAPI SD scale. In the sections that follow, the history of psychological assessment in South Africa is explored. The concept of social desirability is discussed in detail with emphasis on definitions, different models as well as the ways in which social desirability is typically measured. The development of the SAPI project and the progress thus far are also discussed, along with an explanation regarding the development of the Social Desirability scale to be used in the SAPI.

2.2 PSYCHOLOGICAL ASSESSMENT IN SOUTH AFRICA

Psychological assessments were developed in an environment characterised by unequal distribution of resources based on ethnic classification. A situation that was vastly different to that of the USA and UK from where the assessments instruments were imported, which were then adapted for the South African context (Foxcroft & Roodt, 2009). According to Claassen (1997) it is impossible to discuss testing in South Africa without taking into consideration the political, economic and social history, which was substantially impacted by apartheid. During the apartheid era psychological tests were developed along particular and separate cultural lines (Owen, 1991). While a reasonable number of psychological tests were developed for the White (minority) group, considerably fewer tests were developed for the Black (majority) group (Foxcroft, 1997). Discriminatory practices flooded the testing arena because these two ethnic groups were seen as separate under the apartheid regime. For these discriminatory practices to be addressed, separate assessments should be replaced by measures that are designed to redress past injustices (Huysamen, 2002).

The socio-political situation in South Africa began to change during the 1980s and this change accelerated in the early 1990s when discriminatory laws were revoked and applicants from different ethnic groups began competing for the same jobs (Foxcroft & Roodt, 2009). However, apartheid policies and legislation continued to govern the manner in which test development was approached (Foxcroft, 1997). According to Foxcroft (1997) it is important to note that very few new culturally relevant tests were developed due to apartheid. The history of the development and use of psychometric assessments in South Africa has thus been tainted by a legacy characterised by segregation, resulting in the fact that these tests are culturally insensitive and inappropriate (HPCSA, 2006). Meiring and colleagues (2005) supported this notion as they acknowledged that in the 1980s aspects of unfairness, bias and discriminatory practices being employed received attention. Psychologists in South Africa today are aware of the advantages and importance of improving test development by creating, as well as utilising tests that are cross-culturally valid and unbiased and that can therefore be used in a fair manner for multiple ethnic groups (Nel, 2008; Paterson & Uys, 2005). There are four major ethnic groups in South Africa namely; Blacks who are from African descent, Coloureds who are from a mixed descent, Asians/Indians who are from Asian descent and Whites who are from an European descent.

There are 11 official languages spoken in South Africa of which nine are Bantu languages spoken by the Black group and two European languages (English and Afrikaans) spoken as a first language by the other ethnic groups.

South African psychologists are increasingly seeking to use measures that were developed for the South African context and that have norms for each of the ethnic groups, thus ensuring that test performance is interpreted in relation to an appropriate norm group (Foxcroft & Roodt, 2009). A study conducted by Paterson and Uys (2005) found that the inequalities concerning the cross-cultural application of tests could be addressed by ensuring appropriate norming of assessment scores. They further concluded that this would be a step in the right direction for South Africa[3].

There are test developers in South Africa who have attempted to adapt and norm international tests in order to make them applicable to the South African context (Foxcroft, 2004). This is characteristic of a new era in psychological test development where tests are being adapted or developed with a multicultural perspective in mind (Foxcroft, 2004; Paterson & Uys, 2005). Psychologists recognise that existing tests are inappropriate for the multicultural context and therefore acknowledge the necessity of addressing past discrimination by aligning the practice of testing to legal demands (Foxcroft, 2004; Meiring et al., 2005; Taylor, 2008; Van de Vijver & Rothmann, 2004). Aligning assessment with legislation involves either developing new instruments or validating existing instruments for use in South Africa (Foxcroft, 2004; Meiring et al., 2005; Paterson & Uys, 2005; Taylor, 2008; Van de Vijver & Rothman, 2004).

However, since the first democratic election practitioners have been faced with a range of ethical and legal issues in psychological assessment (Bartram, 2004). The application, control and development of assessment shave become a disputed terrain controlled by legislation (Foxcroft & Roodt, 2009). According to Mauer (2000) certain legislative acts (see section 1.1 of this document) regulate and monitor the use of assessments administered to individuals. These pieces of South African legislation have shaped the practice of practitioners in the field of psychological

[3] The use of different ethnic norms can be justified if evidence of statistically significant differences in latent means on the latent variables being measured exist and if evidence of the absence of construct and item bias exists. Separate norms can, however, be contentious in as far as it is reminiscent of earlier apartheid practices. Separate construct-referenced norms can still be a meaningful solution (although a combined norm would also be desirable), a single combined criterion-referenced norm table, that acknowledges differences in the criterion-predictor relationship across groups, is the only viable option.

assessment (Paterson & Uys, 2005). The Constitution, which is designed to serve as the safeguard of a vast number of human rights, provides for the development of national legislation pertaining to labour matters (Mauer, 2000). The LRA no.66 of 1995 and the EEA no. 47 of 2013 were developed based on certain constitutional prescriptions. These items of legislation emphasise fair employment practices as well as the development of equity in the workplace. The spirit embedded in the EEA no. 47 of 2013 prescribed the equitable use of psychological tests (Paterson & Uys, 2005). More specifically, it included a section that governs the use of psychological tests and other assessment instruments in the sphere of work (Mauer, 2000).

EEA no. 47 of 2013 (section 8) stipulated that (Republic of South Africa, 2014, p. 6): “Psychological testing and other similar tests 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, (c) is not biased against any employee or group and (d) has been certified by the Health Professions Council of South Africa established by section 2 of the Health Professions Act, 1974 (Act No. 56 of 1974), or any other body which may be authorised by law to certify those tests or assessments”.

The last clause (clause d) was recently accepted as an inclusion to Section 8 of the EEA no. 47 of 2013 (Republic of South Africa, 2014).There are professionals in the industry who are opposed to the inclusion of the latter amendment. Theron (2010) noted that the inclusion imposes a requirement that will be practically impossible to implement and execute. Test developers and distributors were also not consulted regarding the inclusion of this clause. In order for the main players in the industry to agree and apply a new law, a certain level of ‘buy-in’ is needed and this was not received (Theron, 2010). People Assessment in Industry (PAI) is a group that looks after the interests of the Society for Industrial and Organisational Psychology (SIOPSA) and their key aim is to promote, advise on and support the use of valid, reliable and fair assessments in the workplace as per the EEA no. 47 of 2013. They have also criticised the inclusion of the above amendments as they believe the changes will have a significant practical and ethical implication for practitioners, the organisations for which these practitioners work, the Professional Board of Psychology (the Board) and the profession as a whole (SIOPSA, 2013). The implication of this legislation is that Industrial Psychologist will no longer have the autonomy to practice within their scope. Organisations using Industrial Psychologists will not scrutinise the use of assessments from a risk and compliance perspective as they would need to carry the costs of classifying tests before they are administered or they will require Industrial Psychologists to provide the classification certificate of assessments. This poses a risk to the organisation, as a classified test may be less suitable and reliable for the type of assessment that needs to be conducted than a non-classified test. The amendment places additional responsibility in the hands of the Board in that it will be held jointly liable should a test be found not to be reliable and fair. The implication for the profession is that the legislation may result in less employment opportunities being available. This could result in an increase in the usage of ad-hoc assessment methods, which are bad practice and are conducted by poorly training practitioners, ultimately discrediting the profession and science of psychometrics (SIOPSA, 2013).

JvR Psychometrics (previously known as Jopie van Rooyen & Partners) is a company that has been developing and distributing assessments in South Africa for more than two decades (JvR Psychometrics, 2014). They pride themselves on being a company that provides top quality assessments to appropriately qualified clients. The new inclusion in the EEA no. 47 of 2013 has resulted in a few practical dilemmas for JvR Psychometrics. JvR has concurred with the issues highlighted by SIOPSA (2013) and has further stated that some of the assessments they distribute are not on the gazetted list as they are still collecting South African research data in order to be able to submit the assessment to the Psychometrics Committee of the Professional Board for Psychology of the HPCSA. The company uses assessments that have been validated for the South African industry and thus adhere to the EEA no. 47 of 2013 requirements. However, many of these assessments have not been submitted to the Board due to administrative issues and they can only be submitted when the Board provides clarification and guarantees of their processes. There is a concern that the Board may not have the time and capacity to provide informed evaluations of assessments. JVR are also concerned that the standards against which the submitted tests will be compared have not been made public and there is no clarity as to which tests will not be classified as psychological.

There are still practitioners who have taken it upon themselves to address past discrimination and develop indigenous South African psychological measures that take cultural and language issues into account. The SAPI project aims to develop an indigenous personality measurement that consists of an inclusive approach to accommodate the dynamics of the South African multicultural context. Within this indigenous personality measurement there is a scale that measures social desirability and the construct validity of this scale is of interest to this study. Therefore, the construct of social desirability is discussed in detail in the section which follows. The different models are referred to and the measurements for social desirability are discussed.

2.3 SOCIAL DESIRABILITY

It is important to ascertain whether the answers respondents give on any psychological measure are true reflections of their standing on the latent variable of interest or manipulations of the truth (Van Herk et al., 2004). Response biases are a concern in psychological assessment, particularly when using self-report measures; as individuals respond provide with their own responses to questions measuring their own traits, attitudes and behaviours (Paulhus, 1991).

The concept of social desirability was initially developed over 50 years ago when interviewers made the general observation that what respondents say in interviews is not always entirely true and their answers appear to display a consistent distortion of reality (Johnson & Van de Vijver, 2003). In the 1950s psychologists noticed that the desire to be seen as socially acceptable was a major factor in how people answered psychological tests. Psychologists recognised that socially desirable tendencies have a potentially distorting influence on scores based on an individual's responses to statements in a personality measure (Edwards, 1957). Meehl and Hathaway (1964) reported the tendency in the 1960s, but suggested that at the time very few systematic efforts had been directed towards doing anything about this response tendency.

Fischer (2004) described response bias as a distortion in responses so that the observed score is unrelated to the true score of the individual. The way in which a person responds to items on a test may be due to pressures the situation creates, the nature of the items that the scale uses or the fact that the items themselves introduce error variance when comparing personality across cultures (Grimm & Church, 1999; Odendaal, 2013; Paulhus & Reid, 1991). Cross-cultural researchers, especially those who utilise questionnaires to collect data, have noted that there are cultural variations in several types of response bias (Smith, 2004). Response variations occur because respondents react to items on a questionnaire in a manner that they deem natural or appropriate to their culture (Smith, 2004). Individuals often select the answer to a question that they deem to be flattering or socially acceptable rather than the answer that they deem to be accurate (Paulhus, 1991).

In cross-cultural research, particularly those studies that deal with social variables (such as personality), respondents' scores are compared at face value and an individual's score is expected to relate meaningfully to the individual's future behaviour (Van Herk et al., 2004). This could result in erroneous decisions being made when comparing the test scores of different individuals as the responses could be somewhat biased (Taylor, 2008). Test scores are assumed to be an accurate indication of that individual's standing on a particular attribute. If responses to items are inaccurate and faked then individuals' scores are not true reflections of their standing on the construct (Ziegler et al., 2012). It is therefore important that the presence of response bias in the data is investigated because making erroneous decision has an impact on an individual's future.

There are three prominent response biases, namely social desirability, acquiescence and extremity response bias. This study focuses on the response tendency known as social desirability, which is commonly referred to as the tendency to give answers that make the respondent look good (Paulhus, 1991). Acquiescence is the tendency to agree rather than disagree with items on a questionnaire regardless of item content and is also termed agreement tendency or yea-saying (Van Herk et al., 2004). Extreme response bias is the tendency to make use of extreme response categories on a rating scale regardless of item content (Paulhus, 1991). In the section that follows definitions of social desirability are discussed along with the different models of social desirability. The different scale formats that are used to detect social desirability tendencies in personality assessments are also discussed.

2.3.1 Defining Social Desirability

The definition of social desirability has been much debated in recent research on personality assessment. Espinosa and Van de Vijver (2014) conducted a meta-analysis of the construct and concluded that the debate is due to the ambiguous nature of the construct. According to Taylor (2008) social desirability is one of the most researched response styles and is also the most debated response style.

In his seminal book on the social desirability variable, Edwards (1957) posed an interesting argument. He noted that the descriptors of personality can be characterised in terms of the position the statements have on a single dimension that he termed the social desirability-undesirability scale. He originally defined a socially desirable response as a true response to statement that has a socially desirable scale value or as a false response to a statement that has a socially undesirable scale value (Edwards, 1957). This is similar to Nederhof's (1985) definition of social desirability, which states that individuals either claim socially desirable or deny socially undesirable traits. Therefore, Edwards defined social desirability as the scale value for any personality statement. The scale value indicates the position of the statement on the social desirability continuum. Kuncel and Tellegen (2009) demonstrated that the response options present for each item on a personality inventory can be scaled in relation to their rated desirability's. According to Kuncel and Tellegen (2009) this means that individuals appear to weigh a number of factors when selecting a particular response option that they believe to be most desirable in any given situation. The thought process and the reason behind choosing either a true or a false response to a statement as well as the reason why an individual either claims or denies certain attributes are of major interest when trying to decipher the substance behind socially desirable responses.

The diversity in research approaches to social desirability has resulted in a variety of operational definitions that are used interchangeably (Odendaal, 2013). Crowne and Marlowe (1960) defined social desirability as "a need for social approval and acceptance and the belief that this can be attained by means of culturally acceptable and appropriate behaviours" (p. 353). Social desirability, claiming unlikely virtues, denying faults that are common, exaggerating strengths, portraying a good impression, self-enhancement and faking are terms used to describe the socially desirable response distortion (Ones et al., 1996). Each of these terms relate to the fact that the respondent is in some way or another concealing the truth or acting under a facade (Furnham, 1985). The term social desirability is specifically used as a general term to refer to tendencies to distort self-reports in a favourable direction (Paulhus, 1991). According to Holden and Book (2012) faking, or social desirability, has three key features. The first feature relates to the response being intentional, the second feature relates to displaying some degree of deception and the last feature is that social desirability is orientated towards others (Holden & Book, 2012).

What these definitions have in common is the idea that individuals want others to view them in a positive manner; in other words, they want to be admired and favoured by others. This is a natural human need that is referred to as the need for affiliation and adoration. It is normal for an individual to want the approval of others. It is therefore unsurprising that some individuals deny their negative attributes and highlight their positive traits in order to gain acceptance. Distorting answers on personality questionnaires can be seen as one way for these individuals' strengths to be illuminated and their weaknesses to be less noticeable. While this can be characterised as faking it can also be characterised as a natural response satisfying the basic need to be accepted by significant others. According to Holden and Book (2012) although individuals who fake on personality assessments will score high on a social desirability scale, not all extreme scorers on a social desirability scale are necessarily fakers. It would be ironic if honest, respondents that truly approximate the ideal that most others only aspire to be were characterised as liars (Furnham, 1985). Although it is possible to argue this point quite convincingly it is also important to investigate the underlying meaning of social desirability to better understand why individuals respond desirably. A better understanding of social desirability can be achieved by investigating the various models of social desirability identified by researchers. It is interesting to note how these theories have evolved and how the models have been expanded.

2.3.2 Different Models of Social Desirability

The lack of clear dimensionality has been a concern since the early days of measuring social desirability (Uziel, 2010). A milestone in the history of socially desirable responding was the articulation of the dimensions. Most researchers in this field have adopted a two-dimensional factorial structure of social desirability. Wiggins (1964) factor analysed social desirability scales and found two factors, which were termed Alpha and Gamma. There is consensus in the literature that there is a distinction between social desirability scales that measure impression management (Wiggin's Gamma factor) and scales that measure self-deception (Wiggin's Alpha factor) (Paulhus, 1984). One of the most influential models of social desirability is based on the Alpha and Gamma factors, which are labelled self-deception and impression management. Self-deception refers to the tendency to view the self in an overly positive light and impression management is the tendency to distort responses to create a favourable impression (Paulhus, 1984). In this section various models for social desirability are described and the various terms for social desirability are discussed in relation to the numerous factor models.

2.3.2.1 Two-Factor Models

Social desirability was initially conceptualised as a one-dimensional construct and scales were then developed to measure this construct. Due to the lack of correlation between the scales various two-factor models of social desirability were formulated (Steenkamp et al., 2010). Although Edwards (1957) viewed social desirability as the extent to which behavioural personality denotations were undesirable or desirable expressed as a scale value on a bi-polar continuum (undesirable or desirable) other researchers have used two-factor models to explain social desirability. The difference between Edwards (1957) and the researchers proposing two-factor models, however, go further than mere a difference in opinion about the number of dimensions involved. Edwards (1957) and also Nederhof (1985) regard social desirability as firstly descriptive of the behavioural denotations used to assess a person's standing on the various latent personality dimensions and the social desirability of a respondent's test responses then depends on the social desirability of the items he/she choose to endorse and not to endorse. The researchers advocating a two-, three- or four-factor model of social desirability view social desirability as a multi-dimensional abstract theme in a bundle of related behaviours. This also is true with regards to some of the one-dimensional scales examined by Wiggins (1964).

One of the first two-factor models was developed by Damarin and Messick in 1965 and provides a detailed theoretical interpretation of two distinct factors of social desirability. Their two-factor model is depicted in Figure 1.The first factor (Factor 1 in Figure 1) involves distorting the privately held self-image of the individual. This distortion is as a result of a bias that an individual is employing that is associated with a global belief regarding personality traits (Paulhus, 2002). This factor was labelled autistic bias in self-regard (Damarin & Messick, 1965). The second factor (Factor 2 in Figure 1) relates to a naive tendency to promote a reputation that is desired by others and is labelled propagandistic bias. The underlying motivation for propagandistic bias relates to factors varying from social approval to habitual lying (Paulhus, 2002), which are characteristics of the definitions of social desirability discussed in the previous section.

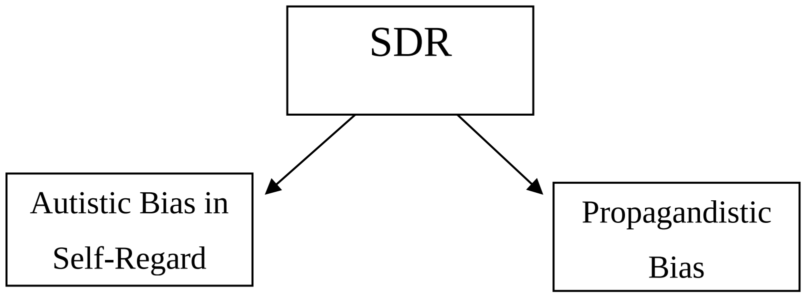

Figure 1: Damarin and Messick's Dimensions of SDR
(Source: Adapted from "Socially desirable responding: The evolution of a construct." By D.L. Paulhus (2002) p. 58)

Sackeim and Gur (1979) also developed a two-factor model of SDR. Their two-factor model is depicted in Figure 2.This model distinguishes between the constructs of self-deception and other-deception. With self-deception the respondents report unrealistic positive self-descriptions about which they appear to be totally convinced, while with other-deception the respondents consciously and deliberately distort their self-description to mislead an audience (Paulhus, 2002).

The first type of social desirability described by Espinosa and Van de Vijver (2014) and Ellingson, Sackett and Smith. (2001), a form of social desirability in which the respondent is convinced that the self they are portraying when answering a self-report measurement is in fact their true self even if others disagree, can be linked to the self-deception factor identified by Sackeim and Gur (1979). Self-deception is measured by including items in the questionnaire that relate to the psychodynamic notion that certain undesirable thoughts (sexual and aggressive) are experienced but often denied (Sackeim & Gur, 1979). If respondents overreact to this offensive content then they are seen to have self-deceptive tendencies (Paulhus, 2002).

Responding in an other-deception manner occurs when the respondent aims to mislead others by lying or faking answers (Ellingson et al., 2001; Espinosa & Van de Vijver, 2014). To measure other-deception, items are constructed to describe desirable behaviours that are so public and blatant that they are not subject to self-deception (Sackeim & Gur, 1979). When an individual makes excessive claims of such behaviours it is assumed that this probably involves some form of conscious dissimilation (i.e. untrue exaggerations; Paulhus, 2002).

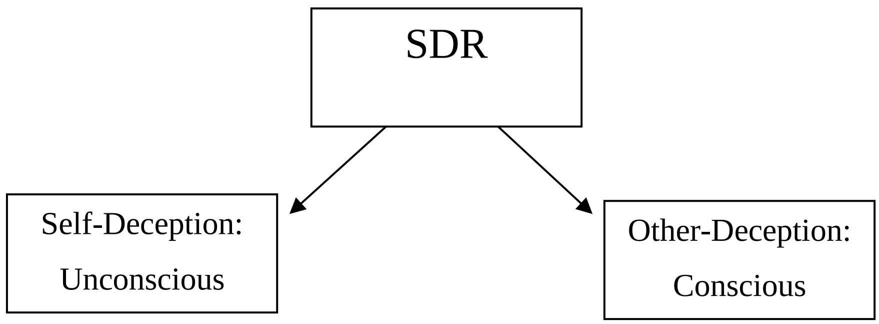

Figure 2: Sackiem and Gur's Dimension of SDR
(Source: Adapted from "Socially desirable responding: The evolution of a construct." By D.L. Paulhus (2002) p. 59)

Self-deception can be linked to Damarin and Messick's (1965) autistic bias in self-regard, as both these factors are based on an individual's tendency to distort the self so that they can protect themselves from being labelled as abnormal with regard to human behaviours that are in fact normal. Other-deception can be linked to Damarin and Messick's (1965) propagandistic bias, in that the individual responds in a manner that is in line with what is deemed to be proper, appropriate or accepted in their community.

The final two-factor model of social desirability was developed by Paulhus (1984) and attempts to link and integrate the concepts and measures developed by Sackeim and Gur (1979) with the structure provided by Damarin and Messick (1965) (Paulhus, 2002). Factor analyses conducted on social desirability scales suggested the existence of two major factors (Paulhus, 1984) that are best interpreted as self-deception and impression management (previously labelled other-deception; Sackeim & Gur, 1979). Paulhus and John (1998) have repeatedly found empirical evidence for these two factors.

According to Paulhus (1984) impression management refers to a conscious and intentional distortion of responses to create a favourable impression that is related to the traditional view that the respondent deliberately alters their answers to present a positive social image (Ferrando, 2008). Research suggests that there is a distinction between the self-deception and impression management components of social desirability (Paulhus, 1984). Self-deceptive positivity refers to an honest but somewhat overly positive view of the self, whereas impression management occurs when the respondent consciously dissembles when responding (Paulhus, 1984; 1991; 2002). The term impression management is favoured over other-deception because the term other-deception implies deliberate lying, which Paulhus (1984) suggested was too presumptuous. Paulhus (1984) agreed with Damarin and Messick's (1965) argument and suggested that the habitual presentation of a specific positive public image could be construed as a meaningful personality construct rather than a response bias. Figure 3 below provides a visual depiction of Paulhus' (1984) model.

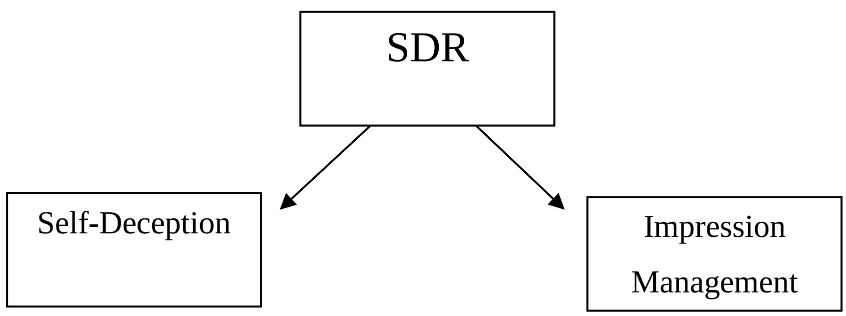

Figure 3: Paulhus' Two Factor Model of SDR

(Source: Adapted from "Socially desirable responding: The evolution of a construct." By D.L. Paulhus (2002) p. 59)

Extensive examination of the social desirability construct has, however, suggested that the two-factor models are overly simplistic and further exploration is required (Paulhus, 2002). This has resulted in the development of three-factor models, which are discussed below.

2.3.2.2 Three-Factor Models

Paulhus and his research team (2002) conducted factor analyses on items of the Balanced Inventory of Desirable Responding (BIDR) and identified one factor for impression management (IM) and two factors for self-deception. Based on these findings the two-factor model was expanded to include two dimensions for one of the factors. Paulhus and Reid (1991) thus concluded that self-deception items can be divided into two distinct factors. The two clusters for the self-deception items appear to involve enhancement, which relates to promoting positive qualities, and denial, which relates to disowning negative qualities (Paulhus, 2002). Figure 4 below presents a depiction of Paulhus' refined three-factor model of SD (adapted from Paulhus, 2002). The two self-deceptive factors could be seen as involving attribution responses that relate to claiming desirable characteristics that enhance an individual's positive features and as denial responses that involve denying negative attributes (Paulhus, 1984). Using the term 'socially desirable' it can be argued that an individual will respond to personality test items using these three response styles to portray themselves in a manner that they view as most 'wanted/desired' by significant others. In other words, individuals can unconsciously distort the favourability of questionnaire responses in two ways, in that they can "*exaggerate their talents and minimise their sins*"(Paulhus & John, 1998, p. 1038).

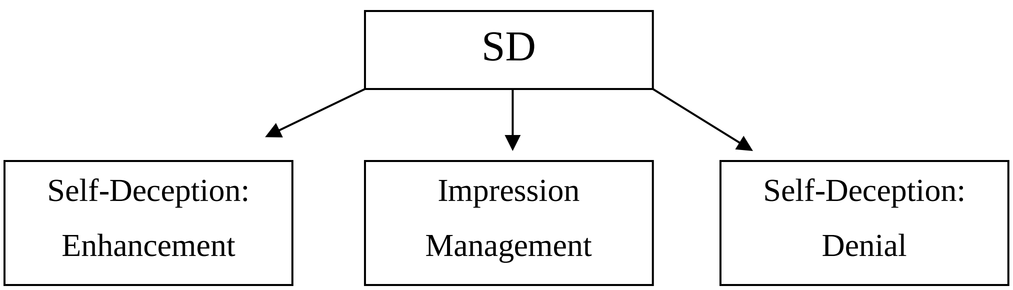

Figure 4: Extended Version of Paulhus' Model of SDR
(Source: Adapted from "Socially desirable responding: The evolution of a construct." By D.L. Paulhus, 2002 p. 61)

Despite the use of different terms the extended version of Paulhus' model of social desirability articulates well with the two-factor models reported earlier. Damarin and Messick's (1965) autistic bias in self-regard is linked to Sackeim and Gur's (1979) self-deception and Paulhus' (1984) self-deception as all three of these constructs measure an individual's unconscious awareness of how they claim positive attributes and deny negative attributes. In the same way, Damarin and Messick's (1965) propagandistic bias is linked to Sackeim and Gur's (1979) other-deception and Paulhus' (1984) impression management. These three constructs all refer to the individuals' tendency to twist the truth about themselves and their abilities in order for others to perceive them as better than what they are in order to make a good impression on some significant other.

The two- and three-factor models discussed above laid the foundation for understanding the conceptualisations of various researchers (Damarin & Messick, 1965; Paulhus, 1984, 1991, 2002; Sackeim & Gur, 1979). Given the ambiguous nature of social desirability it is not surprising that there are different models of social desirability. The section below details further expansion of the two- and three-factor models of social desirability.

2.3.2.3 Four-Factor Models

Paulhus' Two Tier System of Social Desirability is currently the most comprehensive and inclusive model of social desirability[4]. Recent research has increasingly focused on two content domains in which social desirability may be exhibited rather than on emphasising the distinction between forms of social desirability based on the level of conscious awareness (Steenkamp et al., 2010). According to Paulhus and John (1998) self-favouring response tendencies are best understood in the context of two methods (modalities) that are fundamental to human experience. Paulhus and John (1998) referred to these methods as agency and communion. Agency-related modalities involve being dominant, assertive, autonomous, influential, independent and powerful whereas communion-related modalities involve affiliation, intimacy, belonging, love, approval and nurturance. These two modalities impel two corresponding motives, namely the need for power (agency) and the need for approval (communion). According to Paulhus and John (1998) social desirability in agency contexts relates to egoistic response tendencies (ERT) and social desirability in communion contexts relates to moralistic response tendencies (MRT). An egoistic bias is associated with a self-deceptive tendency where an individual tends to exaggerate social and intellectual status, which further leads to unrealistic but positive self-perceptions related to agency traits such as dominance, fearlessness, intellect and emotional stability. Individuals with high scores on this dimension tend to have a narcissistic and somewhat 'superhero' quality (Paulhus & John, 1998). A moralistic bias is associated with a self-deceptive tendency that results in the individual denying socially deviant impulses and claiming attributes that are self-

[4] It needs to be noted that the this may not be the ultimate conceptualisation of social desirability and further research may indicate a structure different to Paulhus's Two Tier System of Social Desirability (Paulhus, 2002).

righteous and somewhat 'saint-like'. These individuals have overly positive self-perceptions on traits such as agreeableness, dutifulness and restraint (Paulhus & John, 1998). These individuals see themselves as altruistic individuals who claim to possess traits indicating moral virtue and respect for social convention (Paulhus & John, 1998). At the impression management level individuals are often motivated to deliberately exaggerate their attainment of agency and communion values depending on their standing in relation to these values (Steenkamp et al., 2010). Individuals who are high on impression management adjust their responses to personality items in order to create a positive ('saint-like') impression (Heggestad, 2012). The same two clusters of traits (egoistic and moralistic) are present but the exaggeration is more conscious at the impression management level than at the self-deceptive level (Paulhus, 2002). In simpler terms, egoistic impression management is referred to as agency management while moralistic impression management is referred to as communion management. The authors further suggested that socially desirable responding consists of two self-favouring tendencies or self-deceptive styles (Paulhus & John, 1998). The first tendency, which is referred to as alpha, is an egoistic tendency to see the self as exceptionally talented and socially prominent in society. The second tendency, which is referred to as gamma, is a moralistic tendency to view the self as an exceptionally good member of society.

The study by Paulhus and John (1998) resulted in the development of the most elaborate conceptualisation of social desirability to date. Paulhus (2002) proposed that social desirability should be classified as positively biased self-perceptions on intellectual, social and emotional qualities (ERT) either perceived as unconscious and honestly held views or as those views that are strategically projected. Positively biased self-perceptions on attributes related to responsibility and interpersonal relationships (MRT) can be sincere and genuinely believed (unconscious deception) or they can be purposefully and instrumentally distorted (conscious deception) (Paulhus, 2002). Paulhus (2002) argued that conscious impression management is vulnerable to situational demands and is therefore not consistent over time, while unconscious self-deception is more dispositional and trait-like. This is depicted in the proposed two-tier system developed by Paulhus and displayed in Figure 5.

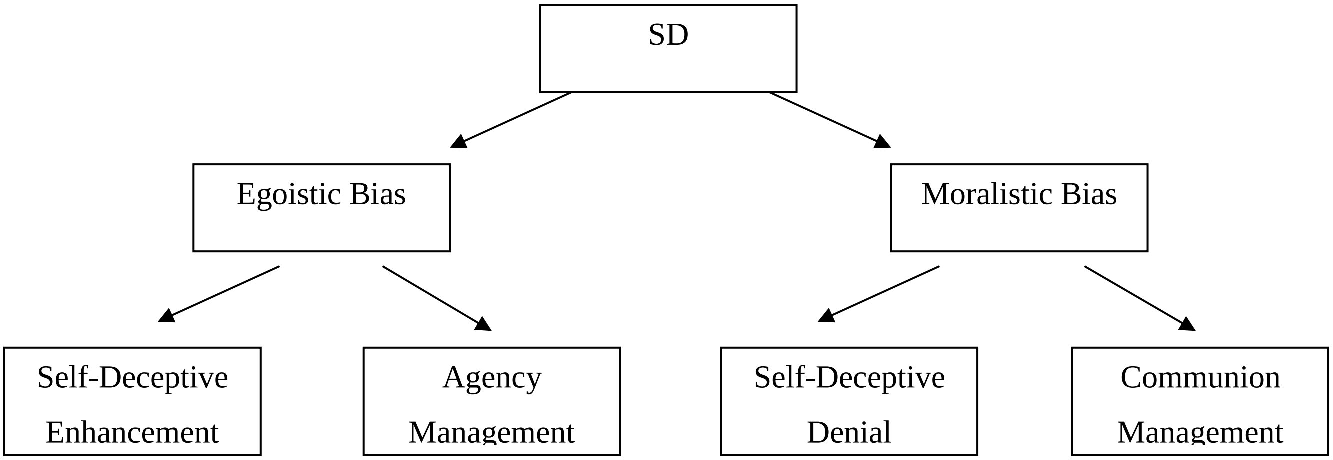

Figure 5: Paulhus' Two Tier System of SDR

(Source: Adapted from "Socially desirable responding: The evolution of a construct." By D.L. Paulhus-, (2002) p. 67).

According to Furnham (1985) individuals who respond in a socially desirable manner favour approved behaviours and deny any association with behaviours and opinions that are not socially approved. This has an implication for individuals responding to questionnaires, as they will alter their responses in relation to the desirability of the behaviour being tested. This response pattern may be conscious or unconscious (Paulhus, 2002). There are cases where the individual may provide responses that they acknowledge as a true representation of their personality traits although their spouses, peers and other observers would disagree (Ellingson et al., 2001). Paulhus and John (1998) noted that defensive biases like the ones discussed above intrude into self-perceptions of personality and ability.

