Investigation of the optimal response scale for personality measurement: Computer-based testing

Elizabeth Maria Classen, Hons.B.Com

Mini-dissertation submitted in partial fulfilment of the requirements for the degree Magister Commercii in Industrial Psychology at the North-West University (Potchefstroom Campus)

Supervisor: Dr Alewyn Nel

Language Editor: Prof. Annette Combrink

Declaration of originality of research

DECLARATION

I, Elizabeth Maria Classen, hereby declare that the dissertation entitled *Investigation of the optimal response scale for personality measurement: Computer-based testing* is my own work and that the views and the opinions expressed in this study are those of the author and relevant literature references as listed in the reference list. I also declare that the content of this research will not be handed in for any other qualification at any other tertiary institution.

_____

ELIZABETH M. CLASSEN

NOVEMBER 2011

COMMENTS

The reader is reminded of the following:

- The editorial style and the references in the mini-dissertation are an accordance with the format accepted and prescribed by the Publication Manual (5th edition) of the American Psychological Association (APA). This is according to the prescribed policy of the Industrial Psychology Programme as presented at the North-West University (Potchefstroom Campus),

- This mini-dissertation is submitted in the form of a research article.

ABSTRACT

Topic: Investigation of the optimal response scale for personality measurement: Computer-based testing

Keywords: personality measurement, computer-based measurement, dichotomous response scale, polytomous response scale, Likert scale, validity, reliability.

To be able to use personality tests in the most reliable and valid manner there are many considerations to be taken into account. Variables such as the population used, the culture of the test-takers, the mode of administration, whether pencil-and-paper or computer-based testing procedures, familiarity with computers when using computer-based tests and the response format to be used when administering the personality questionnaire are but some of the considerations.

Within South Africa it is that much more important to consider the mode of administration, whether pencil-and-paper tests or computer-based tests, as there are many individual groups who have been historically disadvantaged when it comes to the use of computers as a testing method. It is just as important to consider the response scale to be utilised when administering personality testing as this may influence the results obtained and can influence the reliability and validity of these results.

The objective of this study was to determine which response scale, dichotomous or polytomous, was the best to use when conducting computer-based personality testing. The questionnaire that was utilised was the South African Personality Inventory (SAPI) questionnaire; however, only items from the Soft-Heartedness cluster were employed as the objective was not to test the questionnaire but to test the most reliable and valid response scale to be used in conjunction with the questionnaire. A convenience sampling approach was utilised and the questionnaire was administered to students who were available and able to take the test (N = 724). Descriptive statistics, factor analysis and Cronbach Alpha coefficients were used to analyse the data obtained.

By means of a literature review it was found that computer-based testing held many advantages when utilised with personality testing and that it can assist in better testing in the future. Literature also showed that polytomous or Likert-type response scales yielded better results than that of a dichotomous or forced-choice response scale. The factor loadings on the two different response scales also yielded evidence that the polytomous response scale had three distinct factors that measure Soft-Heartedness, namely *generosity*, *compassion* and *appreciation*, which is supported by the findings of the South African Personality Inventory (SAPI), whereas the dichotomous response scale only yielded one overall factor.

The results indicated that the polytomous scale is a more optimal option than the dichotomous scale. It was found that the polytomous response scale had much better inter-item-correlations and item-factor loadings than that of the dichotomous response scale. It was also found that the polytomous response scale had a much higher internal consistency than that of the dichotomous response scale since its alpha coefficient is higher than 0,70, whereas the internal consistency of the dichotomous response scale was lower than 0,70.

This study's limitations were also identified and recommendations were made in terms of future research and for utilisation within the South African Personality Inventory (SAPI) questionnaire.

OPSOMMING

Onderwerp: 'n Ondersoek na die optimale respons-skale vir persoonlikheidsmeting: Rekenaargebaseerde toetsing

Sleutelwoorde: persoonlikheidsmeting, rekenaargebaseerd, dichotomiese responsskaal, poliotomiese responsskaal, Likert-skaal, geldigheid, betroubaarheid

Om persoonlikheidstoetse op die mees betroubare en geldige wyse af te neem is daar baie oorweginge wat in gedagte gehou moet word. Veranderlikes soos die populasie wat gebruik word, die kultuur van die toetsafleggers, die toepassings-modus, die vraag of dit pen en papier of rekenaargebaseerd moet wees, vertroudheid met rekenaars in die aflê van rekenaargebaseerde

toetse en die formaat wat gebruik moet word in die toepas van die persoonlikheidsvraelys is maar ’n paar van die belangrike oorwegings.