Given the large number of models discussed in this section, it is possible to ask which of these models should be regarded as the most fruitful conceptualisation of social desirability. Constructs do not exist as such. They are intellectual constructions developed to allow the development of theoretical explanations of observed phenomena (Kerlinger& Lee, 2000).Theoretical explanations of an observed phenomenon are required to inform rational and purposeful attempts to influence, manage and control the phenomenon. The question therefore is which conceptualisation of social desirability will contribute most fruitfully to the discipline of psychology's explanation of differences in test scores.

Most researchers make use of the two-factor model when measuring social desirability (Uziel, 2010). Social desirability scales form part of most self-report inventories, including personality measures. The impact of social desirability on the test scores of personality assessments has been widely investigated (Grimm & Church, 1999).

2.3.3 Measuring Social Desirability

Personality assessments are usually presented in the form of questionnaires that ask participants to respond to a series of items by indicating the extent to which each item describes their personality (Dilchert, Ones, Viswesvaran & Deller, 2006). These questionnaires thus make use of self-report responses. Validity scales have been developed in an attempt to assess the accuracy of these self-reports and these validity scales examine the responses and then deduce the credibility of the personality profile obtained (Odendaal, 2013). Social desirability in personality inventories is usually assessed through the inclusion of one or more social desirability, impression management or faking scales within the inventory (Odendaal, 2013). These scales are known as validity scales and differ from other response style indicators such as acquiescence and extreme response sets. The validity scale items are usually dispersed amongst the other personality items in the inventory.

Paulhus (1991) found that social desirability scales have low intercorrelations with each other. This suggests that existing social desirability scales and measures tend to measure different underlying social desirability constructs. They are not defining social desirability in the same manner and are definitely not measuring the same construct. Part of the confusion concerning social desirability scales seem to be related to the fact that the term social desirability is often used to tap into both impression management and self-deception without distinguishing between the two concepts (Uziel, 2010). In contrast to other studies, a meta-analysis conducted by Espinosa and Van de Vijver (2014) found that empirically the correlations across scales show some consistency despite the differences in conceptualisation and methodological background. It would therefore seem that deciding on a social desirability scale is a contested area in personality assessment. Dilchert and Ones (2012) noted that there is longstanding empirical evidence to suggest that traditional social desirability scales are ineffective in addressing the issue of faking and impression management. According to McCrae and Costa (1983) studies have repeatedly shown that social desirability scales contain more substance than style. This means that they capture true variance in Emotional Stability, Agreeableness and Conscientiousness (Ones et al., 1996). Therefore, using these scales to disqualify supposed fakers could result in eliminating desirable respondents (Dilchert & Ones, 2012). Paulhus (20102) stressed that no social desirability measure should be used without sufficient evidence that a higher score indicates some difference from reality in terms of how truthful or deceptive the respondent is being when completing a personality assessment.

Many standard personality assessments include a scale that has been developed specifically to detect invalid profiles (Paulhus, 1991). These scales are included to target attempts to distort responses (Dilchert & Ones, 2012). Some of the most popular social desirability scales used in personality inventories are the Edwards Social Desirability Scale (ESD), which was constructed with items from the Minnesota Multiphasic Personality Inventory (MMPI) and is the earliest instrument developed to measure social desirability (Paulhus, 2012). The other popular scales are the Balanced Inventory of Desirable Responding (BIDR) (Paulhus, 1984), which measures self-deception and impression management, and the Marlowe-Crowne Social Desirability Scale (MCSD), which measures a one-dimensional construct called the need for approval.

According to Edwards (1964) it is possible for judges to rate the socially desirability or undesirability of statements in personality scales and inventories. Edwards (1957) asked ten judges to rate whether ‘true’ or ‘false’ was the most desirable response to each of the 79 items from the K, F and Lie scales of the MMPI. The judges unanimously agreed on 39 items and these items were used to form the SD scale (Paulhus, 1991). Participants’ scores can therefore range from 0 to 39 with higher scores indicating more socially desirable responses. However, the MMPI primarily identifies psychopathological aspects of personality symptoms and scores from this scale have been criticised for confounding SD and the absence of psychopathology (Edwards, 1957).

The BIDR measures two constructs, namely self-deception and impression management. Paulhus (1984) introduced the BIDR as an operationalisation of the two-component model of social desirability. The BIDR is a descendant of the Self- and Other-Deception Questionnaires developed by Sackeim and Gur in 1978 (Paulhus, 1991). The BIDR measures two constructs; self-deception enhancement (the tendency to give self-reports that are honest but positively biased) and impression management (deliberate self-presentation to an audience) (Paulhus, 1991). The self-deception items were originally developed based on the assumption that individuals with a propensity for self-deception tend to deny having psychologically threatening thoughts or feelings. The more recent scale emphasises exaggerated claims of positive cognitive attributes and shifted the focus from ego defence to ego enhancement (Paulhus, 1991). The impression management items were rationally developed based on the assumption that some respondents systematically over-report their performance of desirable behaviours and under-report undesirable behaviours. The claims involve overt behaviours and any distortions are assumed to be conscious lies (Paulhus, 1991). The 40 items on the BIDR are stated as propositions and respondents rate their agreement with each statement on a 7-point scale with a balanced scoring key. Total scores on the Self-Deception Enhancement (SDE) and Impression Management (IM) scales range between 0 and 20. This scoring ensures that high scores are attained only by individuals who give exaggeratedly desirable responses. The BIDR is the only multidimensional instrument that differentiates between SDE (assumed to measure unconscious positivity bias) and IM (which is believed to assess deliberate inflation of self-descriptions; Steenkamp et al., 2010). A third sub-scale termed Self-Deception Denial measures unconscious bias but because it usually correlates highly with IM this scale is not frequently used.

Moorman and Podsakoff (1992) conducted a meta-analysis that aimed to determine which social desirability scale is the most popular. They noted that 90% of the studies included in the analysis used some form of the Marlowe-Crowne SD Scale. All of the summed up items of the BIDR yield an overall measure of SDR that correlates highly with the MCSD scale (Paulhus, 1991). The items of this scale describe behaviours that are either culturally acceptable but improbable or culturally unacceptable but probable (Odendaal, 2013). Crowne and Marlowe (1960) set out to develop a measure of socially desirable responding that was an improvement of the Edwards scale. They noted the pathological nature of Edwards' items and decided to focus on ordinary personal and interpersonal behaviours (Paulhus, 1991). Crowne and Marlowe (1960, 1964) were critical of the fact that the Edwards' scale and the MMPI psychopathology scales had strong negative connotations. They argued that the scale did differentiate between respondents who truly lack psychopathological symptoms and respondents who merely denied such symptoms as a result of a need to be viewed as socially accepted (displaying social desirability) (Uziel, 2010). In other words, they argued that high scores on Edwards' SD scale may simply reflect a low frequency of pathological symptoms and not social desirability. To correct for this shortcoming an alternative scale was developed that included two types of items that were free from pathology and referred to infrequent but socially approved behaviours (e.g., "I always try to practice what I preach") and frequent but socially disapproved behaviours (e.g., "I like to gossip at times"). Using these two types of new items individuals who scored high on the approved behaviours and low on the unapproved behaviours were seen to have a high social desirability bias (Uziel, 2010).

During the development of the MCSD scale a number of personality inventories were consulted in order to construct a set of items for the new social desirability scale. According to Crowne and Marlowe (1960) the items selected for inclusion in the scale were all culturally approved and had minimal pathological or abnormal implications. The scale initially contained 50 items but was reduced to 33 items using item analysis and ratings from judges. The items either refer to desirable but uncommon behaviours (e.g. admitting mistakes) or undesirable but common behaviours (e.g. gossiping). Participants respond 'true' or 'false' to each item and scores range between 0 and 33, with high scores indicating that the individual has a higher need for approval by responding in a culturally appropriate manner, a definition which is consistent with Crowne and Marlowe's (1960) definition of social desirability. The internal consistency coefficient for this scale was .88 with a test-retest correlation of .89 (Crowne & Marlowe, 1960). Although this is the scale most commonly used to measure social desirability its use should be questioned as it is a one-dimensional scale and research suggests that social desirability exhibits more than one dimension. This scale confounds the two dimensions of social desirability dimensions and it is therefore unclear what exactly the scale measures.

2.3.4 Social Desirability And Personality Assessment

The use of social desirability scales in personality assessments has been widely debated in international literature with the focus being on the impact of the inclusion of these scales on the validity and utility of personality assessments (Odendaal, 2013). There are two competing views concerning social desirability. The first view suggests that social desirability is a deliberate distortion that makes selection decisions invalid while the other view states that social desirability represents valid variance that can be interpreted as a reflection of the way in which the respondents see themselves. The question of whether social desirability is a response style or a valid personality variable has been extensively debated and various researchers have reported true individual differences associated with social desirability (Odendaal, 2013). The argument that social desirability responding is a valid personality variable is based on the idea that the manner in which an individual distorts results on a personality assessment is a reflection of stable individual differences and therefore represents a potentially useful source of true variance (Rees & Metcalfe, 2003). As noted by the definitions provided previously, socially desirable responding is typically defined as the tendency to give positive self-descriptions (Paulhus, 2002) and can either be studied as a characteristic of questionnaire items or as an aspect of an individual's personality. According to Ellingson (2012) social desirability should be conceptualised as a latent variable characterising manifest behaviours and not as a latent variable describing a characteristic of a person (i.e. a trait). In 2001, Ellingson and colleagues concluded that the factor structure of personality was not influenced in a meaningful manner by either high or low social desirability. This finding differs from the conclusion reached by Espinosa and Van de Vijver's (2014) meta-review, which suggested that social desirability can be considered to be one of the spectrum of distinguishable personality traits. This line of reasoning was supported by Edwards (1964), who regarded an individual's tendency to respond to statements with a socially desirable scale value as a general personality trait. According to Furnham (1985) there are relatively stable and consistent individual differences in socially desirable responding, once again contributing to the debate concerning whether social desirability is a personality trait or a systematic response tendency that is determined by latent variables characterising the individual (possibly including personality variables) and the situation or context. McCrae and Costa (1983) assumed that lie scales and SD scales are indeed tapping into a stable personality trait that is characterised by confirming to social norms and adjustment according to those norms. Ones et al. (1996) also argued that when respondents deliberately fake good this is not a response style but is rather a manner in which the individual distorts results, thus reflecting an individual difference.

Research has found linkages between social desirability and the personality factors of Emotional Stability, Conscientiousness and Agreeableness (Birkeland et al., 2006; Dilchert et al., 2006; Li& Bagger, 2006; Ones, Viswesvaran & Reiss, 1996). In 1996, Ones et al. conducted a meta-analysis with job applicants. The analysis consisted of over 1400 correlation coefficients accumulated between the Big Five personality factors and social desirability measures. Based on the meta-analysis the researchers concluded that there is substantial variance in social desirability scales that can be explained by personality trait measures. The study reported the following population correlations between the five factors of personality and Social Desirability: Emotional Stability, 0.37; Conscientiousness, 0.20; Agreeableness, 0.14; Extraversion, 0.06; and Openness to Experience, 0.00. Ones et al. (1996) concluded that social desirability is linked to true individual differences in personality and is related to Emotional Stability, conscientiousness and, to a lesser extent, Agreeableness. These findings do not really shed light on the nature of the psychological mechanism that brought about these correlations. These findings therefore do not really clarify whether social desirability should be conceptualised as a personality trait related to other personality traits or whether it should be conceptualised as a behavioural latent variable affected by (amongst other) personality variables.

Ones et al.'s (1996) study was criticised as it did not differentiate between the impression management and self-deception components of social desirability. These limitations were addressed in a subsequent meta-analysis conducted by Li and Bagger (2006). This study examined the effects of both impression management and self-deception on the criterion-related validity of personality using the BIDR. Li and Bagger (2006) reported that both self-deception and impression management scales are related to substantive personality scales. Self-deception scales have stronger relationships with Emotional Stability (p = 0.54) and Conscientiousness (p = 0.42) than with the other Big Five personality factors. Impression management scales correlated most notably with Conscientiousness (p = 0.42), Agreeableness (p=0.42) and, to a lesser extent, Emotional Stability (p = 0.35). Li and Bagger (2006) also reported that both self-deception and impression management scales have small unreliability-corrected correlations with job performance (p = 0.12 and p = 0.10, respectively) and are therefore not particularly useful in the prediction of work performance.

A recent study conducted by Valchev, Van de Vijver and Meiring (2013) with the SAPI reported links between social desirability and the Social-Relational scales of Agreeableness and Interpersonal Relatedness as well as the Conscientiousness scale. The correlations for the two groups included in the study (a Black group and a White group) were significant. The results yielded a relatively strong correlation with the mean correlation being 0.39 for the Black group and 0.36 for the White group (Valchev et al., 2013). These correlations were lower for both groups for Agreeableness (0.37 for the Black group and 0.24 for the White group) and higher for Conscientiousness (0.46 and 0.42 respectively). Valchev and colleagues (2013) also noted that the SAPI Social Relational scales were generally strongly related to the Impression Management scales of Social Desirability. The Positive Social Relational scale was related to the Positive Impression Management scale and the Negative Social Relational scale was related to the Negative Impression Management scale.

In conclusion, there is compelling evidence to suggest that traditional social desirability scales are related to substantive personality scales, and that individual differences in social desirability scales reflect meaningful differences in personality.

The SAPI includes a SD scale. The objective of the study is to evaluate the construct validity of the SAPI's SD scale. The development history of the SAPI project is discussed next to contextualise the development of the SAPI SD scale.

2.4. THE SOUTH AFRICAN PERSONALITY INVENTORY

The most commonly used theory of personality, which is usually considered to be universal, is the Big Five theory, also referred to as the Five Factor Model (FFM) (Costa & McCrae, 1992). The five factors are neuroticism, openness to experience, extraversion, agreeableness and conscientiousness. The NEO-Personality Inventory (NEO-PI-R) is a widely used and researched instrument that is based on these five factors (Laher, 2008). Barrick and Mount (1991) noted that the personality traits that make up the FFM are increasingly applied in assessments within the field of Industrial Psychology. According to Cheung (2004) the FFM has been widely studied in cross-cultural personality research over the last two decades. Laher (2008) conducted a study on the structural equivalence of the NEO-PI-R and noted that the FFM is cross-culturally applicable. Allik and McCrae (2004) conducted a study using an international sample consisting of 36 cultures including both Black and White South Africans. The outcome of the study showed that Black South Africans as well as Africans and Europeans scored lower on Extraversion and Openness to experience, and scored higher on Agreeableness (Allik & McCrae, 2004). McCrae, Terracciano (2005) and 78 Members of the Personality Profiles of Cultures Project conducted a study that further explored the universality of the FFM in various countries, including five African countries. They found that the factor solutions in these cultures do not replicate the American normative structure for the FFM (McCrae & Terracciano., 2005). In 2000, Taylor conducted a construct comparability study of the NEO-PI-R for both Black and White employees in the South African workplace. It was found that this assessment worked better for the White population than it did for the Black population; and the Openness factor of the FFM was not replicated in the Black sample (Taylor, 2000). Taylor and De Bruin (2005) noted that attempts to identify the Big Five factors in South African samples have yielded disappointing outcomes. These findings have been attributed to the cultural inappropriateness of some of the items included in the imported questionnaires as they may be poorly understood (Meiring, 2007). There is currently a lack of evidence for the suitability of the FFM model in African countries and this presents an opportunity for further research in this cultural context (Laher, 2008).

A similar development, to the one initiated in the South African context, has occurred in China with the development of the Chinese Personality Assessment Inventory (CPAI), which was described in the previous section. Lin and Church (2004) investigated how the Interpersonal Relatedness dimension replicated in the Chinese American and the European American samples and found evidence indicating that the dimension is non-cultural specific. However, the dimension does appear to be more evident in the characteristics of the Chinese culture (Lin & Church, 2004). It is expected that the development of a personality inventory in South Africa would yield a similar outcome. Research in the last decade suggests that personality assessments are gaining momentum in South Africa (Meiring, Van de Vijver, Rothmann & Barrick, 2005; Taylor, 2000; Visser & Viviers, 2010). The field of personality psychology is increasingly focusing on making assessments more sensitive to ethnic differences (Nel, Valchev, Rothmann, Van de Vijver, Meiring & De Bruin, 2011; Valchev, Van de Vijver, Nel, Rothmann, Meiring & De Bruin, 2011). According to Taylor (2008) the pressure to conform to the stipulations of the EEA no. 47 of 2013 (Republic of South Africa, 2013) is positive as it encourages psychologists to create new local psychological measures specifically for South Africa. This development is important within the South African context and Meiring (2007) suggested that it will enhance the understanding and assessment of the personality structure of African individuals.

The South African Personality Inventory (SAPI) project was initiated in order to address the urgent need for psychological measurements that are in line with legislative requirements in South Africa (Meiring, Van de Vijver & Rothmann, 2006; Nel et al., 2012; Valchev et al., 2011)[5]. The SAPI and the CPAI-2 (revised version of the CPAI) are examples of personality assessments that were developed by psychologists to satisfy the need for indigenous inventories (Branco e Silva, 2012). There is a need to reduce bias in testing and the SAPI project aims to do this by identifying an indigenous personality model and developing an instrument that is able to assess this model across the ethnic groups found within South Africa (Valchev et al., 2014). The two phases that marked the development of this tool are described next.

2.4.1 The Development of the SAPI

The SAPI project was launched in 2005 and began with an indigenous investigation of personality conceptions that are culturally and linguistically appropriate for the 11 official language groups[6] in South Africa (Meiring, 2007). The general goal of the SAPI project is to develop a unified and valid personality inventory for all major language and cultural groups in South Africa (Nel, 2008). The project was not based on well-known personality theories such as the FFM but instead was based on common conceptualisations of personality as found in the South African language groups. The SAPI project follows a combined etic-emic approach and aims to identify an indigenous personality structure by collecting culturally relevant concepts in South Africa. Both cross-culturally common and culture-specific concepts are accommodated in this personality structure (Lotter, 2010). The development of the SAPI occurred in two phases and these are detailed below.

[5] It is thereby not implied that the development of a construct valid personality measure without construct and item bias is a sufficient (and in the case of item bias) even a necessary condition to adhere to the EEA's prohibition of unfair indirect discrimination in selection.

[6] The eleven official languages of South Africa are Afrikaans, English, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, Siswati, isiNdebele, Tshivenda, and Xitsonga.

Phase 1:

The SAPI started with an indigenous investigation of personality conceptions using an emic (culture-specific) approach. The SAPI team used a qualitative, comparative research design in which the personality structure was derived from interviews with individuals from all the groups (Nel, 2008). Personality descriptive terms were collected from the interviews (participants were asked to describe persons they knew well) and attempts were made to understand the personality structure occurring in natural language in the 11 official languages of South Africa. Fieldworkers[7] conducted the interviews using tape recorders and then transcribed the recordings to an answer sheet. During the interviews, the participants were asked to describe their own personality. This included that of a best friend from both genders, a parent, a child, a grandparent, a colleague, or friend from another ethnic group, a person who they view as psychologically different from themselves, a teacher they liked, and a teacher they did not like. Participants had to describe the target person by explaining what kind of person he or she is and describing typical aspects of the person, his or her behaviours, and habits. The interviews were transcribed into Excel and translated into English. Language experts checked the accuracy and corrected the translations where necessary. 53 139 personality descriptive utterances were gathered from 1,216 participants.

Following the transcriptions and translations of these interviews, the structures were compared across the languages to identify common and language specific aspects (Nel, 2008). Using content analysis techniques the researchers identified unique traits (specific to certain languages) and common traits (shared by most or all languages). These traits were further clustered, resulting in nine overall personality clusters. The personality model consisted of a three-tier (hierarchical) structure with 9 clusters at the top, 37 sub-clusters (between two and six sub-clusters per cluster), and 188 personality facets at the lowest level. For a full description of the breakdown and clustering of these traits see Nel et al. (2012.)

The nine factors were labelled conscientiousness, emotional stability, extraversion, facilitating, integrity, intellect, openness, relationship harmony, and soft-heartedness (Nel et al., 2012). Extraversion is a personality factor that characterises the person as having the tendency to be

[7] The field workers were native speakers of the language of the target group.

energetic, upbeat and able to communicate easily. The conscientiousness cluster defines a person as being determined and orientated toward achieving personal goals. These individuals have a personality trait that ensures they are precise and well organised when carrying out tasks. Emotional Stability refers to emotional balance, self-confidence and independence. Extraversion is used to describe a person who is gregarious, outspoken and not intimidated by others. Facilitating refers to a person's tendency to teach, mentor, motivate and guide others in reaching their full. A person who displays the traits of the integrity personality factor is honest, loyal and ethical. Intellect refers to creativity, innovation and being able to share information and to understand others. Openness is a cluster of traits that relate to being outspoken, adventurous, and open to new ideas and learning new things. Being approachable, accessible and cooperative in maintaining good relationships are characteristics of the relationship harmony cluster. The last cluster is called soft-heartedness and it involves being generous and having compassion for the feelings and needs others (Nel et al., 2012). In the first phase of the SAPI the qualitative data was analysed, resulting in the nine-factor structure described in this paragraph. The second phase saw the beginning of the quantitative exploration of these nine higher-order personality factors.

Phase 2:

In the next phase of the SAPI, items were developed to measure the various personality factors contained in the nine-factor structure. The SAPI project collaborators used specific techniques to develop the item pool for the inventory (Lotter, 2010; Janse van Rensburg, 2010). This section describes the items (both personality and social desirability) used in the SAPI. The SAPI project followed an adapted version of the lexical approach. Instead of using South African dictionaries for all the 11 languages, the SAPI project derived its personality descriptors from the qualitative phase's content-representative responses that were then transformed into item stems (Hill, Nel, Van de Vijver, Meiring, & Valchev, 2013). Items were then generated for the personality inventory based on the extraction of content-representative responses. In order to do this all the original responses associated with each SAPI cluster across the 11 languages were first grouped together. The responses within each cluster were categorised by sub-clusters and facets.

Construct maps were then created, after which the content-representative responses were transformed into item stems. A construct map contains a coherent and substantive definition for the content of the construct (Hill et al., 2013). In creating the item stems, the original responses were manipulated to form statements based on the main content of the original response. An example of this process is presented in Table 1.

Table 1*:*

Example of Item Stem Generation

Cluster	**Sub-cluster**	**Facet**	**Original Response**	**Content-representing response**	**Item stem**
Intellect	Social Intellect	Perceptive	She could easily see when you had a problem (Zulu response)	She could easily see when you had a problem	See when someone has a problem
Relationship Harmony	Approachability	Accommodating	Addresses us in English so we could understand (Xhosa response)	Addresses us in English so we could understand	Addressing others in mutually understandable language

Source: Adapted from "Developing and testing items for the South African Personality Inventory (SAPI)" by Hill et al (2013, p. 4).

After the item stems were developed the next stage involved item development. According to Hogan (2007) the item design process consists of four parts:

- Developing stimuli/items to which the examinee responds;
- Deciding on a response format or method;
- Determining conditions governing how the response is made to the stimulus; and
- Establishing procedures for scoring the response.

During the conversion of item stems into items, the following general item writing guidelines were followed to ensure standardisation of the items (Hendriks, Hofstee, & De Raad, 1999):

- The items had to be short, simple and understandable.
- The items had to be written in the first person and then followed by concrete behaviours, object, and context. Items had to refer to concrete behaviours and not beliefs, values or other orientations.
- Items had to describe a single activity, habit, or preference.
- The items could not include psychological trait terms or idiomatic expressions.
- The words often, always, and sometimes were excluded.

- Items were written with translatability in mind.

This procedure was followed with all of the nine clusters with the aim of developing a personality inventory to be used across all the cultural and language groups across South Africa. A total of 2573 items were developed for the nine clusters. During the item development stage the items were compared to the original responses to ensure that the essence of the original response was found in the items and that the item was relevant to the original response (Hill et al., 2013).

Once the item development process was completed, the number of items needed to be reduced to a more manageable number for the development of an experimental questionnaire that covered the different clusters and facets. Language experts were hosted at a workshop in which they were consulted and advised the SAPI team regarding the items in the pool. The language experts were tasked with reviewing that all the items were understandable, meaningful, translatable and culturally appropriate. All items that were not translated correctly or that could not be translated were removed from the item pool. This resulted in a total of 1 583 items.

Pilot studies were then conducted on each cluster. The results from these studies were statistically analysed using hierarchical factor analysis to reduce the number of items (Chrystal, 2012; Flattery, 2010; Janse van Rensburg, 2010; Labuschagne, 2010;Lötter, 2010). The exclusion criteria employed during this stage of item culling involved removing items with extreme mean values and low loadings. The decisions concerning which items to keep and which items to remove were also based on other psychometric considerations including item-total correlations, item loadings in factor analysis and substantive considerations including item formulation, content coverage, and content overlap. At the end of the pilot study phase the SAPI consisted of 606 items. Workshops were then held with cultural and language experts. These experts were able to advise the SAPI team regarding whether the items had the same meaning across the 11 language groups (Hill et al., 2013).A Microsoft Excel sheet containing all the items was sent to 10 language experts who were familiar with Afrikaans, isiXhosa, isiZulu, isiNdebele, SiSwati, Sesotho, Sepedi, Setswana, Xitsonga, and Tshivenda. These language experts received instructions concerning how to proceed with the item culling. The instructions involved checking whether all the items were understandable, meaningful, and translatable and culturally appropriate (Hill et al., 2013). The experts were also instructed to remove any items containing idiomatic expressions. In addition, items containing complex or culture-bound statements were deleted. The aim of this item culling process was to ensure that all items were worded in simple English so that translations to the other 10 languages would most likely be accurate and comparable across all official languages. After this process only 416 items remained. However, the SAPI team decided that further reduction was necessary. All items with more than ten words were thus deleted, leaving 315 items. The SAPI team then decided to delete, as far as possible, all items relating to abstract traits (e.g., items starting with 'I am'). This resulted in the final item pool consisting of 262 items, a figure that included12 social desirability items.

2.4.2 Recent Developments with the SAPI

A recent study (Valchev, et al., 2013) investigated the factor structure of the 262-item SAPI. The 262 items were administered to 1155 participants. The sample consisted of students as well as the general population. Exploratory factor analysis was conducted for each cluster and items with loading of <.30 were removed. This resulted in 156 items remaining, including 10 Social Desirability items[8]. The revised personality model consisted of a three-tier (hierarchical) structure with 9 clusters at the top, 37 sub-clusters (between two and six sub-clusters per cluster), and 188 personality facets at the lowest level. Therefore, the SAPI structure now consists of six factors and not the original nine described above. The new empirical, quantitative model for personality derived from the study was simpler than the qualitative model.

The structure resembled the Big Five, with separate Positive (SR-Positive) and Negative (SR-Negative) Social-Relational factors (Valchev, Van de Vijver, Meiring, Nel, Laher, & Hill., 2014). The six factors other than the two social relational scales are replicated in the FFM and they are; Extraversion, Conscientiousness, Neuroticism and Intellect/Openness. The Social-Relational factor still consisted of the largest number of concepts and Conscientiousness was broader than it had been initially defined. The factorial structure of the SAPI after this analysis was seen to consist of six factors with 18 sub-clusters. The 18 clusters, as described by Valchev et al.,2014 (pp.9) are as follows: Facilitating (10, "I give guidance to people in their life

[8] For the purpose of this study the original 12 items for Social Desirability were retained for all the analyses.

decisions"), Integrity (12, "I acknowledge my mistakes"), Social Intellect (4, "I understand how people feel"), Interpersonal Relatedness (9, "I help people live in peace"), Warm-Heartedness (12, "I support others when they need it"), Deceitfulness (3, "I fail to meet others' expectations"), Conflict-Seeking (6, "I cause fights"), Hostility-Egoism (13, "I make people feel vulnerable"), Emotional Balance (8, "I calm down easily"), Negative Emotionality (10, "I get angry a lot"), Playfulness (6, "I enjoy laughing with others"), Sociability (7, "I chat with many people"), Achievement Orientation (10, "I get motivated by my goals"), Orderliness (11, "I do things with precision"), Traditionalism-Religiosity (4, "I believe in tradition"), Intellect (10, "I learn new things easily"), Broad-Mindedness (5, "I seek new experiences"), and Epistemic Curiosity (6, "I love learning more about the world"). All scales were unipolar, and items were presented in a random order.

Several other data sets then became available within the SAPI project; these data sets provided information regarding both the original 262-item set and the 156 reduced item set. Further analysis was conducted on the factor structure of the SAPI using this new data (Valchev et al., 2014). Valchev et al. (2014) found that the six-factor structure did not replicate very clearly. The factors with unclear replication tended to be in the domains of intellect/openness, emotional stability and conscientiousness. This recent study (Valchev et al., 2014) had a total of 1364 participants from the four major ethnic groups. In this study items were added and removed from the 156 item sets and items from the 262-item set were used to improve the internal coherence and mutual distinctiveness of each factor. Although the number of items remained the same the list of items was different. This structure had 18 facet scales that were formed based on the per-cluster factor analysis of the last stage of item selection. These sub-clusters and the six-factor structure are displayed in Table 2.

Table 2*:*

Revised Factorial Structure of SAPI

Factors	Subscales
Social Relational Positive	• Facilitating • Interpersonal relatedness • Warm heartedness • Social intellect

Social Relational Negative	• Conflict seeking • Deceitfulness • Hostile egoism
Extraversion	• Playfulness • Sociability
Conscientiousness	• Integrity • Orderliness • Achievement orientation • Traditionalism- religiosity
Neuroticism	• Negative emotionality • Emotional balance
Intellect/ Openness	• Intellect • Epistemic curiosity • Broad mindedness

The original SAPI includes 12 Social Desirability items. The development of these items is discussed below.

2.4.3 Development of Social Desirability Items

A variety of different social desirability scales have been developed over the decades with the intention of measuring this construct within personality measurements. The most frequently used social desirability scales used in literature, as mentioned in section 2.3.3, are the Marlowe-Crowne Social Desirability Scale (MCSD) (Crowne & Marlowe, 1960) and the Balance Inventory of Desirability Responding (BIDR) (Paulhus, 1991). The SAPI included items from both these scales. The MCSD conceptualises social desirability as a one-dimensional construct and consists of 33 items describing desirable but uncommon everyday behaviours. The BIDR, on the other hand, conceptualises social desirability as a two-dimensional construct and therefore consist of a two subscales. The items comprising two sub-scales of the BIDR measure; *self-deception enhancement* (SDE) and *impression management* (IM) respectively. The original items needed to be simplified and the researchers, Valchev and colleagues (2013) did this by removing modifiers such as *always* and *sometimes*. Further to this they rephrased and shortened the statements. The reliabilities for the BIDR are; .67-.77 for the SDE subscale and .77-.85 for the IM subscale (Paulhus, 1991). The reliability for the MCSDS ranged from .73 to .88 (Crowne & Marlowe, 1960). The final list of items, from which scale they was extracted and the factor they were measuring in the original scale can be found in Table 3 below.

Table 3*:*

SD items used in the SAPI from the MCSD scale and the BIDR

SAPI SD Items	Original Scale	Factor measured in original scale
Item 27: I sometimes regret my decisions	Item 1 from BIDR	Self-deceptions enhancement
Item 35: I always obey laws even if I am unlikely to get caught	Item 26 from BIDR	Self – deception enhancement
Item 53: I am jealous of others with good fortune	Item 28 from MCSD scale	Social desirability
Item 90: I always do as I say	Item 17 from MCSD scale	Social desirability
Item 102: I think about my options before I make a choice	Item 1 from MCSD scale	Social desirability
Item 109: I admit when I do not know something	Item 16 from MCSD scale	Social desirability
Item 129: I sometimes tell lies if I have to	Item 21 from BIDR	Impression Management
Item 131: I have some bad habits	Item 39 from BIDR	Impression Management
Item 168: It is hard for me to break my bad habits	Item 2 from BIDR	Self – deception enhancement
Item 188: I am very confident of my judgments	Item 17 from BIDR	Self – deception enhancement
Item 223: My first impressions of people usually turn out to be right	Item 1 from BIDR	Self – deception enhancement
Item 252: I have done things that I don't tell other people about	Item 35 from BIDR	Impression Management

The SAPI includes 12 social desirability items. These items are presented in the Table 3 above. Of the 12 items there are six items in the positive direction (e.g., "I consider different options before committing to a choice") and six in the negative (e.g., "I have done things that I keep secret from others"). Within the SAPI questionnaire the 12 social desirability items were presented randomly among the other SAPI items. The negative items were reverse-scored. These items appear in Table 3 above.

Exploratory factor analyses were conducted on these items (Valchev et al., 2013) and two distinct scales were identified. The reliabilities of the two scales were very low with alphas of .40 to .50. The items that loaded on the first factor were the positively worded items namely; 35, 90, 102, 109, 188 and 223. The items that loaded on the second factor where the negatively worded items namely; 27, 53, 129, 131, 168 and 252. The first factor was called *positive impression management* and the second factor was called *negative impression management*. Valchev and colleagues (2013) concluded that both a single factor and a two-factor solution make sense and are supported by the data.

This study aims to evaluate the construct validity of the SAPI social desirability scale. No *a priori* constitutive definition of social desirability seemingly existed that guided the writing or selection of items for the SD scale. This is an unfortunate, regrettable oversight. Measuring instruments are required to obtain valid and reliable information on specific constructs carrying a specific connotative meaning. Constructs are man-made intellectual constructions created to allow man via his abstract thinking capacity to develop theoretical explanations of observed phenomena and to deduce from these practical measures to affect, manage and control these phenomena. The connotative meaning of a construct should be explicated when it is introduced as a latent variable in an explanatory model. The wisdom of developing a measuring instruments in the absence of a clear conceptualisation of the connotative meaning of the construct that the instrument should provide information on, should be questioned. If the connotative meaning of a to-be-measured construct is paramount, the writing of items should be guided by the internal structure of the construct as explicated in the constitutive definition of the construct.