In Suid-Afrika is dit nog baie belangriker om versigtige oorweging te skenk aan die toepassings-modus, die gebruik van pen en papier of rekenaargesteunde toetse in die lig van die baie individuele sosiale groepe wat histories agtergeblewe is in terme van die gebruik van rekenaars binne die konteks van toetsing. Dit is net so belangrik om aandag te gee aan die responsskaal wat gebruik word in die afneem van persoonlikheidtoetse, aangesien dit die resultate kan beïnvloed, wat weer ’n invloed kan hê op die geldigheid en betroubaarheid van die resultate.

Die doel van die studie was om te bepaal watter responsskale, dichotomies of poliotomies, die beste is om te gebruik wanneer rekenaargebaseerde persoonlikheidstoetsing gedoen word. Die vraelys wat gebruik is, is die *South African Personality Inventory (SAPI*), maar slegs items uit die *Soft-Heartedness cluster* is gebruik aangesien die doel van die studie nie was om die vraelys te toets nie maar om te toets wat die betroubaarste en geldigste responsskaal is om te gebruik in samehang met die vraelys. ‘n Gerieflikheidsteekproef is gebruik en die vraelys is afgeneem onder studente wat beskikbaar was en in staat om die toets te skryf (*N*=724).

Beskrywende statistiek, faktoranalise en Cronbach Alpha ko-effisiënte is gebruik om die data wat verkry is te ontleed. Deur ’n literatuurstudie is gevind dat rekenaargebaseerde toetsing baie voordele inhou wanneer dit gebruik word in persoonlikheidstoetsing en dat dit kan lei tot beter toetsing in die toekoms. Literatuur het ook aangetoon dat poliotomiese of Likert-tipe respons beter resultate oplewer as dichotomiese of gedwonge-keuse responsskale. Die faktorladings op die verskillende responsskale het ook aangetoon dat die poliotomiese responsskaal drie spesifieke faktore bevat wat *Soft-Heartedness* meet, naamlik vrygewigheid, empatiese gevoelens en waardering (*generosity, compassion* en *appreciation*). Dit word gesteun deur die bevindings van die SAPI waar die dichotomiese responsskaal slegs een oorkoepelende faktor gelewer het.

Die resultate het ook aangetoon dat die poliotomiese skaal ’n meer optimale opsie is as die dichotomiese skaal. Daar is ook gevind dat die poliotomiese responsskaal baie beter inter-item korrelasies gelewer het die dichotomiese responsskaal. Ook is dit duidelik dat die poliotomiese

responsskaal ’n baie hoër interne konsekwentheid vertoon as die dichotomiese skaal, aangesien die alpha-ko-effisiënt hoër is as 0,70, waar die interne konsekwentheid van die dichotomiese skaal laer is as 0,70.

Die leemtes van die studie is ook geïdentifiseer en aanbevelings is gemaak in terme van toekomstige navorsing en vir gebruik binne die SAPI.

ACKNOWLEDGEMENTS

I would like to thank and give gratitude to the following people for their contribution to this mini-dissertation:

- Firstly my Heavenly Father for the strength, wisdom and talents that He has given to me and providing me with the will to do my best and give my all.
- My fiancé, Werner, for supporting me in all my achievements and for encouraging me when I needed it most.
- My parents, for providing me the opportunity to study at a tertiary institution and believing in me and my abilities.
- My supervisor Dr. Alewyn Nel for his wisdom, expertise and assistance with this mini-dissertation.
- To Dr. Carin Hill for her contribution in the statistical analysis.
- Professor A.L. Combrink who assisted with the language editing.
- The South African Personality Inventory (SAPI) team for providing me with this opportunity to give a contribution to the SAPI project.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This mini-dissertation aims to evaluate whether a dichotomous or polytomous response scale is best when administering computer-based personality tests.

This chapter entails the problem statement as well as the research objectives, which is the general objective as well as the specific objectives, and gives an explanation of the research method used. An overview of the chapters is also provided.

1.1 PROBLEM STATEMENT

1.1.1 Overview of the problem

Personality tests have, for more than 20 years, been used increasingly for personnel selections (Tett & Christiansen, 2007). Due to the potential to increase organisational success and theory-driven advances, personality testing in organisations has evolved into a foremost topic of research (Converse et al., 2008; Scroggins, Thomas & Morris, 2009; Tett & Christiansen, 2007,). Personality measurement has turned into a popular method of selection in high-stakes hiring contexts. Research has proven that personality tests can be used in forecasting job execution in the future (Morgenson et al., 2007a).

According to Goldman (2009), the ability to use personality measurement to identify an individual's causal needs, and the motivations that serve as a driving force within that individual, are an essential part of the selection process. In a recently published sequel article Morgenson et al. (2007b) stated that although there is extensive and rising acceptance of published personality tests and their validity in terms of the prediction of job performance, most of these tests should be abandoned due to low validity in personnel selection procedures. Personality tests may have value for use in employee selection but more recent research is needed to determine the validity of these tests in performance prediction (Holcombe Ehrhart, Roesch, Ehrhart & Kilian, 2008; Lievens, De Corte & Schollaert, 2008).

In South Africa the field of psychology has taken to a more scientific exploration and investigation of personality, and as a result personality theories and factors as well (Laher, 2008; Joseph & Van Lill, 2008). Changes in South Africa's political, economical and social history, in terms of psychological measurement, have led to increased pressure on test developers and test users to only create and apply appropriate and valid test materials (Foxcroft, 2004; Van Eeden & Mantsha, 2007). According to Visser and Viviers (2010), South African psychologists have to take into consideration all the set pointers and guiding principles and with every action keep in mind that the psychometric bias properties of tests and the fairness of the uses of tests have to be addressed in a scientific manner. The Employment Equity Act 55 of 1998 (EEA), that forms part of South African legislation, induces the use of fair testing practices (Joseph & Van Lill, 2008).

The Employment Equity Act No.55 of 1998 (as cited in Davies, Foxcroft, Griessel &Tredoux, 2005) states that:

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

Nationally and internationally a great range of assessments designed specifically for the personnel and organisational environment have emerged in the past few years (Laher, 2008), and have since evolved into the use of technological innovations and the use of newly-surfaced information technology. Joubert and Kriek (2009) also state that the development of computers have led to the use of computer-based measurement instead of the traditional pencil-and-paper measurement.

Kline (2002) notes that it is possible to computerise almost any traditional psychometric test but that it is essential to check the reliability and validity of the computer-based version of each test, no matter how similar it may be to the paper-and-pencil-based test. Bartram (2006) found those organisations using computer-based assessments could benefit from the potential of a higher validity early sift and more cost-efficient selection. In the past ten years there has been a sufficient increase in employment tests available on the Internet for selection, recruitment and development, however, unfortunately these included tests that were not scientifically validated (Joubert & Kriek, 2009).

One of the greatest perceived advantages of computer-based measurement and computer-aided assessment is the provision of focused, faster and timely feedback (Bartram, 2006; Bull, 1999; Conole & Warburton, 2005; Lim, Ong, Wilder-Smith & Seet, 2006). This leads to larger applicant pools, significant reductions in time to hire, reduced cost per hire and the job profiling tools to set competency standards and profiles against which to assess candidate fit (Bartram, 2006). It also allows for more complex scenario designs which act dynamically over time and react to the applicant's decisions directly (Bocij & Greasley, 2007; Funke, 1998; Hornke, 2005).

Due to the fact that the user expectations are that computers should operate reliably and that there is no need for "soft" products, printing, production, warehousing or postal delivery services anymore, a rise in reliability of computer-based measurement has been witnessed (Bartram, 2006). Bartram (2006) and Kline (2002) both agree that some of the advantages of computer-based measurement can include the innovative use of graphics, sound and multi-media, the use of simulations, timed-responses, simplified answering like the point-and-click simulation of the multiple response of a paper-and-pencil test and the recording of very detailed information about the respondent's performance.