In the case of the SAPI 12 items have been harvested and adapted from two existing social desirability scales. In these two scales the 12 items were earmarked and used as indicators of a specific latent social desirability dimension. It can therefore be argued that in the SAPI the these

12 items should reflect the same underlying latent social desirability dimension. These items were presumably not harvested for use in the SAPI SD scale because of their item content as such but rather because of what they are purported to reflect in their original scales. Indirectly therefore the choice of original scales from which to harvest items and the choice of items that were harvested do imply a specific conceptualisation of social desirability. The nature of this conceptualisation, however, depends on the manner in which the one-dimensional social desirability construct measured by the MCSD scale is conceptually seen to relate to the constitutive definitions of the two latent social desirability dimensions measured by the BIDR.

The BIDR (Paulhus, 1984) defines impression management as a conscious and intentional distortion of responses to create a favourable impression. Impression management is therefore conceptually similar to the traditional view of social desirability in terms of which the respondent deliberately alters their answers to present a positive social image (Ferrando, 2008). The MCSD interprets social desirability as responding in a culturally appropriate manner so as to satisfy the need for approval (Crowne & Marlowe, 1960). This in turn suggests that items 53, 90, 102, 109, 129, 131 and 252 all measure impression management. The remaining items (27, 35, 168, 188 and 223) are therefore suggested to measure *self-deception enhancement*. This line of reasoning suggests that the SD scale of the SAPI conceptualises social desirability as a two dimensional construct comprising the two positively correlated latent dimensions of *self-deception enhancement* and *impression management*.

An alternative line of reasoning is that the MCSD items were developed to measure a broad social desirability construct and, although the items of the BIDR were designed to measure narrower, more specific latent dimensions of social desirability, they nonetheless also measure the broader social desirability construct (Canivez, in press). This line of reasoning suggests that the SD scale of the SAPI conceptualises social desirability as a three dimensional construct comprising the two positively correlated latent dimensions of self-deception enhancement and impression management, but also a broad general social desirability latent dimension, uncorrelated with the two narrower social desirability dimensions, representing the common variance shared by all 12 of the SAPI's SD items.

In the original instruments from which the 12 social desirability items were harvested for the SAPI SD scale each item was earmarked to reflect individuals' standing on a specific latent social desirability. This design intention carried across to the SAPI SD scale. The two competing conceptualisations of the social desirability construct that were deduced from the choice of items for the SD scale therefore imply specific measurement models in which the 12 items are hypothesized to load according to a specific factor loading pattern on a specific number of latent social desirability dimensions and method factors. These measurement models will be explicated in Chapter 3.

A measurement model is an hypothesis on the process that produced the observed inter-item covariance matrix. The latent variable or latent variables that the instrument has been designed to measure should form part of this process and should exert statistically significant and strong influence on the observed item responses. They need, however, not be the only influences. Other, non-relevant, factors can also systematically explain variance in the item responses. These can most often not be anticipated beforehand. Sometimes, however, the presence of method bias factors can be convincingly hypothesised up-front. The SD scale of the SAPI comprises 6 positively keyed items and 6 negatively keyed items. This raises the concern that the positively keyed items might share variance simply because they share a design feature essentially unrelated to the construct of interest. The same concern applies to the negatively keyed items. If nature of the wording of the item (positively worded versus negatively worded) would substantially influence the item responses it would imply the existence of two method factors (a positively keyed factor and a negatively keyed factor). This raises the question whether it would be meaningful to add the two method factors to both measurement models. Hypothesising a measurement model in which items responses are dependent on a single substantive factor (a narrow group factor) as well as a method factor seems reasonable. Hypothesising a measurement model in which items responses are dependent on two substantive factors (a broad general factor and a narrower group factor) as well as a method factor seems somewhat less reasonable although not altogether impossible. A more plausible hypothesis is that the differentially keyed items will result in positive and negative factor loadings on the broad general social desirability factor. Although the first conceptualisation of the social desirability construct put forward is not altered by the foregoing argument, the measurement model associated with this conceptualisation is extended in that two additional method factors are added to the two existing narrow substantive social desirability factors.

2.4.4 Summary

The main objective of the SAPI is to develop an indigenous personality measure for all 11 official languages in South Africa. It is hoped that this will help overcome current problems facing personality measurement in South Africa. The general goal is to develop a unified and valid personality inventory that is useful for all language and cultural groups in South Africa. Different language and cultural groups have different perceptions, values and viewpoints and this results in differences in responses to questionnaires that measure an attribute of interest to psychologists. The SAPI is also influenced by these differences (response styles), which relate to the tendencies of individuals to respond in varied ways to a measuring inventory. It is therefore important that the SAPI includes a Social Desirability scale in order to measure the extent to which the respondents of the SAPI questionnaire alter their responses to place themselves in a more positive light. The concept of social desirability needs to be refined. Individuals who respond desirably on a personality measures are not necessarily faking as their responses might be indicative of an underlying personality trait. When respondents provide socially desirable answers this could simply indicate that that they do possess that trait and that they do in fact behave in that specific manner and are not merely faking to make a good impression. Social desirability is a contested construct in personality research and additional research is required, especially in the South African context, in order to more fully understand the construct. There are currently no psychometric properties found in the research data of personality inventories in South Africa that validate the use of the Social Desirability Scale in a personality assessment. Organisational and Industrial Psychologists currently treat high social desirability scores in personality assessments as evidence of faking.

2.5. CONCLUSION

Researchers have long sought to understand human behaviour and psychological assessments were developed as part of this quest in an attempt to quantifiably understand behaviour and its manifestation in the practical world. Personality inventories are one example of such psychological assessments. The SAPI project was initiated with the aim of developing an assessment that is able to measure the behaviour of individuals and thus provide insight into their personalities in the culturally diverse South African context. This endeavour is governed by current South African legislation, which is based on the need for fair and reliable assessments for all South African cultural groups. The factor structure of personality has been tested in the different ethnic groups to ensure that there is no bias present in the personality constructs across these different ethnic groups. One of the constructs in which these researchers are interested is social desirability, and this study aimed to investigate the factor structure of the construct and to further examine its cross-culture applicability. In the chapter that follows the research method is outlined, included the different procedures that were followed. The chapter details the road map that was followed for the exploration phase of the social desirability items, the relationship that these items have with the other six personality factors and how the items are replicated in a factor structure across the four cultural (ethnic) groups in South Africa.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 INTRODUCTION

The current study aims to come to a valid verdict on the construct validity of the Social Desirability scale of the SAPI. The methodology used in this study will determine whether the verdict on the construct validity of the SD scale will be valid. Methodology serves the epistemic ideal through two characteristics of science, namely objectivity and rationality (Babbie & Mouton, 2001). Science is objective in that it is consciously and purposefully focussed on the reduction of error. Science is rational in that it insists that the method used should be subjected to inspection by members of the scientific community who are knowledgeable regarding the research being conducted (Babbie & Mouton, 2001). Knowledgeable peers can, however, only critically evaluate the methodological choices that were made when the method used is described thoroughly and the methodological decisions that were made are clearly motivated (Babbie & Mouton, 2001). This chapter therefore provides a comprehensive description of the specific research process that was followed in this study by focusing on the research design, participants, procedures, measuring instruments and statistical analysis procedures.

3.2 SUBSTANTIVE RESEARCH HYPOTHESES

The main objective of this study was to test whether the SAPI SD scale measures social desirability as it is constitutively defined. In other words, the study aimed to determine whether the items used in the Social Desirability scale provide a construct valid and reliable measure of social desirability as it is defined in the SAPI project. The problem, however, is that the SAPI project seemingly never explicitly conceptualised the connotative meaning of the social desirability construct that the SD scale is meant to measure. In the absence of an *a priori* constitutive definition of social desirability the current study inferred two possible conceptualisations of the social desirability construct as measured by the SD sale of the SAPI.

These two possible conceptualisations were described in Chapter 2, paragraph 2.4.2. Based on these two conceptualisations the following two substantive research hypotheses were formulated.

- Substantive research hypothesis 1: The SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a two dimensional construct comprising the two positively correlated latent dimensions of self-deception enhancement and impression management;
- Substantive research hypothesis 2: The SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a three dimensional construct comprising the two positively correlated latent dimensions of self-deception enhancement and impression management, but also a broad, general social desirability latent dimension, uncorrelated with the two narrower social desirability dimensions, representing the common variance shared by all twelve of the SAPI's SD items.

The two conceptualisations of the social desirability construct, combined with the manner in which the items were assigned to underlying factors in the scales that they were harvested from, imply two specific measurement models. This allowed the derivation of the following two operational research hypotheses.

Operational research hypothesis 1 representing substantive research hypothesis 1.

- Operational hypothesis 1.1: The two-factor measurement model implied by substantive hypothesis 1 can closely reproduce the co-variances observed between the items comprising the two SD subscales (self-deception enhancement and impression management)[9];
- Operational hypothesis 1.2: The factor loadings of the items on their designated latent social desirability dimension(self-deception enhancement or impression management) are statistically significant ($p<.05$) and large ($\lambda_{ij} \leq .50$);
- Operational hypothesis 1.3: The measurement error variances associated with each item are statistically significant ($p<.05$) but small;

[9] In all subsequent Λ matrices the self-deception dimension or factor will consistently be treated as the first factor

- Operational hypothesis 1.4: The social desirability dimensions (self-deception enhancement and impression management) explain large proportions of the variance in the items that represent them ($\lambda^2_{ij} \geq .25$); and
- Operational hypothesis 1.5: The two social desirability dimensions (self-deception enhancement and impression management) correlate low to moderate ($\phi_{ij} < .90$) with each other (i.e., the SAPI social desirability dimensions display discriminant validity).

Operational research hypothesis 2 representing substantive research hypothesis 2.

- Operational hypothesis 2.1: The three-factor measurement model implied by substantive hypothesis 2 can closely reproduce the co-variances observed between the items comprising each of the two SD subscales (self-deception enhancement and impression management);
- Operational hypothesis 2.2: The factor loadings of the items on their designated latent social desirability dimensions (self-deception enhancement and the general social desirability factor or impression management and the general social desirability factor) are statistically significant ($p < .05$) and large ($\lambda_{ij} \geq .50$);
- Operational hypothesis 2.3: The measurement error variances associated with each item are statistically significant ($p < .05$) but small;
- Operational hypothesis 2.4: The social desirability dimensions (self-deception enhancement, impression management and the general social desirability factor) explain large proportions of the variance in the items that represent them ($\lambda^2_{ij} \geq .25$); and
- Operational hypothesis 2.5: The two social desirability dimensions (self-deception enhancement and impression management) correlate low to moderate ($\phi_{ij} < .90$) with each other[10].

The literature study hypothesised the presence of method factors. The SD scale of the SAPI comprises six positively keyed items and six negatively keyed items. The literature study

[10] The correlations between self-deception enhancement and the general social desirability factor and between impression management and the general social desirability factor were fixed to zero to reflect the assumption that the narrower, more specific, social desirability factors measure aspects of social desirability not related to the broad factor. Since these two elements of $\mathbf{\Phi}$ were fixed to zero this cannot form part of operational hypothesis 2.5.

hypothesised that the positively keyed items would share variance simply because they share a design feature essentially unrelated to the construct of interest. The same argument applies to the negatively keyed items. The literature study hypothesised subsequently argued that if the nature of the wording of the item (positively worded versus negatively worded) would substantially influence the item responses it would imply the existence of two method factors (a positively keyed factor and a negatively keyed factor) in addition to the two substantive social desirability factors of interest (self-deception enhancement and impression management). The two conceptualisations of the social desirability construct is not altered by this argument. The substantive research hypotheses are therefore not altered. The operational research hypotheses that are derived from the substantive research hypotheses are, however, altered. Taking the effect of the differentially keyed of the SD scale into account, the first substantive hypothesis translated into the following operational research hypothesis.

Operational research hypothesis 3 representing substantive research hypothesis 1 :

- Operational hypothesis 3.1: The two-factor measurement model implied by substantive hypothesis 1 and the additional assumption of two method factors can closely reproduce the co-variances observed between the items comprising the two SD subscales (self-deception enhancement and impression management);
- Operational hypothesis 3.2: The factor loadings of the items on their designated latent social desirability dimension (self-deception enhancement or impression management) and on their designated method factors (positively or negatively keyed factor) are statistically significant ($p<.05$) and large ($\lambda_{ij}\geq.50$);
- Operational hypothesis 3.3: The measurement error variances associated with each item are statistically significant ($p<.05$) but small,
- Operational hypothesis 3.4: The social desirability dimensions (self-deception enhancement and impression management) and the two method factors explain large proportions of the variance in the items that represent them ($\lambda^2_{ij}\geq.25$); and

- Operational hypothesis 3.5: The two social desirability dimensions (self-deception enhancement and impression management) correlate low to moderate (ϕ_{ij}<.90) with each other[11].

The second conceptualisation of the social desirability construct put forward is also not altered by the argument that the items of the SAPI SD scale are plagued by method bias. Rather than extending the measurement model assumed under operational research hypothesis 2 by assuming the addition of two method factors, hypothesis 2 is rather refined by hypothesising that the differentially keyed items will result in positive and negative factor loadings on the broad general social desirability factor.[12] This line of reasoning assumes that the measurement model is fitted to data in which the negatively keyed items were not reflected.

Operational research hypothesis 4 representing substantive research hypothesis 2.

- Operational hypothesis 4.1: The three-factor measurement model implied by substantive hypothesis 2 can closely reproduce the co-variances observed between the items comprising each of the two SD subscales (self-deception enhancement and impression management);
- Operational hypothesis 4.2: The factor loadings of the items on their designated latent social desirability dimensions (self-deception enhancement and the general social desirability factor or impression management and the general social desirability factor) are statistically significant (p<.05) and large ($\lambda_{ij}\geq$.50);
- Operational hypothesis 4.3: The factor loadings of the positively keyed items on their designated latent social desirability dimensions (self-deception enhancement and the general social desirability factor or impression management and the general social desirability factor) will be statistically significant (p<.05) and positive and the factor loadings of the negatively keyed items on their designated latent social desirability dimensions (self-deception enhancement and the general social desirability factor or

[11] The correlations between self-deception enhancement and the two method factors and between impression management and the two method factors were fixed to zero to reflect the assumption the narrower, more specific, social desirability factors measure aspects of social desirability not related to the broad factor. Since these four elements of Φ were fixed to zero this cannot form part of operational hypothesis 2.5.

[12]This line of reasoning does not deny the possibility that such five-factor model could have been hypothesised under operational hypothesis 4 or even that an additional operational hypothesis 5 could have been formulated that reflects two method factors in addition to the general factor and the two narrower social desirability factors..

impression management and the general social desirability factor) will be statistically significant ($p<.05$) and negative;

- Operational hypothesis 4.4: The measurement error variances associated with each item are statistically significant ($p<.05$) but small,
- Operational hypothesis 4.5: The social desirability dimensions (self-deception enhancement, impression management and the general social desirability factor) explain large proportions of the variance in the items that represent them ($\lambda^2_{ij} \geq .25$); and
- Operational hypothesis 4.6: The two social desirability dimensions (self-deception enhancement and impression management) correlate low to moderate ($\phi_{ij}<.90$) with each other.

Operational research hypothesis 1 translates into matrix equation 1.

$$\begin{pmatrix} X_{27} \\ X_{35} \\ X_{53} \\ X_{90} \\ X_{102} \\ X_{109} \\ X_{129} \\ X_{131} \\ X_{168} \\ X_{188} \\ X_{223} \\ X_{252} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & 0 \\ \lambda_{2,1} & 0 \\ 0 & \lambda_{3,2} \\ 0 & \lambda_{4,2} \\ 0 & \lambda_{5,2} \\ 0 & \lambda_{6,2} \\ 0 & \lambda_{7,2} \\ 0 & \lambda_{8,2} \\ \lambda_{9,1} & 0 \\ \lambda_{10,1} & 0 \\ \lambda_{11,1} & 0 \\ 0 & \lambda_{12,2} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{pmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

The 12x12 measurement error variance-covariance matrix $\mathbf{\Theta}_\delta$ was defined as a diagonal matrix. Only the off-diagonal elements of the 2x2 matrix $\mathbf{\Phi}$ of latent variable variances and covariance were freed to be estimated. The diagonal of $\mathbf{\Phi}$ was fixed to unity.

Operational research hypothesis 2 translates into matrix equation 2.

$$\begin{pmatrix} X_{27} \\ X_{35} \\ X_{53} \\ X_{90} \\ X_{102} \\ X_{109} \\ X_{129} \\ X_{131} \\ X_{168} \\ X_{188} \\ X_{223} \\ X_{252} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & 0 & \lambda_{1,3} \\ \lambda_{2,1} & 0 & \lambda_{2,3} \\ 0 & \lambda_{3,2} & \lambda_{3,3} \\ 0 & \lambda_{4,2} & \lambda_{4,3} \\ 0 & \lambda_{5,2} & \lambda_{5,3} \\ 0 & \lambda_{6,2} & \lambda_{6,3} \\ 0 & \lambda_{7,2} & \lambda_{7,3} \\ 0 & \lambda_{8,2} & \lambda_{8,3} \\ \lambda_{9,1} & 0 & \lambda_{9,3} \\ \lambda_{10,1} & 0 & \lambda_{10,3} \\ \lambda_{11,1} & 0 & \lambda_{11,3} \\ 0 & \lambda_{12,2} & \lambda_{12,3} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{pmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots[2]$$

The 12x12 measurement error variance-covariance matrix $\mathbf{\Theta}_\delta$ was defined as a diagonal matrix. In the 3x3 variance-covariance matric $\mathbf{\Phi}\phi_{13}$ and ϕ_{23} were fixed to zero but ϕ_{12} was freed to be estimated. The diagonal of $\mathbf{\Phi}$ was fixed to unity.

Operational research hypothesis 3 translates into matrix equation 3[13].

$$\begin{pmatrix} X_{27} \\ X_{35} \\ X_{53} \\ X_{90} \\ X_{102} \\ X_{109} \\ X_{129} \\ X_{131} \\ X_{168} \\ X_{188} \\ X_{223} \\ X_{252} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & 0 & 0 & \lambda_{1,4} \\ \lambda_{2,1} & 0 & \lambda_{2,3} & 0 \\ 0 & \lambda_{3,2} & 0 & \lambda_{2,4} \\ 0 & \lambda_{4,2} & \lambda_{4,3} & 0 \\ 0 & \lambda_{5,2} & \lambda_{5,3} & 0 \\ 0 & \lambda_{6,2} & \lambda_{6,3} & 0 \\ 0 & \lambda_{7,2} & 0 & \lambda_{7,4} \\ 0 & \lambda_{8,2} & 0 & \lambda_{8,4} \\ \lambda_{9,1} & 0 & 0 & \lambda_{9,4} \\ \lambda_{10,1} & 0 & \lambda_{10,3} & 0 \\ \lambda_{11,1} & 0 & \lambda_{11,3} & 0 \\ 0 & \lambda_{12,2} & 0 & \lambda_{12,4} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{pmatrix} \quad \ldots\ldots\ldots[3]$$

The 12x12 measurement error variance-covariance matrix $\mathbf{\Theta}_\delta$ was defined as a diagonal matrix. In the 4x4 variance-covariance matric $\mathbf{\Phi}\phi_{13}$, $\phi_{14}\phi_{23}$ and ϕ_{24}were fixed to zero but $\phi_{12}$and ϕ_{34}were freed to be estimated. The diagonal of $\mathbf{\Phi}$ was fixed to unity.

[13] Factor 3 is the positively keyed method factor and factor 4 the negatively keyed method factor

Operational research hypothesis 4 translates into matrix equation 4.

$$\begin{pmatrix} X_{27} \\ X_{35} \\ X_{53} \\ X_{90} \\ X_{102} \\ X_{109} \\ X_{129} \\ X_{131} \\ X_{168} \\ X_{188} \\ X_{223} \\ X_{252} \end{pmatrix} = \begin{pmatrix} -\lambda_{1,1} & 0 & -\lambda_{1,3} \\ +\lambda_{2,1} & 0 & +\lambda_{2,3} \\ 0 & -\lambda_{3,2} & -\lambda_{3,3} \\ 0 & +\lambda_{4,2} & +\lambda_{4,3} \\ 0 & +\lambda_{5,2} & +\lambda_{5,3} \\ 0 & +\lambda_{6,2} & +\lambda_{6,3} \\ 0 & -\lambda_{7,2} & -\lambda_{7,3} \\ 0 & -\lambda_{8,2} & -\lambda_{8,3} \\ -\lambda_{9,1} & 0 & -\lambda_{9,3} \\ +\lambda_{10,1} & 0 & +\lambda_{10,3} \\ +\lambda_{11,1} & 0 & +\lambda_{11,3} \\ 0 & -\lambda_{12,2} & -\lambda_{12,3} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{pmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots [4]$$

The 12x12 measurement error variance-covariance matrix $\mathbf{\Theta}_\delta$ was defined as a diagonal matrix. In the 3x3 variance-covariance matric $\mathbf{\Phi}\phi_{13}$ and ϕ_{23} were fixed to zero but ϕ_{12} was freed to be estimated. The diagonal of $\mathbf{\Phi}$ was fixed to unity.

3.3 RESEARCH DESIGN

In order to empirically investigate the two substantive research hypotheses, a strategy was needed that would provide unambiguous, empirical evidence that could be used to evaluate the four operational hypotheses. A research design constitutes a plan or a strategy that guides the empirical testing of the validity of the claims made by the hypotheses (Kerlinger & Lee, 2000). The hypotheses formulated for this study made specific claims about the measurement model implied by the SD scale of the SAPI. A research design attempts to ensure that the obtained empirical evidence can be interpreted unambiguously for or against the hypothesis (Donnelley, 2009). Babbie and Mouton (2001) defined the research design as a guideline or blue print of how the researcher intends to test the substantive research hypothesis. The research hypothesis and the type of evidence that is required to test the hypothesis dictate the types of design that is considered appropriate for the particular research study. According to Kerlinger and Lee (2000), an *ex post facto* research design is a systematic empirical inquiry in which the researcher does not have control of independent variables as their manifestations have already occurred. More specifically an *ex post facto* correlational design was used with individual items serving as indicator variables of the latent variables.. Correlation research investigates the relationships between two or more variables that have not been manipulated with the aim of establishing the extent to which the variables co-vary (Emlyn, 2006). The *ex post facto* correlational design provides a test of the claims made by the operational research hypotheses in terms of the following logic. Measures are obtained on the 12SAPI SD scale items and the inter-item covariance matrix is calculated. Estimates for the freed measurement model parameters are obtained in an iterative fashion with the purpose of reproducing the observed inter-item co-variance matrix as accurately as possible (Diamantopoulos & Siguaw, 2000). If the fitted model fails to reproduce the observed covariance matrix accurately (Kelloway, 1998) the conclusion invariably has to follow that the fitted measurement model does not provide an acceptable explanation for the observed inter-item covariance matrix (Moyo, 2009). This would than mean that the items of the SD scale of the SAPI do not measure social desirability as intended. The opposite, however, is not true. If the reproduced covariance matrix derived from estimated measurement model parameters closely corresponds to the observed inter-item covariance matrix it cannot be concluded that the processes postulated by the measurement model necessarily *must have* produced the observed covariance matrix(Moyo, 2009). High correspondence between the observed and estimated inter-item covariance matrices would only mean that the dynamic portrayed in the measurement model provides a permissible explanation for the observed covariance matrix. The claim that the SD scale of the SAPI provides construct valid measures of the social desirability dimension construct, given the specific constitutive definition, defines the construct would thereby have survived an opportunity to be falsified (Popper, 1972). This study aims to provide evidence for this.

3.4 STATISTICAL HYPOTHESES

The nature of the statistical hypotheses used to test the operational hypothesis depends on the decision regarding the nature of the envisaged statistical analyses. The argument that the two conceptualisations of the social desirability construct, taken in conjunction with the manner in which the items were assigned to underlying factors in the scales that they were harvested from, imply two specific measurement models points to the need to perform confirmatory factor analysis using structural equation modelling (Diamantopoulos & Siguaw, 2000). More specifically structural equation modelling utilising LISREL (Jöreskog & Sörbom, 1996) will be used to test the hypothesis that the measurement model defined by operational hypotheses 1 to 4 can closely reproduce the observed covariance matrix.

The present study investigated two substantive research hypotheses, linked to four operational research hypotheses. The statistical hypotheses indicated below represent statistical translations of the claims made by the four operational research hypotheses.

The claims made by the four operational research hypotheses that the measurement model implied by substantive hypothesis 1 or 2 can reproduce the co-variances observed between the items comprising each of the two SD subscales (self-deception enhancement and impression management),without and with method bias taken into account, translate into two overarching model fit hypotheses. More specifically the exact fit null hypotheses (H_{01i}; i=1, 2, 3, 4)were tested which represents the rather bold stance that the measurement model implied by the i^{th} operational hypothesis accurately reflects the measurement model in the parameter (Browne & Cudeck, 1993).

H_{01i} : RMSEA = 0; i=1, 2, 3, 4

H_{a1i} : RMSEA > 0; i=1, 2, 3, 4

If it is, however, somewhat more realistically assumed that that the measurement model implied by substantive hypothesis 1 or 2,without and with method bias taken into account, approximates the processes that operated in reality to create the observed co-variance matrix, the following close fit null hypotheses (H_{02i}; 1, 2, 3, 4) were tested (Browne & Cudeck, 1993):

H_{02i} : RMSEA ≤ .05; i=1, 2, 3, 4

H_{a2i} : RMSEA > .05; i=1, 2, 3, 4

If the exact or close measurement model fit would be found for the measurement model fitted under operational research hypothesis 1 (i.e. H_{01i}; i=1 or H_{02i}; i=1 would not be rejected), the following 12 null hypotheses on the slope of the regression of item j on latent social desirability dimension k will be tested for operational research hypothesis 1:

H_{0pi}: $\lambda_{jk}=0$;p=2, 3, …, 13; i=1; j=1, 2, …, 12; k=1, 2[14]

H_{api}: $\lambda_{jk}\neq 0$;p=2, 3, …, 13; i=1; j=1, 2, …, 12; k=1, 2

If the exact or close measurement model fit would be found for the measurement model fitted under operational research hypothesis 2 (i.e. H_{01i}; i=2 or H_{02i}; i=2 would not be rejected), the following 24 null hypotheses on the slope of the regression of item j on latent social desirability dimension k were tested for operational research hypotheses 2:

H_{0pi}: $\lambda_{jk}=0$;p=14, 15, …, 37; i=2,; j=1, 2, …, 12; k=1, 2, 3

H_{api}: $\lambda_{jk}\neq 0$;p=14, 15, …, 37; i=2,; j=1, 2, …, 12; k=1, 2, 3

If the exact or close measurement model fit would be found for the measurement model fitted under operational research hypothesis 3 (i.e. H_{01i}; i=3 or H_{02i}; i=3 would not be rejected), the following 24 null hypotheses on the slope of the regression of item j on latent social desirability dimension k were tested for operational research hypothesis 3:

H_{0pi}: $\lambda_{jk}=0$; p=38, 39, …, 61; i=4; j=1, 2, …, 12; k=1, 2, 3, 4

H_{api}: $\lambda_{jk}\neq 0$; p=38, 39, …, 61; i=3; j=1, 2, …, 12; k=1, 2, 3, 4

If the exact or close measurement model fit would be found for the measurement model fitted under operational research hypothesis 4 (i.e. H_{01i}; i=4 or H_{02i}; i=4 would not be rejected), the following 12 null hypotheses[15] on the slope of the regression of item j on latent social desirability dimension k were tested for operational research hypotheses 4:

H_{0pi}: $\lambda_{jk}=0$;p=64, 65, 68, 69, 79, 71, 72, 73, 80, 81, 82, 83; i=4; j=2, 4, 5, 6, 10,11; k=1, 2, 3

H_{api}: $\lambda_{jk}>0$;p=64, 65, 68, 69, 79, 71, 72, 73, 80, 81, 82, 83; i=4; j=2, 4, 5, 6, 10, 11; k=1, 2, 3

If the exact or close measurement model fit would be found for the measurement model fitted under operational research hypothesis 4 (i.e. H_{01i}; i=4 or H_{02i}; i=4 would not be rejected), the

[14] p is is counted across the Λ matrix in the rows starting from the first freed factor loading for RQ27_SD

[15] These hypotheses apply to the 6 items that are hypothesised to load positively on the two SD factors and the general factor

following 12 null hypotheses on the slope of the regression of item j on latent social desirability dimension k were tested for operational research hypotheses 4:

H_{0pi}: $\lambda_{jk}=0$; p=62, 63, 66, 67, 74, 75, 76, 77, 78, 79, 84, 85; i=4; j=1, 3, 7, 8, 9,12; k=1, 2, 3

Hapi: λjk<0; p=62, 63, 66, 67, 74, 75, 76, 77, 78, 79, 84, 85; i=4; j=1, 3, 7, 8, 9,12; k=1, 2, 3

If the exact or close fit would be found (i.e. H_{01i}; i=1, 2, 3, 4 or H_{02i}; i=1, 2, 3, 4 not be rejected), the following 48null hypotheses were tested with regards to the freed elements in the variance-covariance matrices$\mathbf{\Theta_{\delta i}}$; i=1, 2, 3, 4 for operational research hypothesis 1 to 4:

H_{0pi}: $\Theta_{\delta jj}=0$;p=87, 88, …, 134; i =1, 2, 3, 4; j=1, 2, ..., 12

H_{api}: $\Theta_{\delta ijj}> 0$;p=87, 88, …,134; ; i =1, 2, 3, 4; j=1, 2 ....., 12

If the exact or close fit would be found (i.e. H_{01i}; i=1 or H_{02i}; i=1 not be rejected), the following null hypothesis were tested with regards to the freed elements in the variance-co-variance matrix $\mathbf{\Phi_i}$; i=1for operational research hypothesis 1:

H_{0pi}: $\phi_{jk}=0$; p=135; i =1; j=1; k=2

H_{api}: $\phi_{jk}> 0$;p=135; i =1; j=1; k=2

If the exact or close fit would be found (i.e. H_{01i}; i=2, 4 or H_{02i}; i=2, 4 not be rejected), the following 2 null hypotheses were tested with regards to the freed elements in the variance-co-variance matrix $\mathbf{\Phi_i}$; i=2, 4for operational research hypotheses 2 and 4:

H_{0pi}: $\phi_{jk}=0$; p=136, 137; i =2, 4; j=1; k=2[16]

H_{api}: $\phi_{jk}> 0$;p=136, 137; i =2, 4; j=1; k=2

If the exact or close fit would be found (i.e. H_{01i}; i=3 or H_{02i}; i=3 not be rejected), the following null hypothesis were tested with regards to the freed elements in the variance-co-variance matrix $\mathbf{\Phi_i}$; i=3for operational research hypothesis 3:

H_{0pi}: $\phi_{jk}=0$; p=138, 139; i =3; j=1, 3; k=2, 4[17]

H_{api}: $\phi_{jk}> 0$;p=138, 139; i =3; j=1, 3; k=2, 4

[16] ϕ_{13}andϕ_{23}were fixed to zero

[17] $\phi_{13,}$ $\phi_{14,}$ ϕ_{23}andϕ_{424}were fixed to zero

3.5. PARTICIPANTS AND PROCEDURES

3.5.1 Sampling

The target population of the current study, and the population for which the SAPI was developed and standardised, consists of all adult South Africans. To allow the selection of a representative sample from the target population, the target population needs to be operationalised in the form of a sampling population. The sampling population comprises those final sampling units in the target population that have a non-zero probability of being selected (Babbie & Mouton, 2001). A sample is a smaller collection of units from within a population that can be used to determine certain truths about the population (Field, 2005). Struwig and Stead (2001) identified two major sampling techniques, referred to as probability and non-probability sampling. This study made use of non-probability sampling, which is used in situations where the "probability of any particular member of the population being chosen is unknown" (Struwig & Stead, 2001, p. 11). This study made use of a form of non-probability sampling known as convenience sampling, which involves a sample being selected purely on availability (Struwig & Stead, 2001).

The general rule for research studies is that larger samples are better. This is because the sample size dictates the extent to which observations can be generalised to the target population. In order to allow for generalisation it is not only important to have a large sample size but also to have a sample that is representative of its broader population (De Goede & Theron, 2010). According to Kerlinger (1973), the larger the sample size the smaller the error and this allows for more accurate calculation of statistics. Some researchers suggest that SEM analyses should not be performed on samples smaller than 200 participants, whereas other researchers recommend a sample size of between 100 and 200 participants (Field, 2013). The minimum sample size is also dependent on the number of parameters to be estimated as per the number of variables (Kline, 2005). Bentler and Chou (as cited in Kelloway, 1998) propose that the ratio of sample size to number of parameters estimated should fall between 5:1 and 10:1. In the current study the Bentler and Chou guideline translates to a sample size of between 185 and 370final sampling units given 37 freed parameters in the largest measurement model specified under operational research hypothesis 3. Another aspect that was considered when determining the required sample size was statistical power. In the context of confirmatory factor analysis, statistical power refers to the probability of rejecting the null hypothesis of close fit (H_0: RMSEA $\leq .05$) when in fact it should be rejected (i.e. the model fit is actually mediocre; H_a: RMSEA = .08). Software developed in R by Preacher and Coffman (2006) was used in the current study to derive a sample size estimates for the test of close fit for the measurement specified under operational hypothesis 3. The parameter RMSEA was set to .05 under H_0, under H_a the parameter RMSEA was set to .08, the significance level was set to .05, the desired level of statistical power was set to .80 and the degrees of freedom were calculated as 107 (144-37). Syntax developed by Preacher and Coffman (2006) in R and available from *http://www.quantsy.org/rmsea/rmsea.htm*returned a required sample size of 126 final sampling units to ensure statistical power of .80 when testing the null hypothesis of close fit of the measurement specified under operational hypothesis 3 when in reality the model fits mediocre in the parameter (RMSEA=.08).