According to the International Test Commission Guidelines on Computer-based and Internet delivered Measurement (ITC, 2005) all computer-based measurement and computer-aided assessment must adhere to the following standards:

- It must ensure that the technical aspects of the computer-based measurement are considered, particularly with the hardware and software used for the measurement;
- The quality of the measurement and test materials must be ensured and assured and good practice must be guaranteed throughout the measurement process;
- Delivery of tests, test-taker authentication and prior practice must be controlled; and
- Security of measurement materials, privacy, data protection and confidentiality must be ensured at all times.

Within computer-based personality measurement, as with all psychometric measurement, consideration must be given to the response scale to be used. A scale is a "series of self-report questions, ratings or items that are used to measure a concept" (Stelmack, 2009). Where, a

response scale can be defined as "…the way you collect responses from people on an instrument. You might use a dichotomous response scale like agree/disagree… …or you may use an interval response scale like a 1 to 5 rating" (Trochim2006).

A dichotomous response scale is a structure which forces, a choice between one of two options representing opposite ends of the scale or alternative choices (Litzinger, Lee, Wise & Felder, 2007). Some of the formats of the dichotomous scale can include a yes/no format (Bijttebier, Vertommen & Vander Steene, 2001; Pryce, Munir & Haslam, 2006), an agree/disagree format (González Romá & Espejo, 2003; Kapadia-Kundu & Dyalchand, 2007), a true/false format (Mullins, Polson, Lanch & Kehoe, 2005; Petrides, Frederickson & Furnham, 2004) and others like the bad/good format or the harmful/beneficial format (Wegener, Downing, Krosnick & Petty, 1995).

A polytomous response scale is a multipoint or continuous format that reflects levels of agreement or different probabilities including a single mid-point (Drasgow, 2005). Stelmack (2009) describes it as a list of ordered response categories such as a rating of difficulty or importance. The polytomous response scale includes some formats such as the 5-point rating scale, ranging from 1 = "not at all applicable" to 5 = "strongly applicable" (Bijttebier et al., 2001), formats including either "?", "in between" or "not sure" as the middle categories (González Romá & Espejo, 2003) and formats ranging from 1 = "strongly disagree" to 5 = "strongly agree" (Sá & Stanovich, 2001).

The Likert scaling method is seen as one of the most widely used polytomous response scaling methods (Gardner & Martin, 2007). Generally with the Likert response methods the responses are ordered from one extreme to another and as common practice the responses are coded as whole numbers (Swain, Weathers & Niedrich, 2008). Likert-scale items are usually used to examine the attitudes of the respondents to a series of written or verbal statements, and the respondents are then asked to represent their strength of feeling in a categorical manner (Dittrich, Francis, Hatzinger & Katzenbeisser, 2007). This is mostly a scale from 1 to 5 and is typically defined by extreme points such as "strongly agree" to "strongly disagree" (Converse et al., 2008; Sá & Stanovich, 2001).

Across many different areas of inquiry and a series of disciplines, reliable and valid measures are seen as the foundation of quality research, because confidence in the quality of the research findings is essential (Noar, 2003). Reliability can be described as the consistency with which a measurement measures whatever it measures and validity is about what the test measures and how well it does measure what it is suppose to measure (Foxcroft & Roodt, 2009).

Dichotomous response scales are chosen to avoid "fence sitting" and as a result increase the chances of identifying preferences and opinions. They are seen as forced-choice measures that in reality force the respondent to choose between two options (Converse et al., 2008). However dichotomous items are likely to lead to reduced reliability due to the fact that the two alternatives may not be sufficient for discriminating between differences consistently (Litzinger et al., 2007) and allowing for less variance and covariance than polytomous or Likert type response scales (Kline, 2002; Noar, 2003). The advantage of a polytomous scale over a dichotomous scale is that it tends to have more variability in the data and as a result increased reliability (Bijttebier et al., 2001).

Mullins et al. (2005) identified within their results that respondents who answered their integrity tests using a 5-point response scale observed themselves to have performed better on the test and also perceived the test to have more face validity, than respondents who answered an identical test using a dichotomous response scale. However no such differences were found when using the dichotomous and 5-point versions of a personality inventory.