The participant group for this study included a sample of *n*=1289. The sample size exceeds the minimum sample size requirements derived from the perspective of the ratio of sample size to number of freed parameters and from the perspective of statistical power. This sample consisted of volunteers from various South African ethnic groups from within private security industries, the South Africa health care industry and two South African universities. In the current study a substantial sampling gap existed between the sampling population and the target population. This meant necessarily means that the sample drawn from the sampling population will not be representative of the target population. This problem was further compounded by the use of a non-probability, convenience sampling procedure. The non-representativeness of the validation sample is acknowledged as a limitation of the research study.

The descriptive statistics for the sample are presented in Table 4 below.

Table 4*:*

Demographic Information for Sample

Variable	Value	Frequency	Percentage
Age	10- 20 years	74	10%
	20-30 years	381	49%
	30-40 years	163	21%
	40-50 years	100	13%
	50 – 60 years	46	6%
	60-73 years	11	1%
Educational Level	Grade 9	41	3%
	Grade 12	559	46%
	Certificate	158	13%
	Diploma	110	9%

	Bachelors	180	15%
	Honours	108	9%
	Masters	22	2%
	Doctorate	1	0.80%
	Other	32	2%
Gender	Male	551	44%
	Female	714	56%
English Reading Ability	Very poor	9	0.70%
	Poor	10	0.80%
	Good	517	41%
	Very good	722	57%
First Language	Afrikaans	235	19%
	English	361	30%
	IsiNdebele	11	1%
	IsiXhosa	136	11%
	IsiZulu	151	13%
	Sepedi	72	6%
	Sesotho	68	6%
	Setswana	65	6%
	SiSwati	16	1%
	Tshivenda	25	2%
	Xitsonga	39	3%
	Other	27	2%
Race	White	341	27%
	Black	655	53%
	Indian	87	7%
	Coloured	159	13%

The average age of the participants was 31 years old and Table 4 shows that the majority of participants have an educational level of Grade 12. The majority of the participants were females. Most participants identified their reading ability as good or very good. The majority of the participants identified English as their first language and selected Black as their race. This data was self-reported by the participants.

3.5.2 Data Collection

The data collection process involved physically administering the SAPI experimental questionnaire to participants. This study made use of a survey methodology utilizing a paper-and-pencil approach for data collection. Participants entered their responses to the SAPI questionnaire on optical answer sheets. A company known as CSX, a division of Metrofile (Pty) Ltd, that specialises in the supply, installation and support of business solutions, scanned the answer sheets and produced a data file that was exported into an Microsoft Excel spreadsheet for further analysis.

3.6 MEASURING INSTRUMENT

This study made use of the experimental SAPI questionnaire, which consists of 262 items, including twelve social desirability items.. The full set of social desirability items (12 items) were used. The SAPI questionnaire uses a five-point Likert-type response format with responses ranging from ‘strongly disagree' to ‘strongly agree’. Likert-scale items are commonly used to investigate respondents’ attitudes and feelings by asking the respondents to rate the strength of their feeling on a scale that is easy to understand (Dittrich, Francis, Hatzinger, & Katzenbeisser, 2007).

3.7 STATISTICAL ANALYSIS

This section describes the statistical procedures used in this study and identifies the order in which these procedures were performed. The initial statistical procedure followed was the standard procedure that has been followed by the SAPI project to date. De Bruin (2010) created a manual, referred to as the SAPI data analysis manual, detailing these procedures. The result from these analyses formed the basis for testing the claim made by the developers of the SAPI that the social desirability items measure social desirability, and that they do so in the manner that was initially intended when the scoring key was developed.

The analysis began by treating the data for missing values, examining the data for inadmissible values, the identification of outliers, the identification of poor items. Subsequently confirmatory factor analysis (CFA) was used to test the four operational research hypotheses derived from the two substantive research hypotheses that the SD scale of the SAPI provides a construct valid measure of social desirability as given the constitutive definition derived from the original scales that were consulted when developing the scale.

3.7.1 Data Preparation Phase

According to Hair, Black, Babin and Anderson (2006) data examination is a necessary initial step in any analysis. During this step the researcher examines the impact of missing data, identifies outliers and tests for the assumptions underlying the multivariate techniques to be used. According to De Bruin (2010) it is not uncommon for test-takers to not respond to a particular item. This may occur due to long questionnaires or, in research that focuses on sensitive topics, when respondents exercise their right to not answer a question. In the data preparation phase the data was investigated for missing values and errors before it could be used in subsequent analysis.

Imputation by matching (Jöreskog & Sörbom, 1996) was used in the current study to replace missing values. Imputation by matching involves substituting real values for missing values. The substitute values replaced for a case with a missing value on a given item were derived from one or more other cases that have a similar response pattern over a set of matching variables. The ideal is to use matching variables that are not be utilised in confirmatory factor analysis. This ideal was not feasible in the current study. The items least plagued by missing values were consequently identified to serve as matching variables. The six items that had no missing values were used as matching variables. Cases with missing values that could not be imputed were eliminated from the data set (Jöreskog & Sörbom, 1996).

The imputed dataset was then used in the descriptive analysis, item analysis and confirmatory factor analyses.

3.7.2 Descriptive Statistics

The SAPI items are strictly speaking discrete variables but were considered to approximate continuous variables and therefore descriptive statistics such as means and standard deviations were considered legitimate to use (Pallant, 2007). A continuous variable is a variable that provides a score for each person and can take on any value on the measurement scale that is being used (Field, 2013). Descriptive statistics describe a number of statistical characteristics of the item distributions. Inferences can be derived from these item distribution characteristics how well an item serves its purpose of representing the latent variable it was tasked to reflect (Taylor, 2008). Moreover data capturing errors can be detected through the inspection of specific characteristics of the item distributions. The means and standard deviations of items convey important information. The mean provides information regarding the respondents' general selection tendency for the items, and the standard deviation conveys information about the average deviation of responses from the mean of the item (Taylor, 2000). Descriptive statistics also provide information about the distribution of the scores on continuous variables through the skewness and kurtosis values (Pallant, 2007). Information on the minimum and maximum item values provide important information on inadmissible item values due to data capturing errors. Descriptive statistics were used to determine the quality of the data through the calculation of the mean, standard deviation, skewness, kurtosis and the minimum and maximum values. Skewness is a measure of the symmetry of a frequency distribution (Field, 2009). Kurtosis is a measure of the degree to "which scores cluster in the tails of a frequency distribution" (Field, 2009, p. 788) and can be formally defined as the standardised fourth population moment about the mean (DeCarlo, 1997). An examination of the item descriptive statistics allowed the identification of errors in the data as well as the identification of items that were deemed unsuitable. Items were considered unsuitable indicators of the latent variables they were designated to represent if they failed to discriminate between different levels of the latent variable. The skewness and kurtosis cut-offs for the response scale were set at > 2 and > 4[18] respectively to allow the researcher to use as many respondents as possible and to exclude the items that were unsuitable (DeCarlo, 1997; Field, 2009). When skewness and kurtosis values range between -1.50 and +1.50 the distribution is considered to approximate normality (Muthen & Kaplan, 1985).

Outliers are cases with extreme values on one or more items (Field, 2009). Univariate outliers are cases that fall more than one-and-a-half inter-quartile range above the seventy-fifth percentile (P_{75}) or more than one-and-a-half inter-quartile range below the twenty-fifth percentile (P_{25}). Multivariate outliers are cases that have extreme values on two or more items.

The identification of poor items via the analysis of the item means, standard deviations, skewness, kurtosis and the minimum and maximum item values, was supplemented by classical measurement item analysis. This allowed the further identification of poor items. Items were also considered unsuitable indicators of the latent variables they were designated to represent if they failed to act in unison with the other items that were designated to reflect the same social desirability dimension.

[18] De Bruin (2009) recommended that the cut-off for kurtosis is > 7. Hence the normal kurtosis in the SPSS was subtracted from 7, giving a cut-off of > 4.

3.7.3 Item And Reliability Analysis

The objective of item analysis was to detect and remove poor items from the SD scale (Murphy & Davidshofer, 2005). Poor items were defined as items that do not reflect the latent social desirability dimension that the items have been tasked to reflect and items that do not elicit different responses when relative small differences exist on the latent social desirability dimension. Test developers typically aim to construct uni-dimensional scales or subscales containing items that all to some degree reflect a single common underlying latent variable or latent dimension. If this is the case the items designated to reflect a single common underlying latent variable should at least correlate moderately and the item analysis procedure is essentially an analysis of correlations between each item and between each item and the total score (Kline, 2005). Nunnally (1978) noted that item analysis can be used during test development to make the first selection of items that will be subjected to factor analysis. When new scales in tests or complete tests are being constructed the design intention is that the items, which are selected to represent a latent variable, in fact exclusively measure the intended latent variable (Kriel, 2011). This ideal is, however, in reality never met. Item analysis attempts to determine to what extent this ideal has been achieved. Item analysis was conducted on all the SD scale items by means of the SPSS22 Reliability Procedure and the results are reported in in Section 4.5.

The items of the Social Desirability scale include negatively and positively worded items. The positively keyed and negatively keyed items are indicated in Table 5.This could potentially impact on the reliability of the scale as not all the items are worded in the same direction. An item is meant to be a stimulus to which the respondent would react in a way that expresses the specific underlying latent variable. Therefore, before the items were analysed for reliability statistics the negatively worded items were reflected. In the study conducted by Valchev and colleagues (2013), in which a two-factor structure was found for Social Desirability, both positive and negative scales for impression management emerged and these scales were used in this study to form the Social Desirability scale for the original SAPI questionnaire (see Table 5).

The item analysis was performed separately on the two SD subscales. This decision made good sense in the case of operational research hypothesis 1 but not when viewed from the perspectives of operational research hypotheses 2, 3 and 4. The measurement models specified under the latter three operational research hypotheses argue that all 12 items of the SAPI SD scale are complex items that load on at least two latent dimensions. Moreover it is hypothesised that the general factor is unrelated to the two narrower social desirability factors in the measurement models hypothesised by operational research hypotheses 2 and 4. Likewise, under operational research hypothesis 3 the two method factors are hypothesised to be unrelated to the two narrower social desirability factors in the hypothesised measurement model. Categorising items in smaller groups based on their hypothesised loadings on two factors will therefore not solve the problem. To the extent that the item analysis return results that point to somewhat noisy, low internal consistency data for the two social desirability subscales it could be interpreted to as evidence that testifies against the measurement model hypothesised under operational research hypothesis 1 and evidence that is compatible with the measurement models hypothesised under operational research hypothesis 2, 3 and 4.

Table 5*:*

Social Desirability items for SAPI project

Item Number	**Item Detail**
Negatively worded items	
Item 27	I sometimes regret my decisions.
Item 53	I am jealous of others with good fortune.
Item 129	I sometimes tell lies if I have to.
Item 131	I have some bad habits.
Item 168	It is hard for me to break my bad habits.
Item 252	I have done things that I do not tell other people about.
Positively worded items	
Item 35	I always obey laws, even if I am unlikely to get caught.
Item 90	I always do as I say.
Item 102	I think about my options before I make a choice.
Item 109	I admit when I do not know something.
Item 188	I am very confident of my judgments.
Item 223	My first impressions of people usually turn out to be right.

Source: Adapted from “The use of traits and contextual information in free personality descriptions across ethno cultural groups in South Africa” by Valchev et al., 2013

3.7.4 Dimensionality Analysis Via EFA

Test developers typically aim to construct uni-dimensional scales or subscales containing items that all to some degree reflect a single common underlying latent variable or latent dimension. If this is the case all the items of such a scale of subscale should have reasonably high factor loadings on a single underlying factor. Dimensionality analysis was consequently conducted via exploratory factor analysis (EFA) on each of the two subscales of the SD scale of the SAPI. The objective of the EFA analysis was to explore the number of factors underlying each of the two subscales of the Social Desirability scale. Exploratory factor analysis analyses the shared or common variance amongst the items in order to determine the number of common factors that need to be assumed to explain the observed inter-item correlation matrix. There are a variety of factor extraction methods, and the two most commonly used methods are, principal-components analysis (PCA) and common-factor analysis (FA) (Worthington & Whittaker, 2006). The purpose of PCA, according to these researchers, is to reduce the number of items whilst still maintaining as much of the original item variance as possible (Worthington & Whittaker, 2006). With this SAPI project the aim is to not reduce the number of items measuring the two hypothesised latent dimensions of social desirability but to rather understand the latent factors or constructs that account for the shared variance among items. Therefore, FA is more closely aligned to the objective of the current study and Worthington and Whittaker (2006) recommend that this approach be implemented in the case of the development of new scales.

The unidimensionality assumption was tested for each of the social desirability subscales by subjecting each of the subscales to a principal axis factor analysis(PAF) with oblique rotation using direct oblimin rotation. The statistical technique of PAF was chosen in preference to principal component analysis because the reasons outlined earlier. The decision as to the appropriate number of factors to extract was based on the rule of thumb known as the 'eigenvalue-greater-than-one' criterion. If there are two eigenvalues greater than one this suggest the presence of two factors. Kaiser (1960) recommended that all factors with eigenvalues greater than one should be retained. This criterion is based on the argument that in the standardised item data set the variance of a single item is one, and since eigenvalues represent the amount of variation explained by a factor, only factors that explain more variance than a single item should be considered for retention (Field, 2013). The number of factors to extract was also determined through the use of a scree plot. Using this criterion the number of factors to the left of the elbow indicates the number of factors that should be extracted. It is also possible to set the factor extraction to a specific number. Parallel analysis (Horn, 1965), is another procedure that can be employed to decide how many factors to retain. This analysis is done when a random data set of the same order as the raw data set (i.e. an equal number of observations and items) is created and factor analysis is then conducted on both the original data set and the random data set. The number of factors to retain is determined by looking at the eigenvalues in both data sets and a factor is retained if the eigenvalue for the original data is larger than the eigenvalue from the random data (Worthington & Whittaker, 2006). The eigenvalues-greater-than-unity rule, the scree plot and parallel analysis were employed in this study to determine the number of factors that should be extracted to explain the observed inter-item correlation matrix.

To examine the factor analysability of the two subscales the Kaiser-Meyer-Olkin (KMO) (Kaiser, 1960) measure of sampling adequacy was examined. When the KMO value approaches one, or is at least greater than .60, the inter-item correlation matrix is considered to be suitable for factor analysis. A KMO value close to one indicates that patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors (Field, 2013). If in the parameter the variables in the correlation matrix do not correlate with each other than the matrix would be an identity matrix. The Bartlett test tests the null hypothesis that the correlation matrix in the parameter is an identity matrix meaning that every item correlates only with itself and nothing else. The identity matrix implies that there are no common factors and that every item is unique with no evidence of common variance. Low factor analysability would constitute comment negatively on the subscales of the SD scale as a common factor should emerge if a subscale is successfully measuring a specific latent trait. In factor analysis, when the sample size is large, the Bartlett's test will nearly always be significant (Field, 2013).

Once one or more factors have been extracted it is possible to calculate the degree to which variables load on the factor (Field, 2013). In most rotated factor analytic solutions the variables being factor analysed have high loadings on a single factor and low loadings on the remaining factors. In a single-factor solution and in an orthogonal rotated factor solution the factor loadings can be interpreted a correlation coefficients. The square of the factor loadings therefore reflect the proportion of variance in the item that is explained by the underlying factor. In an oblique solution the interpretation is less straight-forward (Tabachnick & Fidell, 2007). An oblique rotation method assumes that the factors that have being extracted are related to each other and are not independent (Field, 2013). The factors therefore share variance. When utilising oblique rotation the pattern matrix reflects the partial regression coefficients that express the slope of the regression of the items on each extracted factor when controlling for the other factors in the solution (Tabachnick & Fidell, 2007). Direct oblimin rotation was used in this study.

The decision to perform the dimensionality analysis on the two subscales of the SAPI SD scale again made good sense in the case of operational research hypothesis 1, but not when viewed from the perspectives of operational research hypotheses 2, 3 and 4. The same argument that applied to the item analysis also applies to the use of EFA to evaluate the assumption that each subscale measures a single, undifferentiated latent social desirability dimension.

3.7.5 Single Group Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is used when the researcher has sufficient theoretical and empirical basis for specifying the most plausible model or subset of models. This is because CFA allows for focused testing of specific hypotheses about the data (Finch & West, 1997). According to Floyd and Widaman (1995), CFA is used when enough evidence exists to specify which variables should load onto which common factors. CFA is based on the common factor model and aims to represent the structure of correlations among measured variables using a set of latent variables (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Multivariate normality tests were also conducted prior to conducting the CFA. CFA allows the researcher to test the hypothesis that a relationship between the observed variables and their underlying latent construct does indeed exist. In order to do so CFA relies on statistical tests to determine the adequacy of the model fit to the data (Moyo & Theron, 2011).

The most commonly used approach when conducting CFA is structural equation modelling (SEM), which was the statistical technique employed to test the operational research hypotheses on the nature of the factor structure underlying the SD scale of the SAPI. In a confirmatory factor analysis items are freed to load on the particular factor or factors that the items have been designated to reflect and the factor loadings of the items on all the remaining factors that the items were not designed to reflect are fix to zero. The ability of each measurement model reflecting the assumptions made by each of the operational research hypotheses on the number of factors, the correlations between the factors and the loading pattern of the items on the factors, to accurately reproduce the observed inter-item covariances and item variances is then evaluated (Worthington & Whittaker, 2009). If the model is able to accurately reproduce the observed variance-covariance matrix the assumptions made by the operational research hypothesis on the number of factors, the correlations between the factors and the loading pattern of the items on the factors, become a valid (i.e. permissible) stance on the process that produced the observed variance-covariance matrix. SEM is a powerful tool to use when confirming a factor structure as it allows the researcher high levels of control over the form of constraints placed on the items and factors when analysing the hypothesised model (Worthington & Whittaker, 2006).

If the sample size is too small to use the individual items as indicator variables when fitting the measurement model then the items are combined into item parcels. Worthington and Whittaker (2006), however, recommend against the use of item parcelling in SEM for scale development research as it can hide problematic relationships between individual the items included in the parcel and the underlying latent variable they were meant to reflect. Parcelling was not used in the current study as the sample was large enough to use the individual items as indicator variables in the CFA analysis and there were only 12 items to be analysed.

When fitting measurement models via LISREL 8.8 (Du Toit & Du Toit, 2001) two crucial decisions needed to be made that affected the outcomes of the analysis. The first decision pertained to the type of matrix to be analysed, either a covariance or a correlation matrix. If the observed variables are continuous then a covariance matrix should be analysed (Mels, 2003). Data that is obtained from Likert scales that have five or more scale points maybe specified as continuous data (Muthén & Kaplan, 1985). The SAPI utilises a five point Likert scale and it is for this reason that the item measures were regarded as continuous data. The second decision that needed to be made was what on the estimation method to use to estimate the freed model parameters. Maximum likelihood estimation is preferred when continuous data is being used (Mels, 2003; Moyo & Theron, 2011). Maximum likelihood estimation, however, makes the assumption that the data follows a multivariate normal distribution, and if this is not the case, then one would get inappropriate model fit statistics and standard error estimates. Therefore, it is good practice to see if data satisfies the assumption of multivariate normality (Theron, 2011a). Mels (2003) indicated that robust maximum likelihood (RML) should be used when analysing non-normal data. The current study tested the null hypothesis that the multivariate item distribution follows a multivariate normal distribution (Theron, 2011b). The multivariate normality of the multivariate item indicators indicator distribution was evaluated using PRELIS (Jöreskog & Sörbom, 1996). In the case where the null hypothesis of multivariate normality was rejected, normalisation was attempted and if the hypothesis of multivariate normality was still rejected, robust maximum likelihood estimation was used (Mels, 2003).

The measurement models specified under operational research hypothesis 3 and 4 were fitted with specific constrains imposed on **Φ**. Under operational hypothesis 3 the correlation between the positively keyed factor (ξ_3) and the two narrower social desirability factors (ξ_1 and ξ_2) were constrained to zero as well as the correlation between the negatively keyed factor (ξ_1) and the two narrower social desirability factors (ξ_1 and ξ_2).Under operational research hypothesis 4 the correlation between the broad, general social desirability factor (ξ_3) and the two narrower social desirability factors (ξ_1 and ξ_2) were constrained to zero to reflect the assumption that the two narrower latent social desirability dimensions capture unique variance in specific items designated to reflect self-deception enhancement and impression management that is not shared with the broad, general social desirability factor.

Measurement model fit in the parameter was determined by testing the exact and close fit null hypotheses formulated in paragraph 3.4 for each of the four fitted measurement models. The fit of the four measurement models in the sample were further interpreted by using the full spectrum of fit indices reported in LISREL, which provide evidence for the adequateness of the fit of each of the four measurement models specified under the fourth operational research hypotheses to the sample data. The fit indices were interpreted holistically and integrated in order to make an informed decision about whether or not the model fits the sample data. In Chapter 4 the indicators of fit are described in detail during the discussion of the evaluation of the measurement models. An examination of the modification indices presented by LISREL is also reviewed.

For each of the 4 measurement models, specified under the 4 operational research hypotheses, where the close fit null hypotheses were not rejected, the measurement model parameter estimates were interpreted by testing the hypotheses formulated in paragraph 3.4 with regards to $\mathbf{\Lambda}$, $\mathbf{\Theta}_\delta$ and $\mathbf{\Phi}$

3.8 SUMMARY

This chapter outlined the research methodology and research design that were used in this research study. It provided a comprehensive description of the specific research process that was followed in this study by focusing on the research design, participants, procedures, measuring instruments and statistical analysis procedures. The methodology that was employed in this study will determine whether the verdict on the construct validity of the SD scale will be valid. This chapter further described that in the absence of *a priori* constitutive definition of social desirability there were two possible conceptualisations; a two-dimensional construct (self-deception enhancement and impression management) or three-dimensional construct (inclusive of a general social desirability latent variable). These two conceptualisation of the social desirability construct were operationalised in terms of two multi-facetted operational hypotheses. This chapter further explained that there is a possibility of the presence of method factors due to the use of positively and negatively keyed items in the SD-scale. The impact of assuming method factors under each of the two substantive research hypotheses on the previous two operational hypotheses were also described in detail. The matrix equations for the four operational hypotheses were provided. The research design was described in this chapter as a strategy that is employed to find unambiguous, empirical evidence that could be used to evaluate the four operational hypotheses. A description of the participants and procedures was provided. The processes followed to conduct the item analysis, reliability procedures, the exploratory and confirmatory factor analyses for the single group model were explained. The results of these analyses are provided in the section in Chapter 4 that follows.

CHAPTER FOUR

RESEARCH RESULTS

4.1 INTRODUCTION

This chapter presents the statistical findings obtained from the analysis of the SD scale of the experimental SAPI instrument used in this study. These results are presented in a narrative sequence, beginning with the description of the descriptive analysis, the discussion of the missing values imputation, followed by a discussion of the test of multivariate normality. The results of the item analysis and dimensionality analysis are subsequently presented. The CFA results for each of the 4 operational research hypotheses are finally presented. The instrument was administered to a sample of participants from across South Africa (*N*=1289). Both the SPSS 22 and LISREL8.8 (Du Toit & Du Toit, 2001) statistical packages were utilised to perform the analyses.

4.2 MISSING VALUES

Imputation by matching (Jöreskog & Sörbom, 1996) was used in the current study to replace missing values. The number of missing variables for each item of the SAPI SD scale are shown in Table 6.

Table 6*:*

Number of missing values per item of the SAPI SD scale

Item	Number of missing values
RQ129_SD	20
RQ131_SD	3
RQ168_SD	9
RQ252_SD	6
Q035_SD	7
Q090_SD	19
Q102_SD	0
Q109_SD	0
Q188_SD	0
Q223_SD	0
RQ27_SD	0
RQ53_SD	0

The imputation by matching procedure resulted in a dataset with no missing data for n=1263 respondents. The imputation by matching procedure therefore failed to successfully impute 26 missing values.

4.3 DESCRIPTIVE STATISTICS

In this study the data was initially screened for data capturing errors and outliers. The SD scale items were checked for scores that were not in the range of possible answers. The SAPI questionnaire uses a 5-point response scale and therefore all the values should fall between 1 and 5. Values that fall out of this range could be indicative of typing errors that occurred during the data capturing process.

Table 7 displays the means, standard deviations, minimum and maximum values and skewness and kurtosis of the 12 Social Desirability items. Those items with inadmissible extreme scores are indicated in bold. ItemRQ53_SD returned a maximum value of 6.16 which fell outside the range of permissible values. This value was obtained by case 37. The incorrect value was not due to the imputation of missing values. The incorrect value already appeared in the original data set. Case 37 was subsequently deleted from the data set. The descriptive statistics were subsequently recalculate and are shown in Table 8. Inspection of Table 8 indicates that seven of the twelve items distributions (RQ27_SD, RQ53_SD, Q102_SD, Q109_SD, RQ131_SD, Q188_SD, Q223_SD) are statistically significantly ($p<.05$) skewed to the left and four item distributions (Q090_SD, RQ129_SD, RQ168_SD, RQ252_SD) are statistically significantly ($p<.05$) skewed to the right. Regarding kurtosis four item distributions are statistically significantly ($p>.05$) platykurtic (Q035_SD, Q090_SD, RQ168_SD and RQ252_SD) and three item distributions (RQ27_SD, Q188_SD and Q223_SD) are statistically significantly ($p>.05$) leptokurtic.

Table 7:

Descriptive Statistics for the Imputed Data Set

		RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD	RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
N	Valid	1263	1263	1263	1263	1263	1263	1263	1263	1263	1263	1263	1263
	Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean		3.86410	3.18923	3.53554	2.63994	3.81815	3.62515	2.42551	4.01742	3.03866	4.05714	4.03560	2.74188
Median		4.00000	3.00000	4.00000	2.00000	4.00000	4.00000	2.00000	4.00000	3.00000	4.00000	4.00000	2.00000
Mode		4.000	4.000	4.000	2.000	4.000	4.000	2.000	5.000	3.000	4.000	4.000	2.000
Std. Deviation		.753827	1.172105	.912089	1.142901	.966236	.850041	.970578	1.059863	1.066393	.772042	.887792	1.161948
Variance		.568	1.374	.832	1.306	.934	.723	.942	1.123	1.137	.596	.788	1.350
Skewness		-.734*	-.123	-.364*	.355*	-.696*	-.239*	.523*	-.971*	.155*	-.884*	-1.163*	.377*
Std. Error of Skewness		.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069
Kurtosis		1.301*	-.958*	.058	-.746*	.107	-.272*	.024	.267	-.727*	1.555*	1.786*	-.812*
Std. Error of Kurtosis		.138	.138	.138	.138	.138	.138	.138	.138	.138	.138	.138	.138
Minimum		1.000	1.000	.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Maximum		5.000	5.000	6.155	5.000	5.000	5.000	5.000	5.000	5.000	5.311	5.234	5.000

*(p<.05)

Table 8*:*

Descriptive Statistics for the Imputed Data Set (with case 37 deleted)

		RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD	RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
N	Valid	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262
	Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean		3.86320	3.18780	3.53346	2.63807	3.81721	3.62406	2.42347	4.01664	3.03711	4.05640	4.03563	2.74089
Median		4.00000	3.00000	4.00000	2.00000	4.00000	4.00000	2.00000	4.00000	3.00000	4.00000	4.00000	2.00000
Mode		4.000	4.000	4.000	2.000	4.000	4.000	2.000	5.000	3.000	4.000	4.000	2.000
Std. Deviation		.753446	1.171459	.909460	1.141419	.966046	.849496	.968250	1.059922	1.065384	.771891	.888143	1.161868
Variance		.568	1.372	.827	1.303	.933	.722	.938	1.123	1.135	.596	.789	1.350
Skewness		-.735*	-.123	-.380*	.355*	-.696*	-.240*	.519*	-.970*	.155*	-.884*	-1.163*	.379*
Std. Error of Skewness		.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069
Kurtosis		1.306*	-.957*	.038	-.744*	.107	-.269	.021	.265	-.725*	1.557*	1.783*	-.80*9
Std. Error of Kurtosis		.138	.138	.138	.138	.138	.138	.138	.138	.138	.138	.138	.138
Minimum		1.000	1.000	.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Maximum		5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.311	5.234	5.000

* (p<.05)

Table 9*:*

Descriptive Statistics for the Standardised Imputed Data Set

	Zscore(RQ129_SD)	Zscore(RQ131_SD)	Zscore(RQ168_SD)	Zscore(RQ252_SD)	Zscore(Q035_SD)	Zscore(Q090_SD)	Zscore(Q102_SD)	Zscore(Q109_SD)	Zscore(Q188_SD)	Zscore(Q223_SD)	Zscore(RQ27_SD)	Zscore(RQ53_SD)
N Valid	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262	1262
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000
Std. Deviation	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000
Variance	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Minimum	-1.47014	-2.84610	-1.91209	-1.49835	-1.86758	-1.43512	-2.91623	-3.08897	-3.95962	-3.41795	-3.80014	-3.88523
Maximum	2.66102	.92777	1.84242	1.94438	1.54696	2.06929	1.22436	1.61971	1.62523	1.34971	1.50880	1.61254

To identify outliers the item scores were transformed to z-scores. Minimum and maximum z-values for each item on the SD scale are shown in Table 9. Observations with z-scores equal to or larger than 3.0 were yellow-carded as outliers and those with z-scores equal to or larger than 3.5 were red-carded as extreme outliers. The yellow-carded and red- observations are shown in Table 10 below.

Table 10*:*

Outliers for the Standardised Imputed Data Set

		Zscore(Q109_SD)	Case number
		-3.08897	4.00
		-3.08897	6.00
		-3.08897	26.00
		-3.08897	31.00
		-3.08897	33.00
		-3.08897	51.00
		-3.08897	90.00
Total	N	7	7
		Zscore(Q188_SD)	Case number
		-3.95962	1.00*
		-3.95962	2.00
		-3.95962	4.00
		-3.95962	5.00
		-3.95962	8.00
		-3.95962	9.00
		-3.95962	36.00
		-3.95962	49.00
		-3.95962	101.00
		-3.95962	105.00
		-3.95962	118.00
Total	N	11	11
		Zscore(Q223_SD)	Case number
		-3.41795	3.00
		-3.41795	6.00
		-3.41795	10.00
		-3.41795	13.00
		-3.41795	15.00
		-3.41795	19.00
		-3.41795	20.00
		-3.41795	25.00
		-3.41795	32.00
		-3.41795	41.00
		-3.41795	48.00
		-3.41795	53.00
		-3.41795	60.00
		-3.41795	62.00
		-3.41795	66.00
		-3.41795	68.00
		-3.41795	73.00
		-3.41795	74.00
		-3.41795	78.00

		-3.41795	103.00
		-3.41795	108.00
		-3.41795	114.00
		-3.41795	120.00
		-3.41795	140.00
		-3.41795	143.00
		-3.41795	149.00
		-3.41795	180.00
		-3.41795	211.00
Total	N	28	28
		Zscore(RQ27_SD)	Case number
		-3.80014	3.00
		-3.80014	16.00
		-3.80014	27.00
		-3.80014	31.00
		-3.80014	40.00
		-3.80014	46.00
		-3.80014	47.00
		-3.80014	87.00
		-3.80014	102.00
		-3.80014	137.00
		-3.80014	182.00
Total	N	11	11
		Zscore(RQ53_SD)	Case number
		-3.88523	**14.00**
Total	N	1	1

(* red-carded extreme outliers are shown in bold)

All the red-carded, extreme outlier were subsequently deleted from the data set. This reduced the sample size to 1240 observations. The descriptive statistics were recalculated for the reduced sample and are shown in Table 11. Kurtosis reflects the thickness, height or heaviness of the tails of the distributions rather than the "peakedness" of the item distributions (Guilford & Fruchter, 1978). In the case of leptokurtic item distributions the tails tend to be higher/heavier which in turn suggest that more of the item variance is due to infrequently occurring outliers rather than more frequently occurring observations that deviate more moderately from the mean. This is reflected in the reduced magnitude of the kurtosis for Q188_SD, RQ27_SD and RQ53_SD as shown in Table 11 after the deletion of the red-carded extreme outliers. The kurtosis for Q223_SD remained high because yellow-carded extreme outliers were not deleted.

Table 11:

Descriptive Statistics for the Reduced data Set

Statistics

		RQ129_SD	RQ131_SD	RQ168_SD	RQ252_SD	Q035_SD	Q090_SD	Q102_SD	Q109_SD	Q188_SD	Q223_SD	RQ27_SD	RQ53_SD
N	Valid	1240	1240	1240	1240	1240	1240	1240	1240	1240	1240	1240	1240
	Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean		2.43179	4.02500	3.04341	2.74032	3.19032	2.63810	3.82123	3.63352	4.08401	4.04150	3.88819	3.54212
Median		2.00000	4.00000	3.00000	2.00000	3.00000	2.00000	4.00000	4.00000	4.00000	4.00000	4.00000	4.00000
Mode		2.000	5.000	3.000	2.000	4.000	2.000	4.000	4.000	4.000	4.000	4.000	4.000
Std. Deviation		.963549	1.054689	1.062871	1.161803	1.168280	1.137726	.961048	.840754	.719601	.882936	.705085	.900502
Variance		.928	1.112	1.130	1.350	1.365	1.294	.924	.707	.518	.780	.497	.811
Skewness		.516	-.980	.152	.387	-.125	.355	-.702	-.221	-.542	-1.172	-.430	-.367
Std. Error of Skewness		.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069	.069
Kurtosis		.030	.291	-.725	-.804	-.951	-.744	.134	-.313	.296	1.841	.314	.033
Std. Error of Kurtosis		.139	.139	.139	.139	.139	.139	.139	.139	.139	.139	.139	.139
Minimum		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2.000	1.000	2.000	.000
Maximum		5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.000	5.311	5.234	5.000	5.000

4.4 MULTIVARIATE NORMALITY TESTS

Multivariate statistics in general, and particularly structural equation modelling, are based on a number of critical assumptions and it is imperative that the data be assessed to determine the extent to which it complies with these assumptions. This should be done before proceeding with the main analyses (Tabachnick & Fidell, 2007). When analysing continuous data structural equation modelling uses maximum likelihood estimation to estimate the freed measurement and structural model parameters. Maximum likelihood estimation assumes that the indicator variables used to operationalise the latent variables in the structural model follow a multivariate normal distribution. The use of maximum likelihood estimation can result in inappropriate chi-square estimates of model fit and inappropriate standard error estimates (Diamantopoulos & Siguaw, 2000). In the current study a 5-point Likert scale was used to record item responses and hence the data was assumed to be continuous (Muthén & Kaplan, 1985). The null hypothesis that the multivariate normality assumption is satisfied was formally tested in PRELIS.

The outcome of the test for multivariate normality for the reduced imputed data is depicted in Table 12. Special attention should be paid to the chi-square value and the p-value.

Table 12:

Test of Multivariate Normality for Continuous Variables before Normalisation

Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
7.910	26.326	0.000	191.272	15.742	0.000	940.871	0.000

The p-value associated with the chi-square value for skewness and kurtosis indicates that the null hypothesis of multivariate normality was rejected. The multivariate normality assumption made by maximum likelihood estimation techniques that LISREL chooses as a default to obtain parameter estimates in the case of continuous variables was therefore not satisfied. The use of maximum likelihood estimation on the current data could have resulted in an inappropriate chi-square estimate and inappropriate standard error estimates. This potentially could have resulted in incorrect decisions based on the significance of the measurement model parameter estimates. A decision was consequently taken to attempt to normalise the dataset with PRELIS. The result for the normalised data is illustrated below (Table 13).

Table 13*:*

Test of Multivariate Normality for Continuous Variables after normalisation

Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
4.620	15.368	0.000	189.926	15.094	0.000	464.004	0.000

Normalising the data would typically improve the symmetry and kurtosis of the indicator variable distributions. Table 13 indicates that the multivariate normality assumption made by maximum likelihood estimation techniques was still not satisfactory after normalisation. The decrease in the chi-square value indicated that, although the attempt at normalising the data was not successful (p<.05), it did improve the situation as there was a decrease in the chi-square value. Based on these findings it was concluded that the normalisation did not have the desired effect and that the null hypothesis that the multivariate item distribution follows a multivariate normal distribution for the normalised data still had to be rejected (p<.05). The use of an alternative method of estimation that is more suited to data that does not follow a multivariate normal distribution was considered. In the case of non-normal data it is recommended that robust maximum likelihood estimation rather be used to derive model parameter estimates from the observed covariance matrix (Mels, 2003). An asymptotic covariance matrix was consequently computed via PRELIS from the normalised data to enable the more appropriate Satorra-Bentler chi-square fit statistic to be calculated in LISREL.

4.5 ITEM ANALYSIS

The item analysis was performed separately for the *self-deception enhancement* and *impression management* subscales of the SAPI SD-scale. The objective of the item analysis was to evaluate the extent to which the items of the SD-scale successfully fulfil the task they were allocated under the design intention assumed under operational research hypothesis 1. The critical cut-off value for Cronbach alpha was set at .80. According to Field (2013) an alpha value of .70 is also seen as suitable. According to Nunnally and Bernstein (1994) reliability coefficient values are acceptable when they are equal to or above .95 and unacceptable at values below .70. Vorster (2010) claimed that the researcher can accept a reliability coefficient value of .65 if an exploratory research paradigm is used, however a reliability coefficient value lower than .60 cannot be tolerated (Vorster, 2010).

As seen in Table 14 the Cronbach alpha values for both subscales were below the critical cut-off value of .80. Especially the *self-deception enhancement* subscale returned an extremely low Cronbach alpha value of .444. The *impression management* subscale returned a Cronbach alpha value of .553. This resulted in the examination of the inter-item correlations and the items statistics in order to determine whether the internal consistency of the subscales could be improved through the refection or deletion of any items.

Table 14*:*

Reliability Statistics for the SD Subscales

	Subscale	
	Self-deception enhancement	
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.444	.504	5
	Impression management	
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.553	.629	7

In Table 15 the inter-item correlations for the two Social Desirability subscales are given. For both subscales the inter-item correlations were generally low but especially so in the case of the *self-deception enhancement* subscale. The mean inter-item correlation for the *self-deception enhancement* subscale was .145 (table not shown) and for the *impression management* scale, .220 (table not shown). The item-total statistics are shown in Table 16. In the case of the *self-deception enhancement* subscale deletion of none of the items would bring about an increase in the Cronbach alpha value. In the case of the *impression management* subscale the deletion of RQ53_SD was suggested.

According to Worthington and Whittaker (2006) it is unusual to retain all the initial research items and researchers should remove items that fail to contribute meaningfully to any potential factor solutions. Not deleting items that need to be removed makes it difficult to arrive at a final decision regarding the numbers of factors to retain. Worthington and Whittaker (2006) recommended that researchers retain only the potentially meaningful items early in the process in order to optimise scale length once the factor solution is clear.

In the current study, however, all the items were retained for further analysis. The reasoning behind this decision is argued below.

Earlier it was argued that operational research hypotheses 2, 3 and 4 assume that all 12 items of the SAPI SD scale are complex items that load on at least two latent dimensions.

Table 15*:*

Inter-item correlations for the two SD Subscales

Self-deception enhancement					
	RQ27_SD	Q035_SD	RQ168_SD	Q188_SD	Q223_SD
RQ27_SD	1.000	.098	.123	.251	.157
Q035_SD	.098	1.000	.225	.117	.103
RQ168_SD	.123	.225	1.000	.127	.104
Q188_SD	.251	.117	.127	1.000	.166
Q223_SD	.157	.103	.104	.166	1.000

Impression management							
	RQ53_SD	Q090_SD	Q102_SD	Q109_SD	RQ129_SD	RQ131_SD	RQ252_SD
RQ53_SD	1.000	.024	.071	.163	.037	.018	.026
Q090_SD	.024	1.000	.167	.120	.203	.191	.279
Q102_SD	.071	.167	1.000	.198	.073	.171	.163
Q109_SD	.163	.120	.198	1.000	.112	.202	.168
RQ129_SD	.037	.203	.073	.112	1.000	.203	.235
RQ131_SD	.018	.191	.171	.202	.203	1.000	.264
RQ252_SD	.026	.279	.163	.168	.235	.264	1.000

Moreover it is hypothesised that the general factor is unrelated to the two narrower social desirability factors in the measurement models hypothesised by operational research hypotheses 2 and 4. Likewise, under operational research hypothesis 3 the two method factors are hypothesised to be unrelated to the two narrower social desirability factors in the hypothesised measurement model.

Table 16*:*

Item-Total Statistics for the two SD Subscales

Self-deception enhancement					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
RQ27_SD	14.35925	5.361	.244	.085	.390
Q035_SD	15.05712	4.158	.237	.065	.396
RQ168_SD	15.20403	4.390	.254	.071	.374
Q188_SD	14.16343	5.279	.259	.091	.381
Q223_SD	14.20594	5.060	.205	.051	.409

Impression management					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
RQ53_SD	19.28997	12.176	.091	.029	.578
Q090_SD	20.19398	9.974	.318	.121	.498
Q102_SD	19.01086	10.977	.260	.080	.522
Q109_SD	19.19856	11.185	.296	.103	.511
RQ129_SD	20.40030	10.861	.278	.095	.515
RQ131_SD	18.80709	10.187	.335	.130	.492
RQ252_SD	20.09176	9.548	.370	.159	.474

In terms of operational research hypotheses 2, 3 and 4 relatively lower levels of internal consistency should therefore be expected. Categorising items in smaller groups based on their hypothesised loadings on two factors will therefore also not solve the problem since the two factors involved are assumed to be orthogonal. The noisy, low internal consistency data for the two social desirability subscales serves as evidence against the position hypothesised under operational research hypothesis 1 that the SD-scale of the SAPI comprises a set of simple items that measure one of two correlated social desirability factors. The results depicted in Table 14, Table 15 and Table 16 rather unequivocally lead to the conclusion that the items of the SD-scale do not successfully fulfil the task they were allocated under the design intention assumed under operational research hypothesis 1. Although the low internal consistency reliability findings are consistent with operational hypotheses 2, 3 and 4, these findings cannot be regarded as sufficient to conclude support for any these hypotheses. Neither do they assist in deciding which position provides a more plausible account of the mechanism that produced the observed inter-item variance-covariance matrix.

4.6 EXPLORATORY FACTOR ANALYSIS/DIMENSIONALITY ANALYSIS

4.6.1 Dimensionality Analysis Of The Self-Deception Enhancement Subscale

As discussed in Chapter 3, the assumption of unidimensionality was tested in relation to the two subscales of the Social Desirability scale of the SAPI.

Table 17 returned a sufficiently large sample estimate for the Kaiser-Meyer-Olkin measure of sampling adequacy(KMO) for the *self-deception enhancement* subscale, above the recommended value of .60 (Tabachnick & Fidell, 2007) (KMO=.632) and therefore the correlation matrix was considered to be suitable for factor analysis. The identity matrix null hypothesis for the Bartlett's test was rejected (p<.05).

Table 17*:*

KMO and Bartlett's Test for the Self-deception Enhancement Subscale

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.632
Bartlett's Test of Sphericity	Approx. Chi-Square	254.559
	df	10
	Sig.	.000

Table 18 shows that there were two factors with eigenvalues greater than one. In terms of Kaiser (1960) rule of thumb, two factors therefore had to be retained to adequately explain the pattern of inter-item correlations obtained for the *self-deception enhancement* subscale.

Table 18*:*

Total Variance Explained for the Self-deception Enhancement Subscale

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	1.592	31.837	31.837	.813	16.263	16.263	.732
2	1.021	20.417	52.254	.250	5.007	21.270	.613
3	.866	17.323	69.576				
4	.775	15.499	85.075				
5	.746	14.925	100.000				

The results of the parallel analysis performed in SPSS 23 are shown in Table 19. The results of the parallel analysis also suggested the extraction of two factors.

Table 19*:*

Parallel analysis: Self-deception enhancement subscale

Raw Data Eigenvalues		Random Data Eigenvalues		
Root	Eigen.	Root	Means	Prcntyle
1.000000	.666789	1.000000	.083737	.124975
2.000000	.093640	2.000000	.036663	.065988
3.000000	-.074914	3.000000	.001385	.020782
4.000000	-.153728	4.000000	-.032937	-.011470
5.000000	-.168714	5.000000	-.072702	-.043125

The scree plot depicted in Figure 6 provided a rather ambiguous indication of the number of factors that should be extracted. The most prominent elbow seemed to occur at two factors and therefore, in contrast to the directive of the eigenvalue-greater-than-one rule and the results obtained from the parallel analysis, the extraction of only a single factor seemed to be indicated.

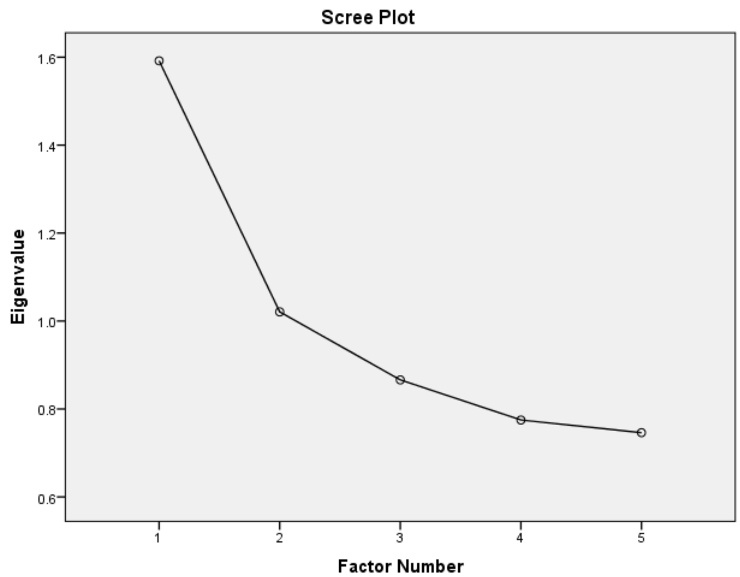

Figure 6. Scree Plot for Factor Analysis(Self-deception enhancement subscale)

The above results indicated that there is sufficient evidence to substantiate the extraction of two factors. The pattern matrix for the rotated two-factor solution is displayed below in Table 20.

The factor matrix interprets how well each item reflects each of the two extracted common factors and from this matrix it is possible to establish whether all the items are satisfactory in terms of the proportion of item variance that can be explained by the factors. An appropriate cut-off for this matrix would be a factor loading of .50 as the factor that the item has been earmarked to represent should explain at least 25% of the variance in the item. Anything less than 25% would mean that more than 75% of the variance in the item is explained by non-relevant error variance and this much error variance should not be present in items. The pattern matrix reflects the partial regression weights when regressing each item on the extracted factors. The oblique rotation allows for correlated factors. The partial regression coefficients reflect the influence of an factor on the item responses when the effect of the other extracted factors in the factor solution are controlled.

The pattern matrix displayed in Table 20 meant that the unidimensionality assumption that all 6 items allocated to the *self-deception enhancement* subscale measure a single undifferentiated dimension of social desirability is not supported. The identity of the two factors were not clear from the wording of the items that loaded on them (see Table 5). The factor loadings were generally moderate to low.

Table 20*:*

Pattern Matrix: Two-factor Solution (Self-deception enhancement subscale)

	Factor	
	1	2
Q188_SD	**.519**	-.016
RQ27_SD	**.511**	-.037
Q223_SD	.290	.066
Q035_SD	-.016	**.485**
RQ168_SD	.027	**.460**

The factor solution shown in Table 20 provided a credible explanation for the observed inter-item correlation matrix in that 0% of the residual correlations were larger than .05. The reproduced correlations derived from the two-factor solution and the residual correlations are shown in Table 21. The two-factor structure displayed in Table 21 therefore provided a valid (i.e. permissible) and credible explanation for the observed inter-item correlation matrix. A lack of support for the unidimensionality assumption was therefore obtained. Confidence in substantive research hypothesis 1 and operational research hypothesis 1 was thereby eroded.

Table 21*:*

Reproduced and residual correlation matrix

		RQ27_SD	Q035_SD	RQ168_SD	Q188_SD	Q223_SD
Reproduced Correlation	RQ27_SD	.243[a]	.102	.118	.251	.157
	Q035_SD	.102	.227[a]	.225	.114	.099
	RQ168_SD	.118	.225	.225[a]	.130	.108
	Q188_SD	.251	.114	.130	.261[a]	.165
	Q223_SD	.157	.099	.108	.165	.108[a]
Residual[b]	RQ27_SD		-.005	.005	1.803E-5	-.001
	Q035_SD	-.005		.000	.003	.004
	RQ168_SD	.005	.000		-.003	-.004
	Q188_SD	1.803E-5	.003	-.003		.001
	Q223_SD	-.001	.004	-.004	.001	

a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 0 nonredundant residuals with absolute values greater than 0.05.

4.6.2 Dimensionality Analysis Of The Impression Management Subscale

Table 22 indicates that a sufficiently large sample estimate for the Kaiser-Meyer-Olkin Measure of Sampling Adequacy(KMO),above the recommended value of .60 (Tabachnick & Fidell, 2007) (KMO = .713) and therefore the correlation matrix was considered to be suitable for factor analysis. The identity matrix null hypothesis for the Bartlett's test was rejected (p<.05).

Table 22*:*

KMO and Bartlett's Test for the Impression Management Subscale

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.713
Bartlett's Test of Sphericity	Approx. Chi-Square	543.221
	df	21
	Sig.	.000

Table 23 shows that there were two factors with eigenvalues greater than one. In terms of Kaiser's rule of thumb two factors should therefore be retained.

Table 23*:*

Total Variance Explained for the Impression Management Subscale

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	1.949	27.842	27.842	1.206	17.229	17.229	1.129
2	1.099	15.707	43.549	.345	4.929	22.158	.676
3	.921	13.150	56.699				
4	.839	11.988	68.687				
5	.761	10.869	79.556				
6	.734	10.492	90.049				
7	.697	9.951	100.000				

The results of the parallel analysis performed on the *impression management* subscale in SPSS 23 are shown in Table 24. The results of the parallel analysis also suggested the extraction of two factors.

Table 24*:*

Parallel analysis: impression management subscale

Raw Data Eigenvalues		Random Data Eigenvalues		
Root	Eigen.	Root	Means	Prcntyle
1.000000	1.067913	1.000000	.113317	.152806
2.000000	.172293	2.000000	.068710	.098788
3.000000	-.008742	3.000000	.033289	.058010
4.000000	-.048031	4.000000	.002778	.023942
5.000000	-.133044	5.000000	-.027364	-.006554
6.000000	-.164677	6.000000	-.059683	-.038200
7.000000	-.169304	7.000000	-.097037	-.070061

The scree plot depicted in Figure 7 provided a rather ambiguous indication of the number of factors that should be extracted. The most prominent elbow seemed to occur at two factors and therefore, in contrast to the directive of the eigenvalue-greater-than-one rule and the results obtained from the parallel analysis, the extraction of only a single factor seemed to be indicated.

.

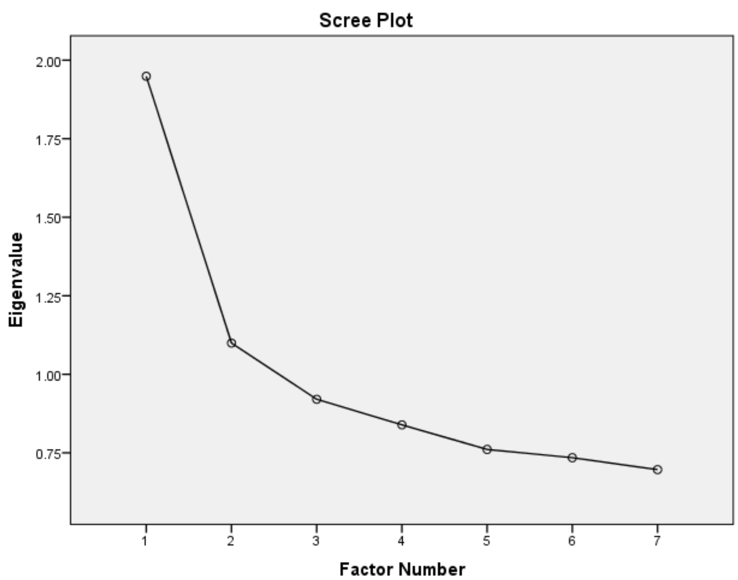

Figure 7. Scree Plot for Factor Analysis(Impression Management subscale)

The above results indicate that there is sufficient evidence to substantiate the extraction of two factors.

The factor matrix displayed in Table 25 indicates the presence of two factors for the *impression management* subscale. The factor loadings were generally moderate to low. Confidence in substantive research hypothesis 1 and operational research hypothesis 1 was thereby eroded.

Table 25*:*

Factor Matrix: Two-factor Solution (Impression Management subscale)

	Factor	
	1	2
RQ252_SD	**.586**	-.025
Q090_SD	**.490**	-.036
RQ131_SD	**.426**	.098
RQ129_SD	**.421**	-.034
Q102_SD	.223	.214
Q109_SD	.098	**.593**
RQ53_SD	-.041	**.270**

The factor solution shown in Table 25 provided a credible explanation for the observed inter-item correlation matrix in that 0% of the residual correlations in Table 26 were larger than .05. The position that the items reflect testees' standing on two underlying common factors rather than a single undifferentiated *impression management* factor is therefore a permissible and credible stance. Q102_SD is a complex item that loaded, although very low, on both of these factors.

Table 26*:*

Reproduced and residual correlation matrix

		RQ53_SD	Q090_SD	Q102_SD	Q109_SD	RQ129_SD	RQ131_SD	RQ252_SD
Reproduced Correlation	RQ53_SD	.067[a]	.018	.067	.157	.015	.049	.026
	Q090_SD	.018	.228[a]	.136	.129	.196	.217	.276
	Q102_SD	.067	.136	.129[a]	.203	.116	.156	.168
	Q109_SD	.157	.129	.203	.403[a]	.109	.194	.166
	RQ129_SD	.015	.196	.116	.109	.168[a]	.186	.237
	RQ131_SD	.049	.217	.156	.194	.186	.221[a]	.264
	RQ252_SD	.026	.276	.168	.166	.237	.264	.334[a]
Residual[b]	RQ53_SD		.006	.004	.005	.022	-.031	-.001
	Q090_SD	.006		.031	-.009	.008	-.026	.003
	Q102_SD	.004	.031		-.005	-.043	.015	-.006
	Q109_SD	.005	-.009	-.005		.003	.008	.002
	RQ129_SD	.022	.008	-.043	.003		.017	-.002
	RQ131_SD	-.031	-.026	.015	.008	.017		-8.588E-5
	RQ252_SD	-.001	.003	-.006	.002	-.002	-8.588E-5	

a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 0 nonredundant residuals with absolute values greater than 0.05.

The dimensionality analysis was performed separately on the two SD subscales. This decision made logical sense in the case of operational research hypothesis 1 but less so when viewed from the perspectives of operational research hypotheses 2, 3 and 4. The measurement models specified under the latter three operational research hypotheses argue that all 12 items of the SAPI SD scale are complex items that simultaneously load on at least two latent dimensions. Moreover it is hypothesised that the general factor is unrelated to the two narrower social desirability factors in the measurement models hypothesised by operational research hypotheses 2 and 4. Likewise, under operational research hypothesis 3 the two method factors are hypothesised to be unrelated to the two narrower social desirability factors in the hypothesised measurement model. To the extent that the dimensionality analyses returned results that fail to confirm the unidimensionality assumption for both the *self-deception enhancement* subscale and the *impression management* subscale it could be interpreted to as evidence that testifies against the measurement model hypothesised under operational research hypothesis 1 and evidence that is compatible with the measurement models hypothesised under operational research hypothesis 2, 3 and 4.

Confirmatory factor analysis was subsequently performed to test the four explanations for the SD scale as hypothesised above.

4.7. EVALUATING THE SINGLE-GROUP MEASUREMENT MODEL FIT VIA CONFIRMATORY FACTOR ANALYSIS IN LISREL

The extent to which the social desirability latent variable was successfully operationalised in terms of the individual items comprising the SD scale of the SAPI was ultimately determined via a series of confirmatory factor analysis (CFA). The SAPI project regrettably never explicitly conceptualised the connotative meaning of the social desirability construct that the SD scale is meant to measure. In the absence of an *a priori* constitutive definition of social desirability the current study inferred two possible conceptualisations of the social desirability construct as measured by the SD scale of the SAPI. Based on these two conceptualisations two substantive research hypotheses were formulated. The literature study, in addition, acknowledged that method bias might explain variance in the SD scale items. The two possible conceptualisations combined with the inclusion and exclusion of the method bias factors resulted in four operational research hypotheses. Each operational research hypothesis postulates a specific measurement model. Each of these four operational research hypotheses were formally tested by evaluating the fit of the corresponding measurement model via CFA.

The operational hypothesis was considered corroborated if its corresponding measurement model fitted the data, (i.e. if the measurement model could successfully reproduce the observed inter-item covariance matrix), if the unstandardised factor loadings were statistically significant ($p<.05$), the completely standardised factor loadings were large (i.e. $\lambda_{ij}>.50$), the unstandardized measurement error variances were statistically significant ($p<.05$), the completely standardised measurement error variances were small (i.e. $\Theta_{\delta}<.75$) and the covariance estimates in Φ were statistically significant but all inter-latent variable correlations (ϕ_{ij}) were small (i.e. $\phi_{ij}< .90$).

4.7.1 Testing operational research hypothesis 1

Operational research hypothesis 1 represents substantive research hypothesis 1 but without assuming method factors. Substantive research hypothesis 1 postulates that the SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a two-dimensional construct comprising the two positively correlated latent dimensions of *self-deception enhancement* and *impression management*. A two-factor measurement model was therefore fitted to the data by having the Social Desirability scale items load on the two factors as they were categorised in Table 5. This model was therefore based on the factor solution implied by the social desirability factors that the items were measuring in the original scales from which they were harvested. A measurement model is a description of the process that is claimed to have brought about the observed covariance matrix. If estimates for the freed model parameters can be found that allows the observed covariance matrix to be accurately reproduced, the measurement model fits the data. The measurement model may then be regarded as a plausible description of the process that brought about the observed covariance matrix. The parameter estimates may then be regarded as credible and worthy of interpretation.

LISREL 8.8 was used to fit the two-factor measurement model without method factors in which each SD-scale item measures one of two positively correlated latent dimensions of *self-deception enhancement* and *impression management*. A visual representation of the measurement model that was fitted to the reduced dataset (n=1240) is displayed in Figure 8.

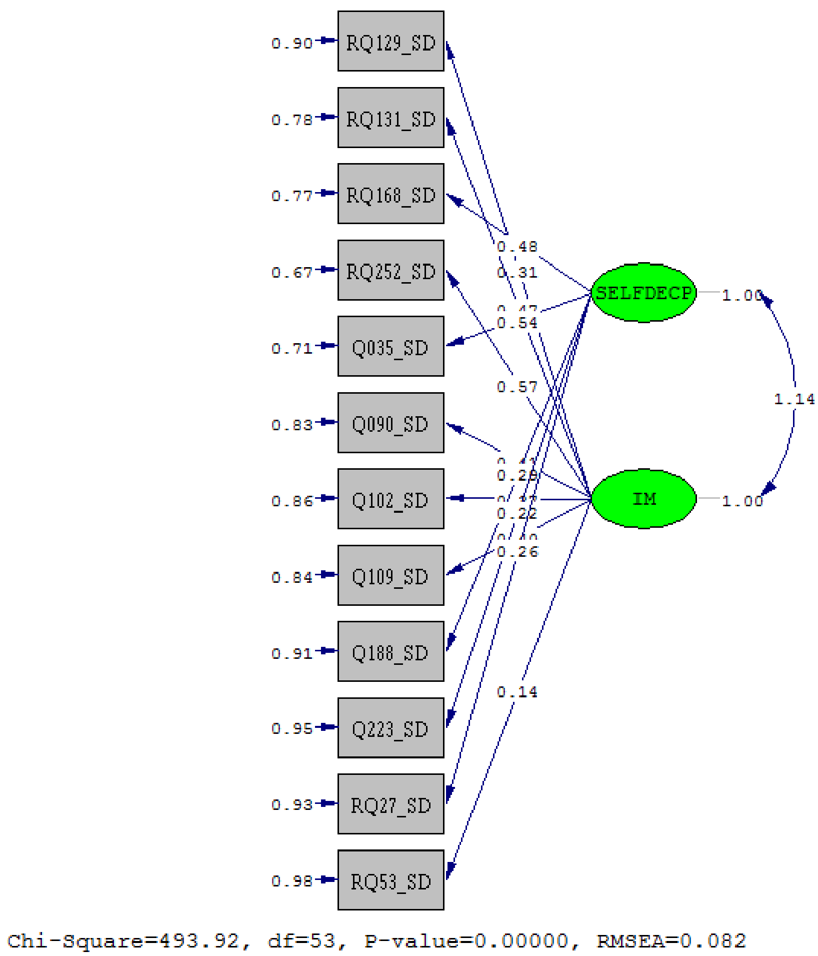

Figure 8: Representation of the fitted two-factor social desirability measurement model hypothesised by operational research hypothesis 1 (completely standardised solution)

The fit statistics for the two-factor measurement model are reported below in Table 27. The indices that are to be reported on are presented in bold in Table 27.

Table 27*:*

Goodness of Fit Statistics for the two-factor Social Desirability Measurement Model (operational hypothesis 1)

Goodness of Fit Statistics
Degrees of Freedom = 53
Minimum Fit Function Chi-Square = 465.062 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 572.396 (P = 0.0)
Satorra-Bentler Scaled Chi-Square = 493.919 (P = 0.0)
Chi-Square Corrected for Non-Normality = 266.333 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 440.919
90 Percent Confidence Interval for NCP = (373.485 ; 515.816)
Minimum Fit Function Value = 0.375
Population Discrepancy Function Value (F0) = 0.356
90 Percent Confidence Interval for F0 = (0.301 ; 0.416)
Root Mean Square Error of Approximation (RMSEA) = 0.0819
90 Percent Confidence Interval for RMSEA = (0.0754 ; 0.0886)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.000
Expected Cross-Validation Index (ECVI) = 0.439
90 Percent Confidence Interval for ECVI = (0.385 ; 0.499)
ECVI for Saturated Model = 0.126
ECVI for Independence Model = 2.148
Chi-Square for Independence Model with 66 Degrees of Freedom = 2637.792
Independence AIC = 2661.792
Model AIC = 543.919
Saturated AIC = 156.000
Independence CAIC = 2735.266
Model CAIC = 696.991
Saturated CAIC = 633.584
Normed Fit Index (NFI) = 0.813
Non-Normed Fit Index (NNFI) = 0.787
Parsimony Normed Fit Index (PNFI) = 0.653
Comparative Fit Index (CFI) = 0.829
Incremental Fit Index (IFI) = 0.829
Relative Fit Index (RFI) = 0.767
Critical N (CN) = 201.289
Root Mean Square Residual (RMR) = 0.0562
Standardized RMR = 0.0654
Goodness of Fit Index (GFI) = 0.929
Adjusted Goodness of Fit Index (AGFI) = 0.895
Parsimony Goodness of Fit Index (PGFI) = 0.631

The exceedence probability associated with the Satorra-Bentler Scaled chi-square test statistic indicated that the null hypothesis of the exact fit (H_{011}: RMSEA=0) is rejected. This result was expected as exact fit represents a somewhat unrealistic position in that it states that the single-group measurement model is able to reproduce the observed sample covariance matrix to a degree of accuracy that could be explained solely in terms of sampling error.

The operational hypothesis that is represented by the close fit null hypothesis (H_{021}) assumes that the measurement model describes an approximation of the process that operated in reality to create the observed covariance matrix (Browne & Cudeck, 1993). The Root Mean Square of Error of Approximation (RMSEA) is a test statistic that is of importance. It is a popular measure of fit that articulates the difference between the observed and the estimated sample covariance matrices. According to Diamantopoulos and Siguaw (2000) it is one of the most informative fit indices as it takes the complexity of the model into consideration.

In Table 27 the RMSEA value is reported as .0819 and this indicates that the measurement model shows only reasonable fit in the sample. The p-value for the test of close fit provides further evidence for the hypothesis (H_{021}: RMSEA $\leq$ .05) and is reported in Table 27 as approximating 0. Therefore, the close fit null hypothesis H_{021} : RMSEA $\leq$ .05 was rejected. This meant that the measurement model does not show close fit in the parameter although it did show reasonable fit in the sample. It therefore had to be concluded that the two-factor measurement model did not offer a plausible description of the process that created the observed covariance matrix in that the model failed to reproduce the observed covariance matrix to a sufficient degree of accuracy. Confidence intervals assess the precision of the RMSEA estimates and the fit statistics in Table 27 reported the upper and lower bounds of the 90% interval to be .0754 to .0886. Interpretation of the confidence interval indicated that the true RMSEA value in the population fell within the relatively narrow bounds of .0754 and .0886, which presents a high degree of precision (Byrne, 2001) but lack of close fit since the confidence interval excluded the critical close fit cut-off value for RMSEA of .05.

The parameter estimates of the two-factor measurement model were consequently not interpreted.

A visual representation of the modification indices calculated for the two-factor measurement model that was fitted to the reduced dataset (N=1240) is displayed in Figure 9. The large number of statistically significant ($p<.05$) modification index values obtained for the currently fixed measurement error covariances in $\boldsymbol{\Theta}_{\delta}$were noteworthy. Allowing the measurement error terms associated with the SD-scale items to correlate would statistically significantly ($p<.05$) improve the fit of the two-factor measurement model.

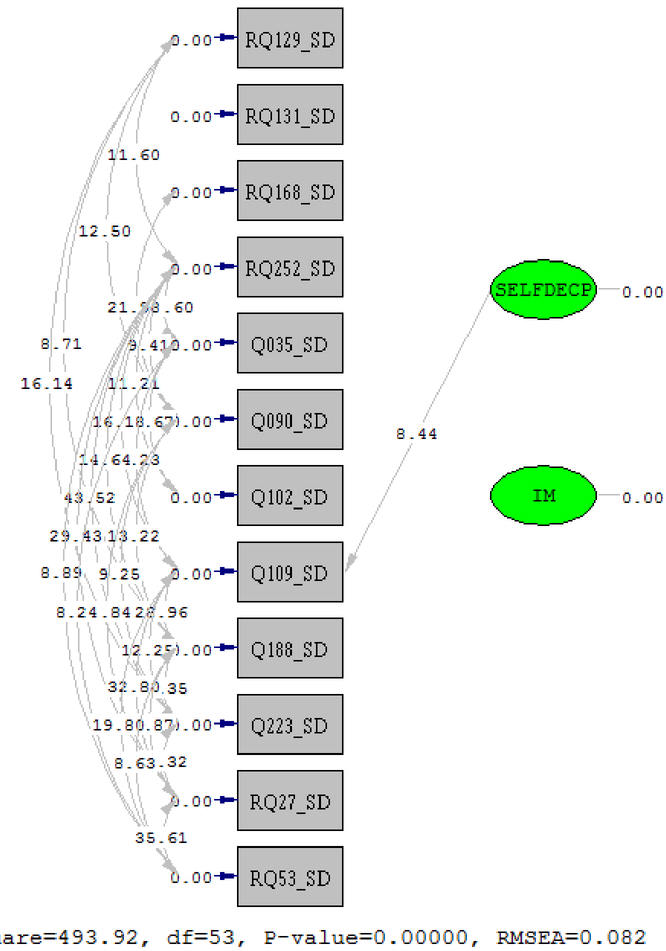

Figure 9: Representation of the modification indices calculated for the fitted two-factor social desirability measurement model

These large number of statistically significant modification index values for Θ_δ suggested that the items of the SD-scale share one or more other common source of variance that the current measurement model ignores. This in turn bolstered confidence in operational research hypotheses 2, 3 and 4.