In the study of Litzinger et al. (2007) the Felder-Solomon Index of Learning Styles (ILS) measure was changed from a dichotomous response scale to a 5-point scale to determine reliability and construct validity. There was no substantial change in construct validity although changing the response scale improved the reliability. This indicates that some types of measurement methods may be more dependent and affected by the number of response points than other selection measures.

By comparing a dichotomous response scale with a Likert-type response scale Kotsalis (1995) found in his study that the internal consistency reliability coefficients of the two scales did not

differ much, respectively at 0,87 and 0,88. The correlations obtained between scores with the dichotomous response scale were actually of a higher magnitude than that of the Likert-type response scale, although on measures of the scale's convergent validity the difference was not statistically significant. Kotsalis (1995) also concluded that the use of a dichotomous scale did not fully capture all the distinctions in the attitudes of the respondents.

There are sufficient studies in which dichotomous response scales and polytomous response scales are compared, within the subject of learning styles, integrity tests and attitude tests respectively (Bijttebier et al., 2001; Kline, 2002; Kotsalis, 1995; Litzinger et al., 2007; Mullins et al., 2005; Noar, 2003). Unfortunately, there is avisible lack in research done where dichotomous response scales and polytomous response scales are compared within the subject of personality measurement and as a result also a lack of research done on computer-based personality measurement.

Due to the fact that personality measurement can benefit the organisation tremendously in terms of selection and assessment procedures, as well as person-job fit, (Converse et al., 2008; Goldman, 2009; Hausknecht, 2010; McDonald & Edwards, 2007; Morgenson et al., 2007a; Scroggins et al., 2009) it is important to explore every aspect of these tests that can influence the validity and reliability of the tests. This then focuses the attention on the response scales, in part, because it can influence the respondent reactions to the test and, in part, because it can influence the validity and/or reliability of the data collected.

1.1.2 Research questions

The following research questions emanated from the problem statement:

- According to literature, how are personality and computer-based measurement conceptualised?
- To what degree does the type of response scale influence the reliability and validity of the personality test?
- What recommendations can be made for future research and practice?

1.1.3 Expected contribution of the study

1.1.3.1 Expected contribution for the individual

By identifying the optimal response scale to be used within computer-based personality measurement, the maximum validity and reliability of the personality test can be achieved in relation to the response scale used, thus resulting in quality research findings and increased confidence in the data used from the measurement which will benefit applicants when applying for employment by identifying the most qualified individual. This in turn can only benefit the individual in the sense that it will convey accurate results during selection and measurement procedures.

1.1.3.2 Expected contribution for the organisation

Personality measurement plays such an important role in organisational selection and assessment procedures today that it has to be as reliable and valid as possible. Especially with the use of computer-based assessments and selection via the Internet it is essential to increase the reliability and validity of the personality test by choosing the right response scale. Also, by gaining insight into the advantages of computer-based measurement, personality measurement will benefit immensely. Organisations can benefit from highly accurate selection procedures and as a result choose the correct candidate, which will in turn lead to more productivity and performance in the long run. It will also contribute to the organisation's image in the sense that applicants will perceive the organisation as fair in its selection procedures.

1.1.3.3 Expected contribution to the field of Industrial Psychology

By identifying the optimal response scale to be used for the South African Personality Inventory (SAPI), the developing SAPI questionnaire will yield more reliable and valid results, thus contributing to psychometric measurement within South Africa across all 11 official languages. This in turn will benefit the field of Industrial Psychology by contributing to South African measurement practices in an appropriate, valid and fair manner.

1.2 RESEARCH OBJECTIVES

The research objectives are divided into a general objective and specific objectives.

1.2.1 General objective

The general objective of this research is to identify the optimal response scale, dichotomous or polytomous, to be used during computer-based personality measurement.

1.2.2 Specific objectives

The specific objectives of this research were:

- To conceptualise personality and computer-based measurement according to literature.
- To identify to what degree the response scale influences the reliability and validity of the personality test.
- To identify recommendations for future research and for practice.

1.3 RESEARCH DESIGN

1.3.1 Research approach

A quantitative research approach is followed. It is a form of conclusive research and involves great representative samples as well as moderately structured procedures of data collection (Struwig & Stead, 2001). For the purpose of this study a cross-sectional design is employed which entails a study that takes place at a single point in time (Du Plooy, 2002). The data used consists of primary data and the approach is factor analysis and the examination of alpha coefficients.