4.7.2 Testing Operational Research Hypothesis 2

Operational research hypothesis 2 represents substantive research hypothesis 2 but without assuming method factors. Substantive research hypothesis 2 postulates that the SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a three dimensional construct comprising the two positively correlated latent dimensions of *self-deception enhancement and impression management*, but also a broad, general social desirability latent dimension, uncorrelated with the two narrower social desirability dimensions, representing the common variance shared by all twelve of the SAPI's SD items. A three-factor measurement model was therefore fitted to the data where the SD-scale items loaded on the two narrow factors social desirability factors as they were categorised in Table 3 and where all items loaded on the broad, general social desirability factor.

The three-factor model without method factors converged in 34 iterations. The completely standardised solution is shown in Figure 10.

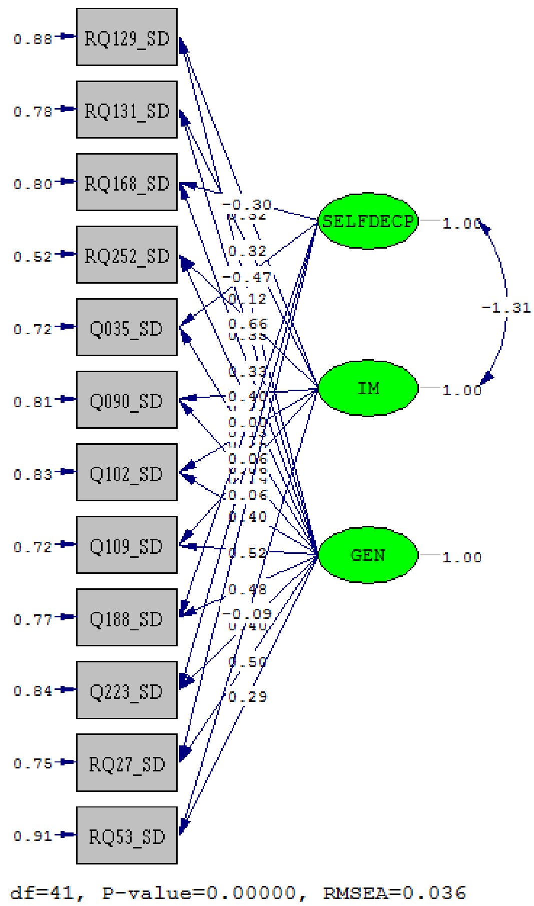

Figure 10: Representation of the fitted three-factor social desirability measurement model hypothesised by operational research hypothesis 2 (completely standardised solution)

For the three-factor measurement model the goodness of fit statistics are reported below in Table 28.

Table 28*:*

Goodness of Fit Statistics for the three-factor Social Desirability Measurement Model (operational hypothesis 2)

Goodness of Fit Statistics
Degrees of Freedom = 41
Minimum Fit Function Chi-Square = 120.158 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 123.862 (P = 0.00)
Satorra-Bentler Scaled Chi-Square = 107.799 (P = 0.000)
Chi-Square Corrected for Non-Normality = 99.154 (P = 0.000)
Estimated Non-centrality Parameter (NCP) = 66.799
90 Percent Confidence Interval for NCP = (39.821 ; 101.448)
Minimum Fit Function Value = 0.0970

Population Discrepancy Function Value (F0) = 0.0539
90 Percent Confidence Interval for F0 = (0.0321 ; 0.0819)
Root Mean Square Error of Approximation (RMSEA) = 0.0363
90 Percent Confidence Interval for RMSEA = (0.0280 ; 0.0447)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.997
Expected Cross-Validation Index (ECVI) = 0.147
90 Percent Confidence Interval for ECVI = (0.125 ; 0.175)
ECVI for Saturated Model = 0.126
ECVI for Independence Model = 2.148
Chi-Square for Independence Model with 66 Degrees of Freedom = 2637.792
Independence AIC = 2661.792
Model AIC = 181.799
Saturated AIC = 156.000
Independence CAIC = 2735.266
Model CAIC = 408.346
Saturated CAIC = 633.584
Normed Fit Index (NFI) = 0.959
Non-Normed Fit Index (NNFI) = 0.958
Parsimony Normed Fit Index (PNFI) = 0.596
Comparative Fit Index (CFI) = 0.974
Incremental Fit Index (IFI) = 0.974
Relative Fit Index (RFI) = 0.934
Critical N (CN) = 747.516

Root Mean Square Residual (RMR) = 0.0295
Standardized RMR = 0.0278
Goodness of Fit Index (GFI) = 0.984
Adjusted Goodness of Fit Index (AGFI) = 0.969
Parsimony Goodness of Fit Index (PGFI) = 0.517

The exceedence probability associated with the Satorra-Bentler Scaled chi-square test statistic indicated that the null hypothesis of the exact fit (H_{012}: RMSEA=0) is rejected. This result was expected as exact fit represents a somewhat unrealistic position in that it states that the single-group measurement model is able to reproduce the observed covariance matrix to a degree of accuracy that could only be explained by sampling error.

In Table 28 the RMSEA value is reported as .036 and this indicates that the measurement model shows close fit in the sample. The p-value for the test of close fit provides further evidence for the hypothesis (H_{022}: RMSEA >.05) and is reported in Table 28 as .997. The probability of observing an RMSEA estimate of .0363 in the sample if the RMSEA value in the parameter was .05, is therefore quite high. The close fit null hypothesisH_{022}: RMSEA $\leq$ .05 was therefore not rejected. This meant that the measurement model does show close fit in the parameter in the sample. It therefore had to be concluded that the three-factor measurement model did offer a plausible description of the process that created the observed covariance matrix as the model succeeded in reproducing the observed covariance matrix to a sufficient degree of accuracy. The remaining model fit indices provide further support for this.

Confidence intervals assess the precision of the RMSEA estimates and the fit statistics in Table 28 reported the upper and lower bounds of the 90% interval to be .0280to .0447. Interpretation of the confidence interval indicated that the true RMSEA value in the population fell within the relatively narrow bounds of .0280 and .0447 which presents a high degree of precision (Byrne, 2001), but more importantly, fell below the critical cut-off value of .05.

If a researcher needed to evaluate whether a model that fits the sample in a current study would also fit an in independent sample of the same size, from the same population they would take the expected cross-validation index (ECVI) into consideration. The reason being is that the ECVI expresses the difference between the reproduced sample covariance matrix derived from fitting the model on the sample at hand, and the expected covariance matrix that would be found in an independent sample (Byrne, 1989; Diamantopoulos & Siguaw, 2000). In this model the ECVI (.147) is smaller than the value obtained for the independence model (2.148) but larger than the ECVI value associated with the saturated model (.126).A model more closely resembling the saturated model therefore seems to have a better chance of being replicated in a cross-validation sample than the fitted model.

There may be cause for improvements in model fit through the addition of more paths to model and the estimation of more parameters until perfect fit is achieved in the form of a saturated or just-identified model with no degrees of freedom (Kelloway, 1998). This is acknowledged by the parsimonious fit. According to Jöreskog and Sörbom (1996) it is essential to find the most parsimonious model that achieves satisfactory fit with as few model parameters as possible when defining and fitting models.

The parsimonious normed fit index (PNFI = .596) and the parsimonious goodness-of-fit index (PGFI = .517) approach model fit from this perspective. PNFI and PGFI range from 0 to 1, with higher values indicating a more parsimonious fit. There is no standard for how high either index should be to indicate parsimonious fit (Kelloway, 1998). However, the PNFI is not close to reaching the .90 cut-off used for other fit indices. According to Kelloway (1998) these indices are more meaningfully used when comparing two competing theoretical models and are not very useful indicators in this CFA analysis.

The values for this model's Akaike information criterion (AIC= 181.799) suggested that the fitted measurement model provided a more parsimonious fit than the independent model (2661.792) but not the saturated model (156.000) since smaller values on these indices indicate a more parsimonious model (Kelloway, 1998). This indicated that the measurement model may lack influential paths. Values for the consistent Akaike information criterion (408.346) implied that the fitted measurement model provided a more parsimonious fit than both the independent model (2735.266) and the saturated model (633.584). This provided further support for the fitted model.

There are indicators of comparative fit that contrast the ability of the model to reproduce the observed covariance matrix with that of a model known *a priori* to fit the data poorly if the baseline used for this is an independent or null model. The fit indices presented include the normed fit index (NFI= .959), the non-normed fit index (NNFI= .958), the comparative fit index (CFI= .974), the incremental fit index (IFI=.974) and the relative fit index (RFI =.934).The closer the values are to unity, the better the fit. However, .90 could be considered indicative of a well-fitting model (Diamantopoulos & Siguaw, 2000; Kelloway, 1998). In the current results, all of these indices substantially exceeded the .90 level, which would be indicative of satisfactory comparative fit relative to the independence model.

The critical sample size statistic (CN) represents Hoelter's critical N (Bollen & Liang, 1988) and refers to the largest sample size that the sample could reach in order to not reject the exact fit null hypothesis given the obtained sample RMSEA estimate due to the $\chi 2$ statistic not being significant at the .05 significance level (Bollen, 1988; Diamantopoulos & Siguaw, 2000). The CN statistic acknowledges the fact that trivial differences between the population covariance matrix and the population estimated covariance matrix (i.e small RMSEA values in the parameter) can result in a significant χ^2 sample estimate when the sample size is large. When the estimated discrepancy between the two matrices as expressed by RMSEA is so small that even at reasonably large samples the exact fit null hypothesis would not rejected, it boosts confidence in the model. Hoelter (as cited in Bollen & Liang, 1988, p. 494) suggested the critical CN value to be 200. Bollen (1988) reports that models with CN values substantially larger than the critical cut-off value of 200 are regarded as better fitting than models with CN values that fall only marginally above the cut-off. The estimated CN value (747.516) in the current study fell way above the recommended threshold value of 200. This threshold is regarded as indicative of the model providing an adequate representation of the data (Diamantopoulos &Siguaw, 2000) although this proposed threshold should be used with caution (Hu & Bentler, 1995).

The standardized RMR may be considered a summary measure of standardized residuals which represents the average difference between the elements of the sample covariance matrix and the fitted covariance matrix. If the model fit is good, the fitted residuals (**S**^ ̂) should be small in comparison to the magnitude of the elements in **S** (Diamantopoulos & Siguaw, 2000). The RMR (.0295) and standardized RMR (.0278) indicated reasonable fit as values less than .05 on the latter index suggest the model fits the data well (Kelloway, 1998).

The goodness-of-fit index (GFI) and the adjusted goodness-of-fit index (AGFI) reflect how closely the model comes to perfectly reproducing the sample covariance matrix (Diamantopoulos & Siguaw, 2000). The AGFI (.969) adjusts the GFI (.984) for the degrees of freedom in the model (Diamantopoulos & Siguaw, 2000; Jöreskog & Sörbom, 1993) and should be between zero and 1.0 with values exceeding .90 indicating that the model fits the data well (Jöreskog & Sörbom, 1993; Kelloway, 1998). For the fit of this model, both the GFI and AGFI are above the acceptable cut-off level.

In conclusion, when the abovementioned model fit statistics were considered holistically they seem to suggest a well-fitting model. In addition, when taking the fitted model, independence model and saturated model into account, evidence was provided in support of the fitted model, however, the model fit may possibly benefit from the inclusion of a few additional paths.

Should the researcher have any questions regarding whether the currently fixed parameters, when freed in the model, would significantly improve the parsimonious fit of the model then the modification indices (MI) should be looked at. These indices provide an indication of whether or not the chi-square fit statistic will statistically significantly decrease if the model is re-estimated when a currently fixed parameter in the model is freed (Jöreskog & Sörbom, 1993). Large modification index values (>6.6349) would be indicative of parameters that, if set free, would improve the fit of the model significantly ($p<.01$) (Diamantopoulos & Siguaw, 2000; Jöreskog &

Sörbom, 1993). The evaluation of the modification indices calculated for $\mathbf{\Lambda}$ and $\mathbf{\Theta}_\delta$ in the current study did not constitute an attempt to improve the fit of the model but rather represented an additional perspective on the fit of the model. If only a limited number of ways would exist to improve the fit of the model, then this would reflect favourably on the substantive measurement hypothesis represented by this specific measurement model. Freeing two of the twelve fixed elements in $\mathbf{\Lambda}$ (17%) and 10 of the 30 fixed elements in $\mathbf{\Theta}_\delta$ (33%) would have resulted in a statistically significant (p<.05) in model fit. These results comment relatively favourably on the fit of the three-factor model.

As the model fitted the data closely, the model parameter estimates were considered credible, valid and therefore worthy of interpretation. This model represented the hypothesis that each SD item measures two constructs; either self-deception and the general SD factor or *impression management* and the general SD factor. The questions that had to be answered through the interpretation of the parameter estimates were whether both factor loadings for each item were statistically significant (p<.05) and sufficiently large in magnitude and whether the measurement error variance for each item was statistically significant (p<.05) but sufficiently small.

4.7.2.1. Interpreting the freed measurement model parameter estimates

The unstandardized factor loading matrix is shown in Table 29. These unstandardized lambda-X coefficients represent the factor loadings of the items on their designated factors.. As such they reflect the slope of the regression of the item on one underlying factor when holding the other factor constant. In $\mathbf{\Lambda}^{\mathbf{X}}$ shown in Table 29 three values are given. The top value is the unstandardized estimate of the loadings of the item on the latent variable. The second value is the standard error of the factor loading estimate. The third value is a test statistic value that is used to evaluate the statistical significance of the unstandardized estimate. The critical test statistic value is determined by the nature of the alternative hypothesis against which the null hypothesis is evaluated. Since H_{api}: $\lambda_{jk} \neq 0$;p=14, 15, …, 37; i=2, 4; j=1, 2, …, 12; k=1, 2, 3 were formulated as non-directional alternative hypotheses, a critical test statistic value of z=|1.96| was used on a 5% significance level.

From the output in Table 29 one can see that all the path coefficients were statistically significant looking at z≥|1.96| except for Q188_SD, Q223_SD and RQ27_SD for the factor of *self-deception enhancement* and Q109_SD and RQ53_SD for the factor of impression management. All items loaded statistically significantly ($p<.05$) on the general SD factor. $H_{015,2}$, $H_{017,2}$, $H_{019,2}$, $H_{021,2}$,$H_{023,2}$, $H_{025,2}$, $H_{027,2}$, $H_{029,2}$, $H_{031,2}$, $H_{033,2}$, $H_{035,2}$ and $H_{037,2}$, therefore had to be rejected. Only 2 of the 5 *self-deception enhancement* items loaded statistically significantly ($p<.05$) on the *self-deception enhancement* factor while 5 of the 7 *impression management* items loaded statistically significantly ($p<.05$) on the *impression management* factor.$H_{016,2}$ and $H_{030,2}$ could be rejected whereas $H_{014,2}$, $H_{032,2}$ and $H_{034,2}$ could not be rejected. Items RQ168_SD and Q35_SD loaded statistically significantly ($p<.05$) on the *self-deception enhancement* factor as hypothesised under operational hypothesis 2, whereas items Q188_SD, RQ27_SD and Q223_SD loaded statistically insignificantly on the self-deception factor. $H_{020,2}$, $H_{022,2}$,$H_{026,2}$, $H_{028,2}$, $H_{036,2}$ could be rejected whereas $H_{018,2,}$and $H_{024,2}$could not be rejected., Q90_SD, RQ129_SD, RQ131_SD and RQ252 loaded statistically significantly ($p<.05$) on the *impression management* factor as hypothesised under operational hypothesis 2. Item Q109_SD and RQ53_SD were the only items that should have reflected the *impression management* factor according to operational hypothesis 2 that loaded statistically insignificantly on the second factor. This result erodes, to some degree, the confidence in the model. The items that load insignificantly on the separate factors need to be attended to if this model would be accepted as the most plausible model and the recommendation is that perhaps they are reworded as they are tapping into the general factor but not the sub-factor.

Table 29*:*

SAPI SD 3-factor measurement model unstandardised lambda-X matrix (operational hypothesis 2)

	SELFDECP	IM	GEN
RQ27_SD	0.045	- -	0.352*
	(0.046)		(0.028)
	0.965		12.494
Q035_SD	-0.544*	- -	0.297*
	(0.070)		(0.092)
	-7.794		3.232
RQ53_SD	- -	-0.083	0.260*
		(0.050)	(0.038)
		-1.660	6.797
Q090_SD	- -	0.450*	0.217*

		(0.053)	(0.081)
		8.553	2.687
Q102_SD	- -	0.122*	0.381*
		(0.062)	(0.039)
		1.969	9.794
Q109_SD	- -	0.054	0.440*
		(0.069)	(0.032)
		0.777	13.798
RQ129_SD	- -	0.312*	0.118*
		(0.040)	(0.060)
		7.769	1.970
RQ131_SD	- -	0.332*	0.368*
		(0.063)	(0.063)
		5.256	5.868
RQ168_SD	-0.315*	- -	0.350*
	(0.060)		(0.063)
	-5.286		5.576
Q188_SD	0.000	- -	0.347*
	(0.046)		(0.025)
	0.005		13.765
Q223_SD	0.050	- -	0.351*
	(0.048)		(0.033)
	1.050		10.488
RQ252_SD	- -	0.763*	0.258*
		(0.062)	(0.127)
		12.268	2.040

* (p<.05)

In Table 30 the completely standardised solution for lambda-X is provided in which both the latent and observed variables are standardised to have a mean of zero and a standard deviation of one. The factor loadings represent the slope of the regression of the item on the latent variables it was earmarked to reflect when controlling for the other factors in the measurement equation. In the unstandardized solution the metric of the latent variables are unknown and it could potentially differ across latent variables. In the completely standardised solution the standard deviation becomes the metric of both the observed and latent variables (Diamantopoulos & Siguaw, 2000). The completely standardised factor loadings in Table 30 therefore represent the average change on the item, expressed in standard deviation units, associated with one standard deviation change in the latent SD dimension being measured when holding the other latent SD dimension being measured constant.

Table 30 indicates that all the completely standardised factor loadings fell below the set critical factor loading value of .50. This, along with the findings on the statistical significance of the factor loadings on the two narrower SD factors, testify against operational hypothesis 2.2. Support is therefore not obtained for operational hypothesis 2.2 that posited that the factor loadings of the items on their designated latent social desirability dimensions (*self-deception enhancement* and the general social desirability factor or *impression management* and the general social desirability factor) are statistically significant ($p<.05$) and large ($\lambda_{ij}\geq 50$). As the items reflect two factors, it may seem unreasonable to assume or to expect that they would load as high as .50 on each factor they are meant to represent.

Table 30*:*

SAPI SD 3-factor measurement model completely standardised solution lambda-X matrix (operational hypothesis 2)

	SELFDECP	IM	GEN
RQ27_SD	0.064	- -	0.499
Q035_SD	-0.466	- -	0.254
RQ53_SD	- -	-0.092	0.289
Q090_SD	- -	0.396	0.191
Q102_SD	- -	0.127	0.396
Q109_SD	- -	0.064	0.523
RQ129_SD	- -	0.323	0.122
RQ131_SD	- -	0.315	0.349
RQ168_SD	-0.296	- -	0.329
Q188_SD	0.000	- -	0.482
Q223_SD	0.057	- -	0.398
RQ252_SD	- -	0.657	0.222

In Table 31 the estimates for the freed theta-delta coefficients have been provided and these represent the sample estimates of the variance in measurement error terms. Table 31indicates that all the items are statistically significantly ($p<.05$) plagued by measurement error. Since H_{api}: $\Theta_{\delta ijj}> 0$; p=14, 15, 16…37; i =1, 2, 3, 4; j=1, 2 ....., 12 have been formulated as directional hypotheses, H_{0pi}: $\Theta_{\delta jj}=0$; p=14, 15, 16…37; i =1, 2, 3, 4; j=1, 2, ..., 12 were tested via one-sided statistical tests. A critical z-value of 1.6449 was therefore used as the critical value to evaluate the statistical significance of the $\theta_{\delta ii}$ estimates shown in Table 31. This finding is welcomed since a finding of perfectly valid and reliable measures of the three SD factors would have been a too improbable outcome.

Table 31*:*

Three-factor SAPI SD measurement model unstandardised theta-delta matrix (operational hypothesis 2)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.371*	0.981*	0.736*	1.045*	0.764*	0.510*
(0.022)	(0.064)	(0.034)	(0.047)	(0.038)	(0.029)
16.774	15.270	21.408	22.252	20.302	17.691
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.817*	0.866*	0.908*	0.398*	0.654*	0.701*
(0.035)	(0.039)	(0.042)	(0.022)	(0.033)	(0.057)
23.631	21.989	21.399	18.440	19.566	12.294

* ($p<.05$)

The completely standardised $\theta_{\delta ii}$ estimates are shown in Table 32. Table 32 indicates that 11 of the 12 SAPI SD items provide measures of the SD factors they were earmarked to reflect in terms of substantive research hypothesis 2, that are generally highly contaminated by systematic and/or random measurement error. RQ252_SD is the only exception. In all the remaining items more than 50% of the variance in the item responses are due to systematic and/or random measurement error. In the case of 9 of the 12 items, more than 70% of the variance in the item responses cannot be attributed to the latent variables that the items were meant to reflect. Table 32 displays the magnitude of the error variance and this shows that the items are rather noisy in that more than half of the item variance (if the $\theta_{\delta ii}$ value is > .5) is not due to the latent variables/factors the item was earmarked to reflect.

Table 32*:*

Three-factor SAPI SD measurement model completely standardised solution theta-delta matrix (operational hypothesis 2)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.747	0.718	0.908	0.807	0.827	0.722
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.880	0.779	0.804	0.768	0.839	0.519

In Table 33 the completely standardised phi matrix is provided. This matrix shows the correlations between the latent variables and is therefore the observed correlations between the observed scores on the *impression management* and *self-deception enhancement* dimensions corrected for the attenuating effect of measurement error. Table 33 indicates that *impression* *management* correlates statistically significant ($p<.05$) and negatively with *self-deception enhancement*. H_{0pi}: $\phi_{jk}=0$; p=136,; i =2; j=1; k=2 could therefore not be rejected despite the fact that the ϕ_{21} estimate was statistically significant ($p<.05$) since H_{api}:.$\phi_{jk}>0$; p=136,; i =2; j=1; k=2 was formulated as directional and positive. Even more disturbing though is the fact that ϕ_{12} has an inadmissible value that exceeds unity. This seriously erodes confidence in the model hypothesised under operational hypothesis 2. Table 33 also reflects the fact that the measurement model under operational hypothesis 2 fixed the correlations between the two narrow social desirability factors and the general social desirability factor to zero. According to operational hypotheses 2 there is no correlation between the general SD factor and the other two factors.

Table 33*:*

Three-factor SAPI SD measurement model completely standardised phi matrix (operational hypothesis 2)

SELFDECP	IM	GEN
SELFDECP	1.000	
IM	-1.311* (0.120) -10.900	1.000
GEN	- -	1.000

* (p<.05)

In Table 34 the squared multiple correlations for X-variables are provided and these correlations indicate the proportion of variance in the observed variable explained by the latent variable/ variables linked to it in the measurement model. High correlation values are preferred. Table 34 echoes the results obtained in Table 32. The latent variables that the items were designed to reflect generally explain less than 50% of the variance in each item. RQ252_SD is the only exception.

Table 34*:*

Three-factor SAPI SD measurement model squared multiple correlations for X-variables (operational hypothesis 2)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.253	0.282	0.092	0.193	0.173	0.278
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.120	0.221	0.196	0.232	0.161	0.481

This operational hypothesis postulated that SAPI SD scale measures a three-dimensional social desirability construct comprising two positively correlated narrow social desirability factors (*self-deception enhancement and impression management*) and a broad, general social desirability factor, uncorrelated with the two narrower social desirability dimensions. The three-factor measurement model did offer a plausible description of the observed variance-covariance matric S in that estimates for the freed model parameters could be obtained that allowed the reproduction of S to a sufficient degree of accuracy. In addition support was obtained for the position that all the items in the SD scale measure the broad, general social desirability factor. However, support was not obtained for five items (Q109_SD, Q188_SD,Q223_SD and RQ53_SD) that the SD items at the same time also measure a specific narrow social desirability factors. This eroded confidence in the measurement model postulated by operational hypothesis 2 as the most plausible account of the process underpinning the SD scale. In addition the inadmissible and negative ϕ_{12} estimate also wore away confidence in operational hypothesis two as a plausible description of the mechanism that produced the observed variance-covariance matrix.

4.7.3 Testing Operational Research Hypothesis 3

Operational research hypothesis 3 represents substantive research hypothesis 1 but also assumes two method factors. Substantive research hypothesis 1 postulates that the SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a two-dimensional construct comprising the two positively correlated latent dimensions of *self-deception enhancement* and *impression management*. In addition to the two substantive social desirability factors, operational research hypothesis 3 postulates that two method factors also explain systematic variance in the SD scale items due to the positive versus negative wording of the items. A four-factor measurement model was therefore fitted to the data where the SD-scale items loaded on the two narrow factors social desirability factors as they were categorised in Table 5 and where the six positively worded items loaded on a positively keyed factor and the six negatively worded items loaded on a negatively keyed factor. The correlations between the two method factors and the two narrow SD factors were fixed to zero.

The four-factor model converged in 70 iterations. The completely standardised solution is shown in Figure 11.

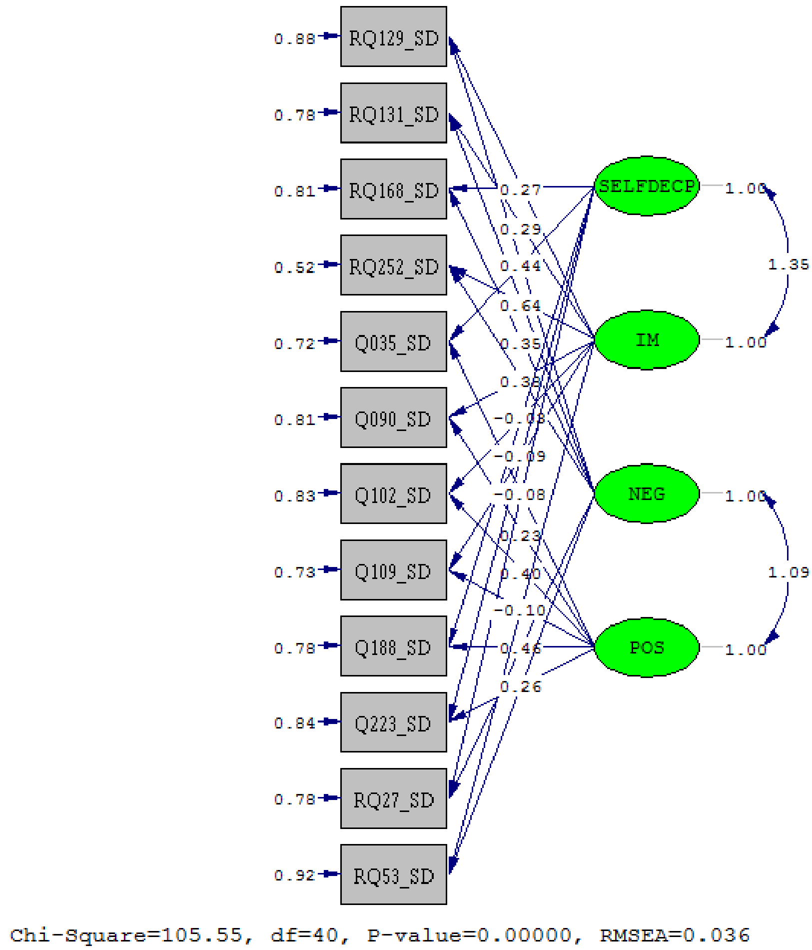

Figure 11: Representation of the fitted two-factor social desirability measurement model hypothesised by operational research hypothesis 3 with method factors (completely standardised solution)

Table 35:

Goodness of Fit Statistics for the 4-factor Social Desirability Measurement Model (operational hypothesis3)

Goodness of Fit Statistics
Degrees of Freedom = 40
Minimum Fit Function Chi-Square = 117.342 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 120.999 (P = 0.00)
Satorra-Bentler Scaled Chi-Square = 105.546 (P = 0.000)
Chi-Square Corrected for Non-Normality = 97.054 (P = 0.000)

Estimated Non-centrality Parameter (NCP) = 65.546
90 Percent Confidence Interval for NCP = (38.876 ; 99.885)
Minimum Fit Function Value = 0.0947
Population Discrepancy Function Value (F0) = 0.0529
90 Percent Confidence Interval for F0 = (0.0314 ; 0.0806)
Root Mean Square Error of Approximation (RMSEA) = 0.0364
90 Percent Confidence Interval for RMSEA = (0.0280 ; 0.0449)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.996
Expected Cross-Validation Index (ECVI) = 0.147
90 Percent Confidence Interval for ECVI = (0.125 ; 0.174)
ECVI for Saturated Model = 0.126
ECVI for Independence Model = 2.148
Chi-Square for Independence Model with 66 Degrees of Freedom = 2637.792
Independence AIC = 2661.792
Model AIC = 181.546
Saturated AIC = 156.000
Independence CAIC = 2735.266
Model CAIC = 414.215
Saturated CAIC = 633.584
Normed Fit Index (NFI) = 0.960
Non-Normed Fit Index (NNFI) = 0.958
Parsimony Normed Fit Index (PNFI) = 0.582
Comparative Fit Index (CFI) = 0.975
Incremental Fit Index (IFI) = 0.975
Relative Fit Index (RFI) = 0.934
Critical N (CN) = 748.669
Root Mean Square Residual (RMR) = 0.0287
Standardized RMR = 0.0277
Goodness of Fit Index (GFI) = 0.984
Adjusted Goodness of Fit Index (AGFI) = 0.969
Parsimony Goodness of Fit Index (PGFI) = 0.505

The exceedence probability associated with the Satorra-Bentler Scaled chi-square test statistic indicated that the null hypothesis of the exact fit (H_{013}: RMSEA=0) is rejected. This result was expected as exact fit represents a somewhat unrealistic position in that it states that the single-group measurement model is able to reproduce the observed covariance matrix to a degree of accuracy that could only be explained by sampling error.

In Table 35 the RMSEA value is reported as .0364 and this indicates that the measurement model shows close fit in the sample. The p-value for the test of close fit provides further evidence for the hypothesis (H_{023} : RMSEA >.05) and is reported in Table 35 as .996. Therefore, the close fit null hypothesis H_{023} : RMSEA ≤ .05 was not rejected. This meant that the claim that the measurement model shows close fit in the parameter may be regarded as permissible. It therefore had to be concluded that the four-factor measurement model did offer a plausible description of the process that created the observed covariance matrix as the model succeeded in reproducing the observed covariance matrix to a sufficient degree of accuracy. Confidence intervals assess the precision of the RMSEA estimates and the fit statistics in Table 35 reported the upper and lower bounds of the 90% interval to be .0280 to .0449. Interpretation of the confidence interval indicated that the true RMSEA value in the population fell within the relatively narrow bounds of .0280 and .0449 which presents a high degree of precision (Byrne, 2001). In addition both the upper and lower bound of the confidence interval fell below the critical cut-off value of .05

A model more closely resembling the saturated model seems to have a better chance of being replicated in a cross-validation sample than the fitted model. This is deduced from the fact that the model ECVI (0.147) was smaller than the value obtained for the independence model (2.148) but larger than the ECVI value associated with the saturated model (0.126).

The parsimonious normed fit index (PNFI) and parsimonious goodness-of-fit index (PGFI) range from 0 to 1, with higher values indicating a more parsimonious fit. The PNFI for this model was .582 which is quite low when taking into account the .90 cut-off used for other fit indices. The PGFI was slightly lower at .505. According to Kelloway (1998) these indices are more meaningfully used when comparing two competing theoretical models and are not very useful indicators in this CFA analysis.

The values for this model's Akaike information criterion (AIC = 182.546) suggest that the fitted measurement model provided a more parsimonious fit than the independent model (2661.792) but not the saturated model (156.000) since smaller values on these indices indicate a more parsimonious model (Kelloway, 1998). This indicated that the measurement model may lack influential paths. Values for the consistent Akaike information criterion (414.215) imply that the fitted measurement model provides a more parsimonious fit than both the independent model (2735.266) and the saturated model (633.584). This provided further support for the fitted model.

The comparative fit indices returned the following values: normed fit index (NFI = .960), the non-normed fit index (NNFI = .958), the comparative fit index (CFI = .975), the incremental fit index (IFI = .975) and the relative fit index (RFI = .934). In the current results, all of these indices exceed the .90 level, which would be indicative of satisfactory comparative fit relative to the independence model.