1.3.2 Research method

1.3.2.1 Literature review

In phase 1 a complete review regarding personality measurement, computer-based measurement, dichotomous response scales and polytomous response scales is done. The sources that are consulted include:

- SAePublications
- EbscoHost
- Google Scholar
- Sagepub
- PsychInfo
- Jopie Van Rooyen Database
- South African Personality Inventory (SAPI) Database
- Nexus database
- BSCW (SAPI website containing previous masters' and doctorate studies on the SAPI, presentations, and other relevant literature)

Keywords that are used in the literature review include: personality measurement, computer-based measurement, computer-aided measurement, dichotomous response scale, polytomous response scale, Likert scale, 5-point response scale, forced-choice response scale, validity and reliability.

1.3.2.2 Research participants

The target population consists of undergraduate and post-graduate University students (N = 724) from a South African tertiary institution. The participants differ in terms of age, race, language and gender. The sampling method used is convenience non-probability sampling since participants are easily accessible and available. The total participants are split into two groups where different types of response scales (dichotomous or polytomous scale) are administered.

1.3.2.3 Measuring instrument

An adapted version of the South African Personality Inventory (SAPI) is used in this study. Nine personality clusters were identified in the first phase of the SAPI project (refer to Nel et al., in press) namely: Extraversion, Soft-Heartedness, Conscientiousness, Emotional Stability, Intellect, Openness, Integrity, Relationship Harmony and Facilitating. It further consists of 37 sub-clusters and 191 facets. For the purpose of this study, which is to identify the optimal response scale to be used, three facets from the Soft-Heartedness cluster, namely *generosity*, *compassion* and *appreciation* are used (Nel et al., in press) and are measured with a total of 24 items.

The rationale for not using the complete instrument, is that the optimal response scale for a computer-based personality test is the main concern and not how the facets, sub-clusters or clusters combine or coincide. Therefore only the facets that performed well in the initial pilot studies (see Flattery, 2010) conducted on the Soft-Heartedness cluster are utilised as a basic method of gathering data in terms of the response scales.

The dichotomous response scale is in the form of 0 = "disagree" and 1 = "agree", where the polytomous response scale consists of 1 = "strongly disagree", 2 = "disagree", 3 = "somewhat disagree/somewhat agree", 4 = "agree" and 5 = "strongly agree". The test booklets contain identical questions and items, only the response scale is changed between a dichotomous response scale and a polytomous response scale. Examples of statements that are answered under each sub-cluster include: *generosity* - "I give things to people without expecting anything in return", *compassion* – "I feel sympathy for people who have problems", and *appreciation* – "I am happy when I see good things happening in other people's lives"). *Generosity* can be defined as "Liking to spend money on others and sharing things with them.", *compassion* can be defined as: "Being concerned with the well-being of others and showing mercy. Being sensitive to the needs of others, and doing good for others. Being compassionate and empathetic.", and finally, *appreciation* can be defined as: "Expressing thankfulness and gratitude. Commending others on work well done. Enjoying, being grateful and satisfied with what you have". These three sub-clusters performed well during the initial pilot studies of the SAPI project (Flattery, 2010).

1.3.2.4 Research procedure

Authorisation from the tertiary institution was obtained thus allowing the researcher to use students to gather the data. The lecturers of the students are approached and asked for assistance in gathering the data. As agreed upon by their lecturers, students are approached during a class session. The students are asked to sign a consent letter allowing the use of their information and inputs in the study. It is made clear that they have certain rights including: anonymity, the right to withdraw at any time and confidentiality. During the taking of the test the appropriate instructions are given for the answering of the questionnaire. The questionnaire is taken

electronically by means of computers provided by the universities. The students are required to login and complete the questionnaire.

1.3.2.5 Statistical analysis

To identify the optimal response scale statistical analysis is conducted by means of the SPSS programme (SPSS, 2008). Analysis of the data is done by descriptive statistics, factor analysis and Cronbach Alpha coefficient analysis. It determines which of the response scales have the highest functionality, validity, and reliability. Within descriptive statistics the data is analysed by looking at the distribution of the items (means, standard deviation, skewness and kurtosis). With factor analysis the inter-item correlations and item loadings on the overall cluster of Soft-Heartedness are investigated. The reliability is determined by identifying the Cronbach Alpha coefficients of each facet (*generosity*, *compassion* and *appreciation*). A cut-off point of 0,70 for the Cronbach Alpha is utilised (Nunnally & Bernstein, 1994).

1.4 CHAPTER DIVISION

The chapters in this mini-dissertation are presented as follows:

Chapter 1: Introduction.

Chapter 2: Research article.

Chapter 3: Conclusions, limitations and recommendations.

REFERENCE LIST

Bartram, D. (2006). Measurement on the internet: Issues, challenges and opportunities in the field of occupational assessment. In Bartram, D., & Hambleton, R. K. (Eds.), *Computer-based measurement and the internet: Issues and advances* (13-38). Thames Ditton, UK: John Wiley & Sons, Ltd.

Bijttebier, P., Vertommen, H., & Van der Steene, G. (2001). Assessment of cognitive coping styles: A closer look at situation-response inventories. *Clinical Psychology Review, 21*(1), 85-201.

Bocij, P., & Greasley, A. (2007). Can computer-based measurement achieve quality and efficiency in assessment? *International Journal of Educational Technology, 1*(1).

Bull, J. (1999). Computer-assisted assessment: Impact on higher education institutions. *Educational Technology & Society, 2*(3), 123-126.

Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-F, Research in Learning Technology, 13*(1), 17-31.

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16*(2), 155-169.

Davies, C., Foxcroft, C., Griessel, L., & Tredoux, N. (2005). Computer-based and internet delivered assessment. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context* (2nded.), (pp 153-166) Cape Town: Oxford University Press.

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*(1), 3-28.

Drasgow, F. (2005, April). *Computerized measurement and assessment: Boon or boondoggle?* Presidential address at the SIOP conference, Los Angeles, CA.

Du Plooy, G. M. (2002). *Communication research: techniques, methods and applications.* Lansdowne: Juta Publishers.

Flattery, A. (2010). *Developing and validating a hostility, gratefulness and active support measuring instrument*. Unpublished masters' dissertation, North West University, Potchefstroom, South Africa.

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

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

Funke, J. (1998). Computer-based measurement and training with scenarios from complex problem-solving research: Advantages and disadvantages. *International Journal of Selection and Assessment, 6*(2), 90-96.

Gardner, H. J. & Martin, M. A. (2007). Analyzing ordinal scales in studies of virtual environments: Likert or lump it! *Presence, 16*(4), 439-446.

Goldman, R. (2009, September). Personality measurement: Identify your top performers before you reduce your workforce. *Industrial Engineer,* 41-45.

González Romá, V., & Espejo, B. (2003). Measurement the middle response categories «not sure», «in between» and «?» in polytomous items. *Psicothema, 15*(2), 278-284.

Hausknecht, J. P. (2010). Candidate persistence and personality test practice effects: Implications for staffing system management. *Personnel Psychology, 63,* 299-324.

Holcombe Ehrhart, K., Roesch, S. C. Ehrhart, M. G., & Killian, B. (2008). A test of the factor structure equivalence of the 50-item IPIP Five-Factor model measure across gender and ethic groups. *Journal of Personality Assessment, 90*(5), 507-516.

Hornke, L. F.(2005). Response time in computer-aided measurement: A “verbal memory” test for routes and maps. *Psychology Science, 47*(2), 280-293.

International Test Commission, ITC. (2005).*Guidelines on Computer-based and Internet Delivered Measurement*: ITC.

Joseph, L., & Van Lill, B. (2008). Investigating subscale differences among race and language groups on the occupational personality profile. *SA Journal of Psychology, 38*(3), 501-514.

Joubert, T., & Kriek, H. J. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *SA Journal of Industrial Psychology, 35*(1), 1-11.

Kapadia-Kundu, N., & Dyalchand, A. (2007, June). The Pachod Paisa scale: A numeric response scale for health and social services. *Demography India.*

Kline, R. (2002). *Handbook of psychological measurement* (2nd ed.). New York: Routledge.