The estimated CN value (748.669) fell above the recommended threshold value of 200. This threshold is regarded as indicative of the model providing an adequate representation of the data (Diamantopoulos & Siguaw, 2000).

The RMR (.0287) and standardized RMR (.0277) indicated reasonable fit as values less than .05 on the latter index suggest the model fits the data well (Kelloway, 1998).

For the fit of this model, both the GFI (.984) and AGFI (.969) are above the acceptable cut-off level of .90, reflecting that the model does reproduce the same covariance matrix (Diamantopoulos &Siguaw, 2000).

In conclusion, when the abovementioned model fit statistics were considered holistically they seem to suggest a well-fitting model. In addition, when taking the fitted model, independence model and saturated model into account, evidence was provided in support of the fitted model, however, the model fit may possibly benefit from the inclusion of a few additional paths.

4.7.3.1 Interpreting the freed measurement model parameter estimates

The unstandardized factor loading matrix is shown in Table 36. These unstandardized lambda-X coefficients represent the factor loadings of the items on their designated factors as in the previous operational hypothesis.

From the output in Table 36 one can see that all the path coefficients are significant looking at z≥|1.96| except for Q188_SD and RQ27_SD for the factor of *self-deception enhancement* and Q102_SD and Q109_SD for the factor of *impression management*. All the negatively worded items loaded statistically significantly (p<.05) on the negative SD factor and all the positively keyed items loaded statistically significantly (p<.05) on the positive SD factor. A total of 3 of the 5 *self-deception enhancement* items loaded statistically significantly (p<.05) on the *self-deception enhancement* factor while 5 of the 7 *impression management* items loaded on the *impression management* factor. Items RQ168_SD, Q035_SD and RQ223_SD loaded statistically significantly (p<.05) on the *self-deception enhancement* factor as hypothesised under operational hypothesis 3, whereas item Q188_SD and RQ27_SD loaded statistically insignificantly on the *self-deception* factor. $H_{038,3}$, $H_{0566,3}$ and $H_{058,3,}$ could be rejected whereas $H_{040,3}$ and $H_{054,3}$could not be rejected.Q90_SD, RQ53_SD, RQ129_SD, RQ131_SD and RQ252_SD loaded statistically significantly ($p<.05$) on the *impression management* factor as hypothesised under operational hypothesis 3. Item Q102_SD and Q109_SD, were the only items that should reflect the *impression management* factor according to operational hypothesis 2 that loaded statistically insignificantly on the second factor. Therefore, $H_{042,3}$, $H_{044,3}$, $H_{050,3}$, $H_{052,3}$, and $H_{060,3}$could be rejected whereas $H_{046,3}$ and $H_{048,3}$ could not be rejected. RQ27_SD, RQ53_SD, RQ129_SD, RQ131_SD, RQ168_SD and RQ252_SD loaded statistically significantly ($p<.05$) on the *negatively worded* factor as hypothesised under operational hypothesis 3. Q35_SD, Q90_SD, Q102_SD, Q109_SD, Q188_SD and Q223_SD loaded statistically significantly ($p<.05$) on the *positively worded* factor as hypothesised under operational hypothesis 3. For both the *negatively* and *positively* worded factors there were no items that loaded insignificantly therefore the remaining hypotheses could be rejected, namely; $H_{039,3,}$ $H_{041,3}$, $H_{043,3,}$ $H_{045,3}$, $H_{047,3}$, $H_{049,3,}$ $H_{051,3}$, $H_{053,3}$, $H_{055,3,}$ $H_{057,3}$, $H_{059,3}$ and $H_{061,3}$.

Table 36*:*

Four-factor SAPI SD measurement model unstandardised lambda-X matrix (operational hypothesis 3)

	SELFDECP	IM	NEG	POS
RQ27_SD	-0.054	- -	0.323*	- -
	(0.040)		(0.028)	
	-1.369		11.638	
Q035_SD	0.512*	- -	- -	0.340*
	(0.071)			(0.078)
	7.162			4.368
RQ53_SD	- -	-0.094*	0.235*	- -
		(0.045)	(0.036)	
		-2.068	6.517	
Q090_SD	- -	0.428*	- -	0.257*
		(0.054)		(0.069)
		7.961		3.724
Q102_SD	- -	0.088	- -	0.388*
		(0.057)		(0.036)
		1.543		10.877
Q109_SD	- -	0.017	- -	0.438*
		(0.061)		(0.030)
		0.283		14.646
RQ129_SD	- -	0.305*	0.134*	- -
		(0.040)	(0.051)	
		7.605	2.604	
RQ131_SD	- -	0.304*	0.386*	- -
		(0.059)	(0.054)	
		5.125	7.186	
RQ168_SD	0.287*	- -	0.371*	- -

	(0.056)		(0.055)	
	5.127		6.700	
Q188_SD	-0.024	- -	- -	0.340*
	(0.039)			(0.025)
	-0.608			13.538
Q223_SD	-0.077	- -	- -	0.344*
	(0.042)			(0.033)
	-1.807			10.395
RQ252_SD	- -	0.742**	0.311	- -
		(0.063)	(0.104)	
		11.857	2.986	

* (p<.05)

In Table 37 the completely standardised solution for lambda-X is provided in which both the latent and observed variables are standardised to have a mean of zero and a standard deviation of one. The completely standardised factor loadings in Table 37 therefore represent the average change on the item, expressed in standard deviation units, associated with one standard deviation change in the latent SD dimension being measured when holding the other latent SD dimension being measured constant.

Table 37 indicates that all but one of the completely standardised factor loadings fell below the set critical factor loading value of .50, along with the findings on the statistical insignificance of four of the factor loadings on the two narrower SD factors. Support is therefore not obtained for operational hypothesis 3.2 that posited that the factor loadings of the items on their designated latent social desirability dimensions (*self-deception enhancement* or *impression management*) are statistically significant ($p<.05$) and large ($\lambda_{ij} \geq .50$).

Table 37*:*

Four-factor SAPI SD measurement model completely standardised solution lambda-X matrix (operational hypothesis 3)

	SELFDECP	IM	NEG	POS
RQ27_SD	-0.077	- -	0.458	- -
Q035_SD	0.438	- -	- -	0.291
RQ53_SD	- -	-0.104	0.261	- -
Q090_SD	- -	0.376	- -	0.226
Q102_SD	- -	0.092	- -	0.404
Q109_SD	- -	0.021	- -	0.521
RQ129_SD	- -	0.316	0.139	- -
RQ131_SD	- -	0.288	0.365	- -
RQ168_SD	0.270	- -	0.349	- -
Q188_SD	-0.033	- -	- -	0.472
Q223_SD	-0.087	- -	- -	0.390
RQ252_SD	- -	0.638	0.268	- -

In Table 38 the estimates of the freed theta-delta coefficients have been provided and these represent the sample estimates of the variance in measurement error terms. Table 38 indicates that all the items are statistically significantly ($p<.05$) plagued by measurement error. Since H_{api}: $\Theta_{\delta ijj}> 0$;p=38, 39, 40…61; i =1, 2, 3, 4; j=1, 2 ....., 12 have been formulated as directional hypotheses, H_{0pi}: $\Theta_{\delta jj}=0$;p=38, 39, 40…61; i =1, 2, 3, 4; j=1, 2, ..., 12 were tested via one-sided statistical tests. A critical z-value of 1.6449 was therefore used as the critical value to evaluate the statistical significance of the $\theta_{\delta ii}$ estimates shown in Table 38. This finding is welcomed since a finding of perfectly valid and reliable measures of the two SD factors and the two method factors would have been a too improbable outcome.

Table 38*:*

Four-factor SAPI SD measurement model unstandardised theta-delta matrix (operational hypothesis 3)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.390*	0.989*	0.747*	1.046*	0.766*	0.515*
(0.022)	(0.065)	(0.034)	(0.047)	(0.037)	(0.029)
17.652	15.201	22.202	22.318	20.418	17.759
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.818*	0.872*	0.911*	0.402*	0.655*	0.704*
(0.035)	(0.039)	(0.043)	(0.021)	(0.033)	(0.058)
23.614	22.085	21.167	18.846	19.717	12.215

* ($p<.05$)

The completely standardised $\theta_{\delta ii}$ estimates are shown in Table 39. Table 39 indicates that 11 of the 12 SAPI SD items provide measures of the SD factors they were earmarked to reflect in terms of substantive research hypothesis 3, that are generally highly contaminated by systematic and/or random measurement error. RQ252_SD is the only exception. In all the items more than 50% of the variance in the item responses were due to systematic and/or random measurement error.

Table 39*:*

Four-factor SAPI SD measurement model completely standardised solution theta-delta matrix (operational hypothesis 3)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.784	0.724	0.921	0.807	0.829	0.728
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.881	0.784	0.805	0.776	0.841	0.521

In Table 40 the completely standardised phi matrix is provided. This matrix shows the correlations between the latent variables and is therefore the observed correlations between the observed scores on the *impression management* and *self-deception enhancement* dimensions corrected for the attenuating effect of measurement error. Table 40 indicates that *impression management* correlates statistically significantly (p<.05) with *self-deception enhancement* H_{0pi}: ϕ_{jk}=0; p=138, 139; i =3; j=1, 3; k=2, 4 could nonetheless not be rejected despite the fact that the ϕ_{12} and ϕ_{34} estimates were statistically significant (p<.05). Table 40 reflects a warning. This warning is acknowledged as indicating that the correlations have inadmissible values in that the correlation between *self-deception enhancement* and *impression management* is greater than unity. The correlation between the negative method factor and the positive method factors is also greater than 1. These inadmissible estimates seriously erode confidence in the model.

Table 40*:*

Four factor SAPI SD measurement model completely standardised phi matrix (operational hypothesis 3)

	SELFDECP	IM	NEG	POS
SELFDECP	1.000			
IM	1.350*	1.000		
	(0.137)			
	9.843			
NEG	- -	- -	1.000	
POS	- -	- -	1.089*	1.000
			(0.056)	
			19.351	

* (p<.05)

In Table 41 the squared multiple correlations for X-variables are provided and these correlations indicate the proportion of variance in the observed variable explained by the latent variable/variables linked to it in the measurement model. High correlation values are preferred.

Table 41 echoes the results obtained in Table 39. The latent variables that the items were designed to reflect all explain less than 50% of the variance in each item. RQ252_SD is the only exception where the proportion of variance explained in the item approaches 50%.

Table 41*:*

Four-factor SAPI SD measurement model squared multiple correlations for X-variables (operational hypothesis 3)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.216	0.276	0.079	0.193	0.171	0.272
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.119	0.216	0.195	0.224	0.159	0.479

4.7.4 Testing Operational Research Hypothesis 4

Operational research hypothesis 4 represents substantive research hypothesis 2 but also assumes that method bias plays a role. Substantive research hypothesis 2 postulates that the SD scale of the SAPI provides a construct valid measure of social desirability conceptualised as a three dimensional construct comprising the two positively correlated latent dimensions of *self-deception enhancement* and impression management, but also a broad, general social desirability latent dimension, uncorrelated with the two narrower social desirability dimensions, representing the common variance shared by all twelve of the SAPI's SD items. Rather than extending the measurement model assumed under operational research hypothesis 2 by assuming the addition of two method factors, operational research hypothesis 4 rather extended hypothesis 2 by hypothesising that the differentially keyed items will result in positive and negative factor loadings on the broad general social desirability factor. A three-factor measurement model was therefore fitted to the data where the SD-scale items loaded on the two narrow factors social desirability factors as they were categorised in Table 3 and where the six positively worded items were hypothesised to load statistically significantly ($p<.05$) and positively on the general social desirability factor and the six negatively worded items were hypothesised to load statistically significantly ($p<.05$) and negatively on the general social desirability factor. Operational hypothesis 4 therefore differs from operational hypothesis 2 in that under operational hypothesis 4 the alternative hypotheses for Λ^X were formulated as directional hypotheses whereas under operational hypotheses they were formulated as non-directional hypotheses. Operational research hypothesis 4 assumes that the measurement model is fitted to data in which the negatively keyed items were not reflected.

The measurement model converged in 76 iterations. The completely standardised solution is shown in Figure 12. The fit statistics are shown in Table 42.

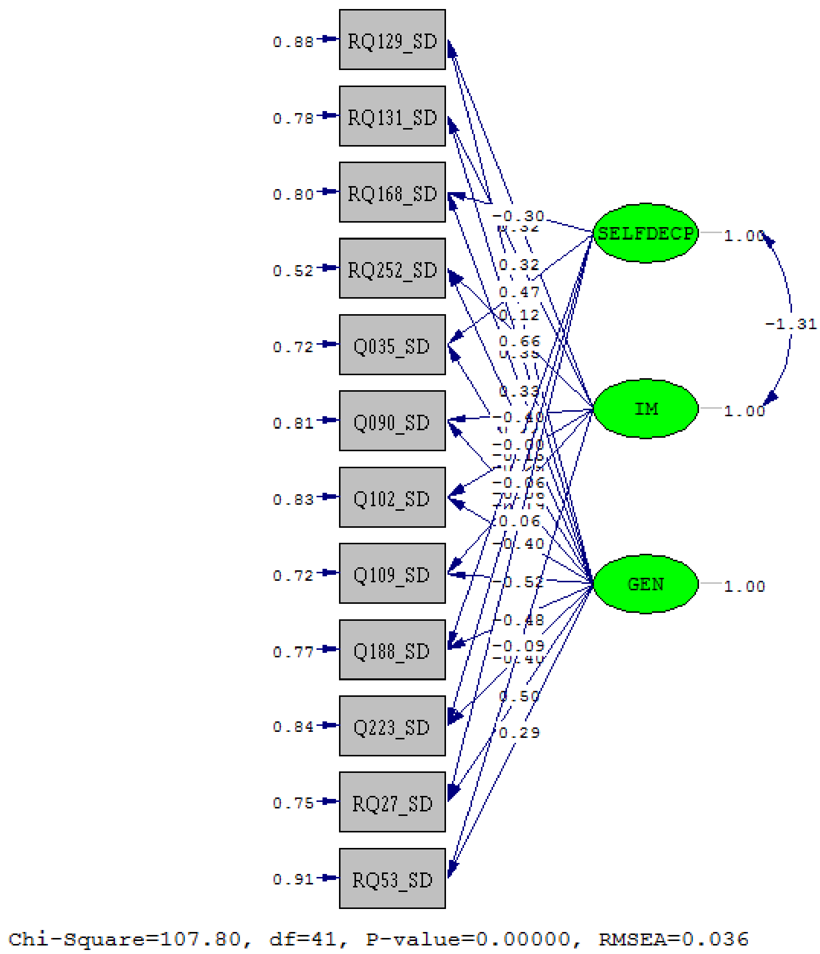

Figure 12: Representation of the fitted three-factor social desirability measurement model hypothesised by operational research hypothesis 4 with method factors (completely standardised solution)

Table 42:

Goodness of Fit Statistics for the 4-factor Social Desirability Measurement Model (operational hypothesis4)

Goodness of Fit Statistics
Degrees of Freedom = 41
Minimum Fit Function Chi-Square = 120.158 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 123.862 (P = 0.00)
Satorra-Bentler Scaled Chi-Square = 107.799 (P = 0.000)
Chi-Square Corrected for Non-Normality = 99.154 (P = 0.000)
Estimated Non-centrality Parameter (NCP) = 66.799
90 Percent Confidence Interval for NCP = (39.821 ; 101.448)
Minimum Fit Function Value = 0.0970
Population Discrepancy Function Value (F0) = 0.0539
90 Percent Confidence Interval for F0 = (0.0321 ; 0.0819)
Root Mean Square Error of Approximation (RMSEA) = 0.0363
90 Percent Confidence Interval for RMSEA = (0.0280 ; 0.0447)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.997
Expected Cross-Validation Index (ECVI) = 0.147
90 Percent Confidence Interval for ECVI = (0.125 ; 0.175)
ECVI for Saturated Model = 0.126
ECVI for Independence Model = 2.148
Chi-Square for Independence Model with 66 Degrees of Freedom = 2637.792
Independence AIC = 2661.792
Model AIC = 181.799
Saturated AIC = 156.000
Independence CAIC = 2735.266
Model CAIC = 408.346
Saturated CAIC = 633.584
Normed Fit Index (NFI) = 0.959
Non-Normed Fit Index (NNFI) = 0.958
Parsimony Normed Fit Index (PNFI) = 0.596
Comparative Fit Index (CFI) = 0.974
Incremental Fit Index (IFI) = 0.974
Relative Fit Index (RFI) = 0.934
Critical N (CN) = 747.516
Root Mean Square Residual (RMR) = 0.0295
Standardized RMR = 0.0278
Goodness of Fit Index (GFI) = 0.984
Adjusted Goodness of Fit Index (AGFI) = 0.969
Parsimony Goodness of Fit Index (PGFI) = 0.517

The fit statistics associated with the measurement model fitted under operational research hypothesis 4 are identical to those obtained for the measurement model fitted under operational research hypothesis 2. The conclusion on the fit of the measurement model fitted under operational research hypothesis 2, derived from the discussion of the array of fit statistics presented in paragraph 4.6.2, therefore also applies here. The three-factor measurement model fits the data well.

4.7.4.1 Interpreting the freed measurement model parameter estimates

The unstandardized factor loading matrix is shown in Table 43. These unstandardized lambda-X coefficients represent the factor loadings of the items on their designated factors as in the previous operational hypothesis.

Table 43*:*

Three-factor SAPI SD measurement model unstandardised lambda-X matrix (operational hypothesis 4)

	SELFDECP	IM	GEN
RQ27_SD	0.045	- -	0.352*
	(0.046)		(0.028)
	0.965		12.494
Q035_SD	0.544*	- -	-0.297*
	(0.070)		(0.092)
	7.794		-3.232
RQ53_SD	- -	-0.083	0.260*
		(0.050)	(0.038)
		-1.659	6.797
Q090_SD	- -	-0.450*	-0.217*
		(0.053)	(0.081)
		-8.553	-2.686
Q102_SD	- -	-0.122*	-0.381*
		(0.062)	(0.039)
		-1.969	-9.794
Q109_SD	- -	-0.054	-0.440*
		(0.069)	(0.032)
		-0.777	-13.798
RQ129_SD	- -	0.312*	0.118*
		(0.040)	(0.060)
		7.769	1.970
RQ131_SD	- -	0.332*	0.368*
		(0.063)	(0.063)
		5.256	5.868
RQ168_SD	-0.315*	- -	0.350*
	(0.060)		(0.063)
	-5.286		5.576
Q188_SD	0.000	- -	-0.347*
	(0.046)		(0.025)
	-0.005		-13.765
Q223_SD	-0.050	- -	-0.351*
	(0.048)		(0.033)
	-1.050		-10.488
RQ252_SD	- -	0.763*	0.258*
		(0.062)	(0.127)
		12.268	2.040

* (p<.05)

The following 12null hypotheses on the slope of the regression of the positively keyed item j on latent social desirability dimension k were tested for operational research hypotheses 4:

H_{0pi}: λ_{jk}=0;p=62, 63, …, 73; i=4; j=1, 2, 4, 6, 10,11; k=1, 2, 3

H_{api}: λ_{jk}>0;p=62, 63, …, 73; i=4; j=1, 2, 4, 6, 10,11; k=1, 2, 3

Since H_{api} were formulated as directional alternative hypotheses, H_{0pi} were tested via a one-sided test using a critical z-value of 1.6449. Table 43 indicates that $H_{0,65,4}$: λ_{23}=0, $H_{0,69,4}$: λ_{43}=0, $H_{0,71,4}$: λ_{53}=0, $H_{0,73,4}$: λ_{63}=0, $H_{0,81,4}$: $\lambda_{10,3}$=0 and $H_{0,83,4}$: $\lambda_{11,3}$=0 could not be rejected. All positively keyed items loaded statistically significantly (p<.05) but negatively on the general social desirability factor in contrast to the positive loadings hypothesised under operational hypothesis 4.

The positively keyed item Q35_SD loaded statistically significantly (p<.05) and positively on the *self-deception* factor as hypothesised under operational hypothesis 4. $H_{0,64,4}$: $\lambda_{21,4}$=0 could therefore be rejected. The other two positively keyed items that are meant to load on the *self-deception* factor (Q188_SD and Q233_SD) returned statistically insignificant loadings. $H_{0,80,4}$: $\lambda_{10,1.4}$=0 and $H_{0,82,4}$: $\lambda_{11,1,4}$=0 can therefore not be rejected. Although λ_{42}, (Q090_SD) and λ_{52} (Q102_SD) were statistically significant (p<.05), the sign of the factor loadings did not agree with the direction hypothesised under $H_{a68,4}$, and $H_{a70,4}$. These two hypotheses could therefore not be rejected. The positively keyed items Q109_SD loaded statistically insignificantly (p>.05) on the *impression management* factor rather than significantly as hypothesised under operational hypothesis 4.

The following 12 null hypotheses on the slope of the regression of negatively keyed item j on latent social desirability dimension k were tested for operational research hypotheses 4:

H_{0pi}: λ_{jk}=0; p=74, 75, …, 85; i=4; j=3, 5, 7, 8, 9, 12; k=1, 2, 3

H_{api}: λ_{jk}<0; p=74, 75, …, 85; i=4; j=3, 5, 7, 8, 9, 12; k=1, 2, 3

Table 43 indicates that $H_{0,63,4}$: λ_{13}=0, $H_{0,67,4}$: λ_{33}=0, $H_{0,75,4}$: λ_{73}=0, $H_{0,77,4}$: λ_{83}=0, $H_{0,79,4}$: λ_{93}=0 and $H_{0,85,4}$: $\lambda_{12,3}$=0could not be rejected. All negatively keyed items loaded statistically significantly (p<.05) but positively on the general social desirability factor in contrast to that which was hypothesised under operational hypothesis 4. The negatively keyed item RQ53_SD, RQ129_SD, RQ131_SD and RQ252_SD all loaded statistically significantly on the *impression management* factor. The positive sign associated with the $\lambda_{72}\lambda_{82}$and $\lambda_{12,2}$estimates did not allow the rejection of $H_{0,74,4}$: λ_{72}=0, $H_{0,76,4}$: λ_{82}=0 and $H_{0,84,4}$: $\lambda_{12,2}$=0,. Only $H_{0,66,4}$: λ_{32}=0 could be rejected.

Table 44 represents the completely standardised solution for lambda-X and it indicates that all the completely standardised factor loadings fell below the set critical factor loading value of .50. except for item Q109_SD for the general social desirability factor and RQ252_SD for the *impression management* factor.

Table 44*:*

Three-factor SAPI SD measurement model completely standardised solution lambda-X matrix (operational hypothesis 4)

	SELFDECP	IM	GEN
RQ27_SD	0.064	- -	0.499
Q035_SD	0.466	- -	-0.254
RQ53_SD	- -	-0.092	0.289
Q090_SD	- -	-0.396	-0.191
Q102_SD	- -	-0.127	-0.396
Q109_SD	- -	-0.064	-0.523
RQ129_SD	- -	0.323	0.122
RQ131_SD	- -	0.315	0.349
RQ168_SD	-0.296	- -	0.329
Q188_SD	0.000	- -	-0.482
Q223_SD	-0.057	- -	-0.398
RQ252_SD	- -	0.657	0.222

In Table 45 the freed theta-delta coefficients have been provided and these represent the sample estimates of the variance in measurement error terms. Table 45 indicates that all the items are statistically significantly ($p<.05$) plagued by measurement error. A critical z-value of 1.6449 was used as the critical value to evaluate the statistical significance of the $\theta_{\delta ii}$ estimates shown in Table 45 since the alternative hypotheses were formulated as directional hypotheses. This finding is welcomed since a finding of perfectly valid and reliable measures of the two SD factors and the general SD factor would have been a too improbable outcome.

Table 45*:*

Three-factor SAPI SD measurement model unstandardized theta-delta matrix(operational hypothesis 4)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.371*	0.981*	0.736*	1.045*	0.764*	0.510*
(0.022)	(0.064)	(0.034)	(0.047)	(0.038)	(0.029)
16.774	15.270	21.408	22.252	20.302	17.691
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.817*	0.866*	0.908*	0.398*	0.654*	0.701*
(0.035)	(0.039)	(0.042)	(0.022)	(0.033)	(0.057)
23.631	21.989	21.399	18.440	19.566	12.294

* (p<.05)

The completely standardised $\theta_{\delta ii}$ estimates are shown in Table 46. Table 46 indicates that 11 of the 12 SAPI SD items provide measures of the SD factors they were earmarked to reflect in terms of substantive research hypothesis 4, that are generally highly contaminated by systematic and/or random measurement error. RQ252_SD is the only exception. In all the items more than 50% of the variance in the item responses are due to systematic and/or random measurement error.

Table 46*:*

Three-factor SAPI SD measurement model completely standardised solution theta-delta matrix (operational hypothesis 4)

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.747	0.718	0.908	0.807	0.827	0.722
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.880	0.779	0.804	0.768	0.839	0.519

In Table 47 the completely standardised phi matrix is provided. Table 47 indicates that *impression management* does correlates statistically significantly (p<.05) with *self-deception enhancement*. H_{0pi}: $\phi_{jk}=0$; p=137; i =4; j=1; k=2 could nonetheless not be rejected, despite the fact that the ϕ_{21} estimate was statistically significant (p<.05), because the sign of the correlation was negative whilst a positive correlation was hypothesised under H_{api}: $\phi_{jk}> 0$; p=137; i =4; j=1; k=2 and because the ϕ_{21} estimate was inadmissible.

Table 47*:*

SAPI SD measurement model completely standardised phi matrix

	SELFDECP	IM	GEN
SELFDECP	1.00		1.000
IM	-1.311 (0.120) -10.900	1.000	1.000
GEN		- -	1.000

In Table 48 the squared multiple correlations for X-variables are provided and these correlations indicate the proportion of variance in the observed variable explained by the latent variable/ variables linked to it in the measurement model. The latent variables that the items were designed to reflect generally explain less than 50% of the variance in each item. RQ252_SD is the only exception.

Table 48*:*

SAPI SD measurement model squared multiple correlations for X-variables

RQ27_SD	Q035_SD	RQ53_SD	Q090_SD	Q102_SD	Q109_SD
0.253	0.282	0.092	0.193	0.173	0.278
RQ129_SD	RQ131_SD	RQ168_SD	Q188_SD	Q223_SD	RQ252_SD
0.120	0.221	0.196	0.232	0.161	0.481

4.8. SUMMARY

These analyses were designed to provide insight into the functioning of the Social Desirability scale (and its factors) within the SAPI questionnaire. The development of this scale made use of items from two different scales that have been noted in the literature to be the best scales measuring social desirability in personality tests used in Western societies. The SAPI collaborators applied the SAPI principles to identify the best items for a Social Desirability scale for the SAPI and these 12 items survived the analyses.

The models that were fitted in this analysis were derived from the original scales. It needs to be noted that the original scales were developed in a Westernised environment and then applied to a South African population. It is therefore, not altogether surprising that the original model did not produce good fit when applied to this population. It in addition needs to be noted that the developers of the SAPI Social Desirability scale did not develop the scale after they have committed to a specific constitutive definition of social desirability that clearly conceptualised the connotative meaning of the construct. The current study had to make specific inferences about the connotative meaning of the construct measured by the scale from the manner in which the scales from which items were harvested for the SAPI SD scale conceptualised social desirability. The current study also had to make specific inferences about the manner in which the scale measures the social desirability construct thus conceptualised.

In making inferences about the connotative meaning of the construct, the current study firstly considered the possibility that the social desirability construct comprised two positively correlated narrowly-defined social desirability factors, namely *self-deception enhancement* and *impression management*. The current study secondly, however, considered the possibility that the social desirability construct comprised two positively correlated narrowly-defined social desirability factors, namely *self-deception enhancement* and *impression management* as well as a broader general social desirability factor that is uncorrelated with the two narrower social desirability factors.

In making inferences about the manner in which the scale measures the social desirability construct conceptualised in these two possible ways, the current study considered two possible method factor that reflect the manner in which the items of the SAPI SD scale were positively and negatively keyed positively.

The two inferences on the connotative meaning of the construct, combined with the possibility that method factors had to be assumed to successfully explain the observed inter-item variance-covariance matrix, lead to the development of four operational research hypotheses[19] on the nature of the measurement model underlying the SAPI SD scale.

The current study explored and tested these four operational hypotheses to best understand the structure of social desirability construct as measured by the SAPI SD scale in the South African population. These four different possibilities were tested by fitting four different measurement

[19] In developing these hypotheses the possibility of a measurement model incorporating both a general factor and two method factors in addition to the two narrower SD factors was not considered.

models with their associated sets of statistical hypotheses on model fit and parameter values. The first operational hypothesis hypothesised a two-factor measurement model based on the manner in which the scales from which SAPI SD items were harvested conceptualised social desirability. This model was cleanly discounted. The other three measurement models all showed close fit but insignificant factor loadings (p>.05) and inappropriate signs prevented all null hypotheses associated with the factor loadings to be rejected. Inadmissible ϕij estimates were also obtained for these three measurement models. No unequivocal verdict was therefore possible on either operational hypothesis 2, 3 or 4.. The findings on the four models are summarised in Table 49.

Table 49*:*

Summary of the findings on the four operational hypotheses on the factor structure underlying the SAPI SD scale

Operational hypothesis number	Sample RMSEA value	p-value	Comments on Λ	Comments on Θ_δ	Comments on Φ
1	.0819	.0000	NA	NA	NA
2	.0363	.997	RQ27_SD, Q188_SD and Q233_SD load insignificantly (p>.05) on *self-deception*. RQ53_SD and Q109_SD load insignificantly (p>.05) on *impression management*. All items load positive and significantly (p<.05) on the general SD factor.	All $\theta_{\delta jj}$ were statistically significant (p<.05) but large ($\theta_{\delta jj}$>.50)	ϕ_{12} negative inadmissible
3	.0364	.996	RQ27_SD, Q188_SD and Q233_SD load insignificantly (p>.05) on *self-deception*. Q102_SD, Q109_SD load insignificantly (p>.05) on *impression management*. All negatively keyed items load positive and significantly (p<.05) on the negatively keyed factor. All positively keyed items load positive and significantly (p<.05) on the positively keyed factor.	All $\theta_{\delta jj}$ were statistically significant (p<.05) but large ($\theta_{\delta jj}$>.50)	ϕ_{12} positive inadmissible ϕ_{34} positive inadmissible
4	.0363	.997	RQ27_SD, Q188_SD and Q233_SD load insignificantly (p>.05) on *self-deception*. Q109_SD load insignificantly (p>.05) on *impression management*. Of the positively keyed items that loaded significantly on their designated narrow SD factor only Q035_SD had an appropriate positive sign. Of all the negatively keyed items that loaded significantly on their designated SD factor only RQ168_SD and RQ53_SD, had appropriate negative signs. All the negatively keyed items load positive and significantly (p<.05) on the general SD factor. All positively keyed items load negative and significantly (p<.05) on the general SD factor.	All $\theta_{\delta jj}$ were statistically significant (p<.05) but large ($\theta_{\delta jj}$>.50)	ϕ_{12} negative inadmissible

The second and the fourth model seems to provide the best account of the observed covariance matrix in that they fitted the data marginally better than model 3 as judged by the RMSEA fit statistic. Model 2 conceptualised the social desirability construct underpinning the SAPI SD scale in terms of the two narrow factors of social desirability; *self-deception enhancement* and *impression management* as well as a general factor of social desirability (orthogonal to the two narrow SD factors). Model 3 conceptualised the social desirability construct underpinning the SAPI SD scale in terms of the two narrow factors of social desirability; *self-deception enhancement* and *impression management* as well as two method factors (orthogonal to the two narrow SD factors). In model 2 the items loaded as hypothesised on the general SD factor and in model 3 the items loaded as hypothesised on the two method factors. In both models though, three *self-deception enhancement* items (the same three items) and two items of *impression management* (one common problematic item) loaded statistically insignificantly ($p>.05$) on their designated narrow SD factors. The fourth model was fitted to the data before the negatively keyed items were reflected. The fourth model set more stringent hypotheses on the measurement model parameters freed in model 2. The negatively keyed items were hypothesised to load negatively on their designated narrow SD factor and on the general SD factor and the positively keyed items were hypothesised to load positively on their designated narrow SD factor and on the general SD factor. The signs of the statistically significant items loading, with two exceptions, differed from the sign that was hypothesised under the alternative hypothesis. In all the measurement models that demonstrated close fit the measurement error variances were large and the proportion of variance that the latent SD dimensions that the items were tasked to reflect explained in the items were small The items of the SAPI SD scale therefore provide very noisy measures of the latent SD dimensions in the various models that showed close fit.

CHAPTER FIVE

DISCUSSION AND LIMITATIONS

5.1. INTRODUCTION

Psychological assessments are regularly used within the South African context. Personality assessments belong to the family of psychological assessments and are frequently used in the workplace. The impact of the country's historical background on the information derived from these assessments is a contested issue. South African legislation protects the population from any discriminatory practices in relation to assessment and this includes ensuring that valid and reliable assessments are used to measure personality traits and that the inferences derived from such personality measures are valid and fair. This is especially important if selection decisions are made based on the results of personality assessments. This places a further demand on the culture appropriateness of personality assessments.

There are several strategies that can be used to develop personality measures to be used in a country like South Africa. Culture influences the decisions an individual makes in specific situations. Culture includes aspects such as the individual's language, belief, values, religion and social organisation. Cross-cultural researchers aim to investigate traits that are universally common to all cultures (etic perspective), traits that are culturally specific (emic) and then identify a combination of the two perspectives (indigenous). This combination approach ensures that even when an assessment tests a relatively universal construct such as personality cultural specific versions are developed based on the outcomes of equivalence studies (Cheung, Van de Vijver, & Leong., 2011). The same applies to the Social Desirability scales that are used in personality assessment.

The SAPI project's main objective is the development of an indigenous personality assessment that addresses the concerns raised by South African legislation and the historical background of the country. The SAPI is a self-report instrument. The use of personality assessments has increased as these assessments have been shown to predict job performance and future behaviour (Barrick & Mount, 1996; Hough et al., 1990; Ones & Viswesvaran, 2001). Personality assessments are vulnerable to respondents making use of self-enhancement strategies like socially desirable responding. The construct of social desirability is typically assessed as part of the personality assessment in an attempt to determine to what extent the personality description that has been obtained has been distorted by the desire to appear socially acceptable. A Social Desirability scale of 12 items had been developed for the SAPI.

Normally, when an instrument like the SAPI SD scale is developed, it is developed to measure a specific construct to which a specific constitutive definition is attached and for which specific items are then designed to reflect individuals' standing on the latent dimensions that comprise the construct in terms of the constitutive definition. The design intention is reflected in the scoring key of the instrument and the combination of this scoring key and the constitutive definition of the construct (social desirability in this case) then implies a very specific measurement model. In the case of the SAPI SD scale, however, the developers seemingly did not first clearly explicate the connotative meaning of the social desirability construct and neither did a specific design intention guide the selection or writing of items. This seriously constrained the attempt to determine whether the SAPI SD scale provides a construct valid measure of the social desirability construct.

In an attempt to circumvent the problem the connotative meaning of the social desirability construct was inferred from the constitutive definitions attached to the construct by the two instruments from which the SAPI SD scale items were harvested. It was thus inferred that the SAPI SD scale conceptualised social desirability in terms of the two narrow SD factors of *self-deception enhancement* and *impression management*. In addition it was hypothesised that the variance in each of the 12 SAPI SD scale items not only reflected an individual's standing on a specific narrow SD factor but also on a broader general social desirability factor that is unrelated to the narrower SD factors (i.e. the general factor explains unique variance in each item not shared by a narrow SD factor). Moreover the possibility was considered that two method factors could possibly explain variance in the 12 SAPI items due to the fact that 6 items were positively keyed and six items were negatively keyed. This resulted in 4 operational hypotheses on the nature of the measurement model that describes the nature of the relationships between the latent variables measured by the SAPI SD scale and the items of the scale and the nature of the correlational relationship between the latent variables.

5.2. DISCUSSION

The SAPI project aims to develop an indigenous personality measurement that consists of an inclusive approach to accommodate the dynamics of the South African multi-cultural context. This indigenous personality measurement includes a scale that measures social desirability and the construct validity of this scale is of interest to this study

The response bias of social desirability is measured by 12 items in the SAPI. In this study the items were exposed to a series of analyses with the end goal of determining whether inferences about individual's standing on the Social Desirability construct may permissibly be derived from the measure obtained on the SAPI SD scale. The study made use of both exploratory and confirmatory factor analysis in its investigation of the Social Desirability scale of the SAPI. Missing values were addressed using imputation by matching which was successful in that this procedure resulted in a dataset of n=1263 respondents with no missing data. Preliminary analyses of the data showed that some items had values outside of the 1-5 range. Further analysis aimed at identifying any outliers indicated that there were a number of extreme outliers and these cases were also deleted from the data set yielding a final sample size of 1240 respondents.

The social desirability items used in the SAPI SD scale were harvested from two already established scales; the BIDR and the MCSD scales. This harvesting implied a specific conceptualisation of social desirability. The manner in which the MCSD scale defined social desirability is similar to the manner in which the BIDR conceptualised *impression management*. The MCSD items and the BIDR items that had been earmarked to measure *impression management* therefore all measure the intentional distortion of responses to create a favourable impression. The remaining items of the BIDR had been developed to measure *self-deception enhancement.* This resulted in the assumption, based on the scales from which the SAPI SD scale items were harvested from, that 7 items (RQ53_SD, Q90_SD, Q102_SD, Q109_SD, RQ129_SD, RQ131_SD, and RQ252_SD) measure *impression management* and the remaining 5 items (RQ27_SD, Q35_SD, RQ168_SD, Q188_SD and Q223_SD) measured *self-deception enhancement* as defined by the BIDR scale.

As part of the exploration phase of the study, item analysis was conducted on each of the two SAPI SD subscales separately. The item analysis yielded unsatisfactory results for the reliability coefficients for each subscale; *impression management* and *self-deception enhancement*. The results for the inter-item statistics for both these subscales indicated that this situation would not improve with the removal of an item except for that of RQ53_SD for the *impression management* subscale. However, the improvement was not substantial enough to warrant the deletion of this item. More importantly the low internal consistency pointed to the need to question operational hypothesis 1 that each of the items of the SAPI SD scale only measure one of two narrow SD factors (*impression management* and *self-deception enhancement*). As part of the exploration phase of the study dimensionality analysis was also conducted separately on the two subscales *impression management* and *self-deception enhancement* to test the assumption that each subscale measures a unidimensional narrow SD factor. For the *self-deception enhancement* scale there was evidence of two factors. The same was found for the *impression management* scale. These results failed to confirm the unidimensionality assumption for both these subscales and provided evidence to testify against operational hypothesis 1.

A series of confirmatory factor analyses were subsequently conducted. These analyses provided further evidence that the two-factor measurement model hypothesised under operational hypothesis 1 did not offer a plausible description of the observed covariance matrix as it did not reproduce this to a sufficient degree of accuracy. In turn this increased the confidence that the researcher had in hypothesis 2, 3 and 4.

Operational hypothesis 2 conceptualised a model that includes the same positively correlated latent dimensions measured in hypothesis 1 with the inclusion of a broad and general SD factor. Operational hypothesis 2 proposed a bi-factor model. In the bi-factor model both the general factor (broad) and the specific group factors/subscales (narrow) have direct influences on the observed indicators and the specific group factors do not mediate the influence of the broad, general factor (Canivez, in press).

This three-factor measurement model was fitted to the data. This measurement model showed good fit in the sample and the position that the model fits closely in the parameter was found to be permissible. The three-factor measurement model therefore did offer a plausible explanation for the observed covariance matrix. This model may however benefit from the possible inclusion of additional paths. Three of the items for *self-deception enhancement* obtained factor loadings that were statistically insignificant ($p>.05$) and two of the items for *impression management* yielded the same results. Interestingly enough all the items loaded significantly on the generalised social desirability factor. Subsequent analyses of **Φ** revealed a negative and inadmissible correlation between the two narrow SD factors that further eroded confidence in this model.

Under operational hypothesis 3 a measurement model was assumed that included two method factors in addition to the two narrow SD factors so as to formally acknowledge that 6 of the SAPI SD scale items were negatively worded and the other 6 were positively worded. This resulted in a four-factor measurement model being fitted to the data. This measurement model also displayed close fit. The model fit statistics indicated that the model was a well-fitting model that would benefit from additional paths. Two of the items for *self-deception enhancement* did not yield statistically significant factor loadings which was one less than hypothesis 2. Two of the items for *impression management* yielded the same results, however, the two items for hypothesis 3 were different from the ones that were not statistically significant for hypothesis 2. All negatively worded items loaded on the negative SD factor and the same was found for the positively worded items and their loading on the positive SD factor. Subsequent analyses of **Φ** revealed a negative and inadmissible correlation between the two narrow SD factors and a positive but inadmissible correlation between the two method factors that further eroded confidence in this model. It was mentioned in the literature review that within the SAPI project studies have also been conducted on the nature of social desirability. These studies resulted in the identification of two factors, which were termed positive and negative impression management (Valchev et al., 2013).

The fourth and final operational hypothesis was an extension of the model proposed in operational hypothesis 2. The only difference was that that the positive items were hypothesised to load positively on the two narrow SD factors and the generalised SD factor and the (unreflected) negative items were hypothesised to load negatively on the two narrow SD factors and the generalised SD factor. This model did fit the data well with close fit. Three of the five *self-deception enhancement* items did not yield statistically significant results and one of the seven *impression management* items yielded the same results. All the negatively worded items loaded positively and significant on the general SD factor and the positively worded items loaded negatively and significantly on the general SD factor. This fourth model and the second model provided the best fit statistics with the RMSEA = .0363 and the p=.997 which is very close to unity. Confidence in these models were, however, eroded by the inadmissible ϕ_{ij} estimates.

The hope was that the empirical investigation of operational hypotheses 1 to 4 would produce unambiguous evidence that testifies in favour of one of the four positions. This unfortunately was not the case. Operational hypothesis 1 was unambiguously discredited. Operational hypotheses 2 to 4 were all to some degree supported but not unambiguously supported. In the case of all three operational hypotheses several statistical null hypotheses formulated with regards to Λ and Φ could not be rejected. There are therefore three competing possibilities that are more or less on par in terms of their ability to explain the observed covariance matrix. Operational hypotheses 2 to 4 are also more or less on par in terms of the extent to which support was obtained for the hypotheses that were developed for the freed measurement model parameters. However, each of these have, to some degree, failed to fully corroborate the operational model. Therefore, this study cannot commit to one model. This study has shown that the SD items do not only measure two narrow factors as they are not measuring the substantive narrow factors clearly. There is evidence of the general factor. It is acknowledged that there is a general factor of SD with the presence of a bi-factor modelling structure. In bi-factor analysis the first factor is called a general factor and the remaining factors are called group factors. Each items measures something specific at the same time it also measure something general.

5.3. LIMITATIONS

All studies have limitations and the limitations of this study are discussed in this section.

Firstly, the SAPI project did not commit to an upfront constitutive definition of the connotative meaning of the SD construct. Construct validity refers to the question whether it is permissible to derive inferences on the construct as constitutively defined from the observed scores obtained on the construct. In the absence of a clear constitutive definition and in the absence of a clear design intention underlying the instrument given the connotative meaning attached to the construct it becomes very difficult to examine the construct validity of the construct inferences derived from observed scores.

Secondly, the items were harvested from two western SD scales and then used in a South African personality inventory. The manner in which the other items for the SAPI were developed followed a totally different approach. A similar approach should have been followed with the SD items. The SAPI should have made sure that the SD items are appropriate South African behavioural expressions of the latent dimension that comprise the SD construct as it is constitutively defined. Besides not clarifying the connotative meaning of the SD construct the SAPI collaborators did not empirically generate critical behavioural incidents that reflect testees standing on the latent dimensions comprising social desirability.

Thirdly, the current study did not examine all conceivable models that could potentially underpin the scale. Other possible measurement models that could have been considered was a model that makes provision for the two narrow SD factors, two method factors and a general factor, a model that makes provision for a broad, general SD factor and two method factors or a model only making provision for two method factors.

Fourthly, the researcher was not in the position to alter the instrument or the underlying measurement model in any way and merely reported on the findings.

Finally, the administration of the personality assessment was pencil-and paper-based rather than electronic. Strictly speaking therefore the findings of the current research study cannot be generalised to SAPI SD scale data collected electronically.

5.4. RECOMMENDATIONS

The findings of the current study really permit only one practical recommendation, namely that the SAPI SD scale scores should as yet not be used to influence practical decisions based on SAPI personality profiles.

The following practical recommendations are made based on the findings and the limitations of the current study:

- The SAPI collaborators should, based on a theoretical stance on what constitutes social desirable responding in personality assessment, what determines it and what the consequences are, develop a formal constitutive definition of the connotative meaning of the social desirability construct.
- Paulhus and John (1998) found evidence for a four-factor model of social desirability and referred to social desirability in agency contexts as egoistic response tendencies (ERT) and social desirability in communion contexts as moralistic response tendencies (MRT). ERT consist of two factors, namely *self-deception enhancement* and agency management, while MRT also consists of two factors, namely self-deception denial and communion management. It is recommended that the SAPI collaborators should explore the option of using the hierarchical four-factor model proposed by Paulhus and John (1998) as the theoretical basis of the conceptualisation of the SAPI SD construct.
- Once the connotative meaning had been explicated the SAPI collaborators should generate items for measuring the various latent dimensions of the SD construct via the critical incident technique as opposed to harvesting the items from other SD scales.
- Once a new experimental SAPI SD scale has been, developed data should be collected on the new scale embedded in the current SAPI.
- The construct validity of the new SAPI SD scale should be evaluated utilising the procedure followed in the current study to empirically evaluate operational hypothesis 1. This analysis should allow the SD scale to be refined.
- The validity of the SD scales should be empirically evaluated experimentally by randomly assigning testees to high faking and low faking conditions on each of the dimensions of the SD construct and then determining whether the experimental manipulation [a] influenced the SD scale rating on the corresponding subscales, and [b] whether it impacted on any personality dimension scores[20].
- It is also recommended that once the construct validity of the new SAPI SD scale has been demonstrated on either a homogeneous or a representative sample from the SAPI target population the measurement invariance and equivalence of the new SAPI SD scale

[20] A similar experimental procedure could also be considered as an alternative approach to the identification of items for the SD subscales.

multi-group measurement model should be established across racial-ethnic, language and gender groups in South Africa.

- The SAPI should also be administered electronically to allow the investigation of construct and item bias across the two modes of administration via multi-group measurement invariance and equivalence analysis.
- Additional measurement models should be explored with regards to the current SAPI SD scale. Perhaps a model with two method factors, 2 narrow factors and one general factor would present a more plausible representation of the operational model.

5.5. CONCLUSION

The inclusion of a social desirability scale within personality measures continues to be an important concern for researchers and practitioners. In the literature of personality assessments, when looking specifically at social desirability, there has been a debate on its nature. This construct has unleashed much debate and research resulting in mixed, and at times contradictory, findings. The nature of these findings appears to be dependent on the operational or constitutive definition of social desirability and the choice of factor modelling. The term social desirability is often used to tap into both impression management and self-deception. The current study added to the body of knowledge and made a number of conclusions by examining the factor structure of the SAPI SD scale within the South African context.

Based on the results of the current study there is no definitive conclusion on which measurement model provides the more appropriate representation of the factor structure of the SD scale in the SAPI. The current study suggests that is probably is most prudent to abandon further attempts to describe the psychological mechanism that produced the observed inter-item covariance matrix obtained for the current SAPI SD scale. The current study suggests that it probably is more prudent to rather focus on the clear conceptualisation of the SD construct, the development of a new SD scale for the SAPI and the psychometric evaluation of the reliability, construct validity and measurement bias of the new SD scale.

REFERENCES

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35,* 13-28.

Babbie, E., & Mouton, J. (2001). *The practice of social research*. Cape Town: Oxford University Press.

Barger, S.D. (2002). The Marlowe-Crowne Affairs: Short forms, psychometric structure and social desirability. *Journal of Personality Assessment*, 79(2),286-305.

Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.

Barrick & Mount (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, *81*, 261-272.

Bartram, D. (2004). Assessment in organizations. *Applied Psychology*, 53, 237-259.

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.

Bollen KA & Liang J (1988). Some properties of Hoelter's CN. *Sociological Methods and Research*, 16(4), 492-503.

Branco e Silva, L. (2012) Exploring the collective dimension of personality using the SAPI and CPAI-2 in a sample of South African students. Unpublished master's thesis, University of Witwatersrand, Johannesburg, South Africa.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A.Bollen& J.S. Long (Eds.), *Testing structural equation models*. Newbury Park: Sage Publications.

Byrne, B. (2001). *Structural equation modelling with AMOS: Basic concepts, application, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Canivez, G. L.(in press). Bifactor modeling in construct validation of multi-factored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer& C. DiStefano (Eds., under contract), *Principles and methods of test construction: Standards and recent advancements*. Gottingen, Germany: Hogrefe Publishing.

Cheung, F.M. (2004). Use of western and indigenously developed personality tests in Asia. *Applied Psychology: An International Review*, 53, 173-191.

Cheung, F. M., Van de Vijver, F. J. R., & Leong, F. T. L. (2011). Toward a new approach to the assessment of personality in culture. *American Psychologist*, 66, 593-603.

Chrystal, E. (2012). The construction of an indigenous emotional stability scale. Unpublished masters thesis, University of Johannesburg, Johannesburg, South Africa.

Claassen, N. (1995). Cross-cultural assessment in the human sciences. Paper delivered at a work session on the meaningful use of psychological and educational tests, Human Sciences Research Council, 24 October 1995. Pretoria, South Africa.

Claassen, N. (1997). Culture differences, politics and test bias in South Africa. *European Review of Applied Psychology*, 42, 297-307.

Costa, P.T., & McCrae, R.R. (1992). *NEO-PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354.

Crowne, D. P. and Marlowe, D. (1964), *The Approval Motive*. New York: John Wiley & Sons.

Damarin, F., & Messick, S. (1965). *Response styles are personality variables: A theoretical integration*. Princeton, NJ: Educational Testing Service.

De Bruin, G.P. (2010). SAPI data analysis manual. Johannesburg, South Africa: University of Johannesburg.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological methods*, 2(3), 292.

De Goede, J., & Theron, C.C. (2010). An investigation into the internal structure of the learning potential construct as measured by the APIL-B test battery. *Management Dynamics*, 19 (4), 30-55.

Diamantopoulos, A., & Siguaw, J.A. (2000). *Introducing LISREL*. London, UK: SAGE Publications, Inc.

Dilchert, S., Ones, D. S., Viswesvaran, C., & Deller, J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, 25(3), 209-225.

Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling*, 7, 3-28.

Dilchert, S., & Ones, D. S. (2012). Application of preventive strategies. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 177-200). New York, NY: Oxford University Press.

Donnelly, C. (2009). A multi-group Structural Equation Modelling investigation of the measurement invariance of the Campbell Interest and Skill Survey (CISS) across gender groups in South Africa. Unpublished masters dissertation, University of Stellenbosch, Stellenbosch, South Africa.

Du Toit, M., & Du Toit, S.H.C. (2001). *Interactive LISREL: User's guide*. Lincolnwood: Scientific Software International.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. New York, NY: Dryden. (1957). *The social desirability variable in personality assessment and research* (2nd ed.). New York, NY: Dryden.

Edwards, A.L. (1964). Social desirability and performance on the MMPI. *Psychometrics*, 29, 295-308.

Ellingson, J.E. (2012). People fake only when they need to fake. In M. Ziegler, C. Maccann & R.D Roberts (Eds.),*New perspective on faking in personality assessment*(pp.19-33). New York, NY: Oxford.

Ellingson, J.E., Sackett, P.R., & Smith, D.B. (2001). Investigating the influence of social desirability on personality structure. *Journal of Applied Psychology*, 86, 122-133.

Emlyn, O. (2006). Experimental and correlation research designs. Power Point presentation. Braintech. [Online]. Retrieved on February, 2016 from, www.braintechllc.com.

Espinosa, A. D. C. D., & van de Vijver, F. J. (2014). An Indigenous Social Desirability Scale. *Measurement and Evaluation in Counselling and Development*, 47(3), 199-214.

Fabrigar, L.R., Wegner, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of Exploratory Factor Analysis in psychological research. *Psychological Methods*, 4, 272-299.

Ferrando, P.J. (2008). The impact of social desirability bias on the EPQ-R item scores: An item response theory analysis. *Personality and Individual Differences*, 44, 1784-1794.

Field, A. (2005). *Discovering Statistics using SPSS*(2nd ed.). London, UK: SAGE Publications.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll* (4th ed.).Los Angeles, USA:SAGE Publications.

Finch, J.F., & West, S.G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality*, 31, 439-485.

Fischer, R. (2004). Standardization to account for cross-cultural bias. *Journal of Cross-Cultural Psychology*, 35(3), 263-282.

Flattery, A. (2010). Developing and Validating a Hostility, Gratefulness and Active Support Measuring Instrument. Masters Dissertation, North-West University, Potchefstroom, South Africa.

Floyd, F.J., & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.

Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspective regarding ethical and fair practices. *European Journal of Psychological Assessment*, 13(3), 229-235

Foxcroft, C. (2004). Planning a psychological test in the multicultural South African context. *South African Journal of Industrial Psychology*, 30(4), 8-15.

Foxcroft, C., Paterson, H., Le Roux, N. & Herbst, D. (2004). Psychological assessment in South Africa: A needs analysis. The test use patterns and needs of psychological assessment practitioners. Human sciences research council, final report.[Online]. Retrieved on June 2013 from, www.hsrc.ac.za/en/research-data/ktree-doc/1186.

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

Foxcroft, C., & Roodt, G. (2009). *Introduction to psychological assessment in the South African context*(3rd ed.). Cape Town, South Africa: Oxford University Press.

Furnham, A. (1985). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7(3), 385-400.

Grimm, S.D., & Church, A.T. (1999) A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, 33, 415-441.

Guilford, J.P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education*. Tokyo: McGraw-Hill.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2006). *Multivariate data analysis: A global perspective*(7thed.). Upper Saddle River, NJ: Pearson.

Health Professions Council of South Africa (2006). Form 208: Policy on the classification of psychometric measuring devices, instruments methods and techniques. Johannesburg, South Africa: Professional Board for Psychology, HPCSA.

Heggestad, E.D. (2012). A conceptual representation of faking. In M. Ziegler, C. Maccann& R.D. Roberts (Eds.) *New perspective on faking in personality assessment*(pp. 1-14). New York, NY: Oxford.

Heine, S. J. & Lehman, D. R. (1997). The cultural construction of self-enhancement: An examination of group-serving biases. *Journal of Personality and Social Psychology*, 72, 1268-1283.

Hendriks, A. A., Hofstee, W. K., & De Raad, B. (1999). The Five-Factor Personality Inventory (FFPI). *Personality and Individual Differences*, 27, 307-325.

Hill, C., Nel, J.A., Van de Vijver, F., Meiring, D., & Valchev, V. (2013). Developing and testing items for the South African Personality Inventory. *South African Journal of Industrial Psychology*, 39(1),

Hogan, R. (2007). In defense of personality measurement: New wine for old whiners. *Human Performance,* 18(4), 331-341.

Holden, R.R. & Book, A.S. (2012). Faking does distort self-report personality assessments. In M.Ziegler, C. Maccann & R.D. Roberts (Eds.) *New perspective on faking in personality assessment* (pp. 71-84). New York, NY: Oxford.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrics*, *30*, 179-185.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance,* 11, 209-244.

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.

Hu, L., & P.M. Bentler (1995). Evaluating model fit. In R.H. Hoyle (Ed.). *Structural Equation Modelling: Concepts, issues and applications.* (pp. 76-99). Thousand Oaks CA: SAGE Publications.

Huysamen, G.K. (1979). Psychological test theory. Cape Town, South Africa: National Book Printers.

Huysamen, G.K. (2002). The relevance of the new APA standards for educational and psychological testing for employment testing in South Africa. *South African Journal of Psychology*, 32(2), 26-33.

Janse van Rensburg, C. (2010). The development of an experimental Conscientiousness measurement instrument within the SAPI project. Unpublished masters dissertation, University of Stellenbosch, Stellenbosch, South Africa.

Johnson, T. P., & Van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195-204). New York, NY:Wiley.

Jöreskog, K.G., & Sörbom, D. (1996). *PRELIS 2: User's reference guide*. Chicago, IL: Scientific Software International.

JVR Psychometrics. (2014). EEA position statement and list of HPCSA assessments. [Online]. Retrieved on September 2014 from http://www.jvrpsychometrics.co.za

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, (20), 141-151.

Kelloway, E.K. (1998). *Using LISREL for structural equation modelling: A researcher's guide*. Thousand Oaks, CA: SAGE Publications, Inc.

Kerlinger, F. N. (1973). *Foundations of behavioural research: Educational, psychological and sociological inquiry*. Holt Rinehart and Winston.

Kerlinger, F.N., & Lee, H.B. (Eds.). (2000). *Foundations of behavioural research*(4th ed.). New York, NY: Harcourt College Publishers.

Kline, R.B. (2005). Principles and practice of structural equation modelling (2nd ed.). New York, NY: Guilford.

Kriel, J. (2011). Measurement invariance of the second edition of the Fifteen Factor Personality Questionnaire (15FQ+) over different ethnic groups in South Africa. Unpublished masters dissertation, University of Stellenbosch, Stellenbosch, South Africa.

Kuncel, N.R., & Tellengen, 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.

Labuschagne, A. (2010). South African Personality Inventory: the development of an investigation into the psychometric properties of the intellect cluster. Masters dissertation, North-West University, Potchefstroom, South Africa.

Laher, S. (2008). Structural equivalence and the NEO-PI-R: Implications for the applicability of the Five-Factor Model of personality in an African context. *South African Journal of Industrial Psychology,* 34, 76-80.

Laher, S., & Cockroft, K. (2011). Psychological assessment in South Africa: Research and application (2000-2010). In S. Rothmann, D. Meiring, F. van de Vijver. &G. P. de Bruin (Eds.) *Development of the South African Personality Inventory* (SAPIhttp://www.intestcom.org/Downloads/ITC2006Brussels/Session3.2.4vandevijver.pdf).

Li, A., & Bagger, J. (2006) Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131-141.

Lin, T. Y., & Church, A. T. (2004). Are indigenous Chinese personality dimensions culture-specific? An investigation of the Chinese Personality Assessment Inventory in Chinese American and European American samples. *Journal of Cross-Cultural Psychology*, 35, 586-605.

Lotter, M. (2010). The development of an experimental Integrity instrument for various cultural groups as conceptualised from the South African Personality Inventory (SAPI) project. Unpublished masters dissertation, University of Stellenbosch, Stellenbosch, South Africa.

Mauer, K.F. (2000). Psychological test use in South Africa. Pretoria, South Africa: Department of Industrial Psychology, University of South Africa.[Adobe PDF]. Retrieved on November 2012 from http://www.pai.org.za/Psychological%20test%20use%20in%20South%20Africa.pdf

McCrae, R.R., & Costa, P.T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.

McCrae, R.R., & Terracciano, A. (2005). Universal Features of Personality Traits from the Observer's Perspectives: Data from 50 Countries. *Journal of Personality and Social Psychology*, 2005, Vol. 88, No. 3, 547–561

Meehl, P.E., & Hathaway, S.R. (1964). The K factor as suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30, 525-564.

Meiring, D. (2007). Bias and equivalence of psychological measures in South Africa. Unpublished doctoral thesis, Tilburg University, Tilburg, The Netherlands.

Meiring, D. (2012). Item development. Workshop presentation conducted 16 April 2012 at North-West University, South Africa.

Meiring, D., Van de Vijver, F. J. R., & Rothmann, S. (2006). Bias in an adapted version of the 15FQ+ questionnaire in South Africa. *South African Journal of Psychology*, 36(2), 340-356.

Meiring, D., van de Vijver, A.J.R., Rothmann, S., & Barrick, M.R. (2005). Construct, item and method bias of cognitive and personality tests in South Africa. *South African Journal of Industrial Psychology*, 31(1), 1-8.

Mels, G. (2003). *A workshop on structural equation modelling with LISREL 8.54 for Windows*. Chicago, IL: Scientific Software International.

Moorman, R.H., & Podsakoff, P.M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, 65, 131-149.

Moyo, S. (2009). *A preliminary factor analytic investigation into the first-order factor structure of the Fifteen Factor Questionnaire Plus on a sample of Black South African managers*. Unpublished master's thesis. Stellenbosch University, Stellenbosch, South Africa.

Moyo, S., & Theron, C. (2011). A preliminary factor analytic investigation into the first-order factor structure of the Fifteen Factor Plus (15FQ+) on a sample of Black South African managers. *South African Journal of Industrial Psychology*, 37(1), 01-22.

Murphy, K.R., & Davidshofer, C.O. (2005). *Psychological testing: Principles and applications* (6thed.). Upper Saddle River, NJ: Pearson Education Inc.

Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.

Nederhof, A.J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 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.

Nel, J.A., Valchev, V.H., Rothmann, S., Van de Vijver, A.J.R., Meiring, D., &De Bruin, G.P. (2011). Exploring the personality structure in the 11 languages of South Africa. *Journal of Personality*, 80(4), 915-948.

Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Odendaal, A. (2013). The measurement of social desirability: A cross-cultural perspective. Unpublished doctoral thesis, University of Stellenbosch, Stellenbosch, South Africa.

Ones, D. S., & Viswesvaran, C. (2001). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In B. W. Roberts, & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 63-92). Washington DC: American Psychological Association.

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.

Owen, K. (1991). Test bias: the validity of the Junior Aptitude Tests (JAT) for various population groups in South Africa regarding constructs measured. *South African Journal of Psychology*, 21 (2), 112-118.

Owen, K. & Taljaard, J.J. (1988). Pretoria: Human Sciences Research Council

Pallant, J. (2007). *SPSS survival manual*(3rd ed.). Two Penn Plaza, NY: McGraw-Hill House.

Paterson, H., & Uys, K (2005). Critical issues in psychological test use in the South African workplace. *South African Journal of Industrial Psychology*, 31(3), 12-22.

Paulhus, D.L. (1984). Two-component models of socially 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. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braund, D.N. Jackson, & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (2012). Overclaiming on personality questionnaires. In M. Ziegler, C. Maccann & R. D. Roberts (Eds.) *New perspective on faking in personality assessment* (pp. 151-165). New York, NY: Oxford.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025-1060.

Paulhus, D.L., & Reid, D.B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317.

Popper, K. (1972). *Objective Knowledge: An Evolutionary Approach*.

Plug, C. (1996). An evaluation of psychometric test construction and psychological services supported by the HSRC. (Unpublished HSRC report). Pretoria, South Africa: HSRC.

Preacher, K.J., & Coffman, D.J. 2006. 'Computing power and minimum sample size for RMSEA' [Online]. Retrieved April 21, 2013: http://www.quantpsy.org/

Rees, C. J.,& Metcalfe, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology*, 18,156-165.

Rossė, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on pre-employment personality testing and hiring decisions. *Journal of Applied Psychology*,83, 634-644.

Sackeim, H.A., & Gur, R.C. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37(2), 147-169.

SIOPSA. (2013). Recommendations for regulating development, control and use of psychological tests. Retrieved from: https://www.siopsa.org.za/pages/interest_groups

Smith, P.B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35(1), 50-61.

Steenkamp, J.E.M., De Jong, M.G., & Baumgartner, H. (2010). Socially desirable response tendencies in survey research. *Journal of Marketing Research*, 47, 199-214.

Struwig, F.W., & Stead, G.B. (2001). *Planning, designing and reporting research*. Cape Town, South Africa: Pearson Education South Africa.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education Inc.

Taylor, I. A. (2000) The construct comparability of the NEO-PI-R questionnaire for black and white employees. Unpublished doctoral thesis, University of the Orange Free State, South Africa.

Taylor, N., & De Bruin, G. P. (2005). *Basic Traits Inventory: Technical manual*. Johannesburg, South Africa: Jopie van Rooyen & Partners SA.

Taylor, N. (2008). Construct, item and response bias across cultures in personality measurement (Unpublished doctoral thesis). Rand Afrikaans University, South Africa.

Theron, C. (2010). Lecture 6: Psychometrics, measurement theory, test construction and decision-making. [PowerPoint Slides]. Retrieved from, BS743, Stellenbosch University, Stellenbosch, South Africa.

Theron, C. (2011a). Lecture 1-3: Intermediate statistics and computer usage. [PowerPoint Slides]. Retrieved from, BS815, Stellenbosch University, Stellenbosch, South Africa.

Theron, C. (2011b). SPSS and LISREL Output when fitting a measurement and structural models in single group and multi-group SEM analysis. [Microsoft Word Document]. Retrieved from, BS815, Stellenbosch University, Stellenbosch, South Africa.

Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5(3),243-262

Valchev, V. H., Van de Vijver, F. J. R., & Meiring, D. (2013, July). The SAPI project: Development and validation of an indigenous personality assessment tool in South Africa. Presentation at the 12th European Conference on Psychological Assessment, San Sebastian, Spain.

Valchev, V.H., Van de Vijver, A.J.R., Meiring, D., Nel, J.A., Laher, S., Hill, C. (2014). The South African Personality Inventory (SAPI): An indigenous personality tool for the different ethno cultural groups of South Africa. Manuscript submitted for publication.

Valchev, V. H., Van de Vijver, F. J. R., Nel, J. A., Rothmann, S., Meiring, D., & De Bruin, G. P. (2011). Implicit personality conceptions of the Nguni cultural-linguistic groups of South Africa. *Cross-Cultural Research*, 45, 235-266.

Valchev, V. H., Van de Vijver, A. J. R., Meiring, D., Nel, J. A., Laher, S., Hill, C., & Adams, B. G. (2014). Beyond agreeableness: social-relational personality concepts from an indigenous and cross-cultural perspective. *Journal of Research in Personality*, 48 1: 17-32.

Van de Vijver, A. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.

Van de Vijver, A.J.R., & Rothmann, S. (2004). Assessment in multicultural groups: the South African case. *South African Journal of Industrial Psychology*, 30(4), 1-7.

Van Herk, H., Poortinga, Y.H., & Verhallen, T.M.M. (2004). Response styles in rating scales. *Journal of Cross-Cultural Psychology*, 35, 345-360.

Visser, D., & Viviers, R. (2010). Construct equivalence of the OPQ32N for Black and White people in South Africa. *South African Journal of Industrial Psychology*, 36(1), 1- 11.

Vorster, P. (2010). A factor analytic comparison of dichotomous and polyamorous response categories in personality inventories. Unpublished Masters' dissertation, University of Johannesburg, Johannesburg, South Africa.

Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement*, 24, 551-562.

Worthington, R.L., & Whittaker. T.A. (2006). Scale development research. A content analysis and recommendations for best practices. *The Counselling Psychologist*, 34(6), 806-838.

Ziegler, M., Maccann, C., & Roberts, R.D. (2012). Faking: Knowns, unknowns and points of contention. In M. Ziegler, C. Maccann & R.D. Roberts (Eds.) *New perspective on faking in personality assessment*(pp. 3-18). New York, NY: Oxford.