Kotsalis, L. (1995). *A comparison of the internal consistency reliability and convergent validity of two response formats for the Beck hopelessness scale.* Unpublished masters' dissertation, University of British Columbia, Columbia, Canada..

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. *SA Journal of Industrial Psychology, 34*(1), 76-80.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*(2), 268-279.

Lim, E. C. H., Ong, B. K. C., Wilder-Smith, E. P. V., & Seet, R. C. S. (2006). Computer-based versus pen-and-paper measurement: Students' perception. *Annals Academy of Medicine, 35*(9), 599-603.

Litzinger, T. A., Lee, S. H., Wise, J. C., & Felder, R. M. (2007). A psychometric study of the index of learning styles. *Journal of Engineering Education, 96*(4), 309-319.

McDonald, S., & Edwards, H. M. (2007). Who should test whom? *Communications of the ACM, 50*(1), 67-71.

Morgenson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60*(4), 1029-1049.

Morgenson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729.

Mullins, M. E., Polson, J. M., Lanch, T., & Kehoe, K. (2005). Respondent perceptions of integrity and personality measures: Does response format make a difference? *Applied HRM Research, 11*(2), 107-118.

Noar, S. M. (2003). The role of structural equation modelling in scale development. *Structural Equation Modelling, 10*(4), 622-647.

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

Petrides, K. V., Frederickson, N., & Furnham, A. (2004). The role of trait emotional intelligence in academic performance and deviant behaviour at school. *Personality and Individual Differences, 36,* 277-293.

Pryce, J., Munir, F., & Haslam, C. (2006). Cancer survivorship and work: Symptoms, supervisor response, co-worker disclosure and work adjustment. *Journal of Occupational Rehabilitation, 1,* 1-9.

Sá, W. C., & Stanovich, K. E. (2001). The domain specificity and generality of mental contaminations: Accuracy and projection in judgments of mental content. *British Journal of Psychology, 92,* 281-302.

Scroggins, W. A., Thomas, S. L., & Morris, J. A. (2009). Psychological measurement in personnel selection, part III: The resurgence of personality measurement. *Public Personnel Management, 38*(1), 67-77.

SPSS Inc. (2008). *SPSS 16.0 for Windows*. Chicago, IL: SPSS Inc.

Stelmack, J. (2009). Quality of life assessment and research on health care outcomes. In Stelmack. J. (Ed.) *Optometric care within the public health community* (1-14). New York: Old Post Publishing.

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

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing the three sources of misresponse to reversed Likert items. *Journal of Marketing Research, 15,* 116-131.

Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgenson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt. *Personnel Psychology, 60,* 967-993.

Trochim, M. K. (2006). *Research methods knowledge base.* Retrieved June 17, 2010, from http://scholar.google.co.za/scholar?start=20&q=Trochim+%22Research+Methods+Knowledge+Base%22+author:Trochim&hl=en&as_sdt=2006.

Van Eeden, R., & Mantsha, T. R. (2007). Theoretical and methodological considerations in the translation of the 16PF into an African language.*SA Journal of Psychology, 37*(1), 62-81.

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

Wegener, D. T., Downing, J., Krosnick, J. A., & Petty, R. E. (1995). Measures and manipulations of strength-related properties of attitudes: Current practice and future directions. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (1-18).

CHAPTER 2

RESEARCH ARTICLE

Investigation of the optimal response scale for personality measurement: Computer-based tests

ABSTRACT

The general objective of this study was to determine which response scale, polytomous or dichotomous, would be best to use within computer-based testing when using a personality measurement. The sample (N = 724) consisted of undergraduate and post-graduate students from a tertiary institution in South Africa. The measuring scale instrument that was used was an adapted version of the Soft-Heartedness questionnaire from the SAPI project. When looking at the skewness and kurtosis rating of the two response scale versions it was found that very few items from the computer-based dichotomous response scale version were normally distributed; however, all the items from the computer-based polytomous response scale version showed normal distribution. It was also found that although both versions of the response scales showed an average factor loading above 0,30, the polytomous response scale had a general higher average factor loading (0,57) than the dichotomous response scale (0,50). The polytomous response scale also showed a higher internal consistency than that of the dichotomous response scale, which was eminent from the Cronbach Alpha coefficients (α) being α = 0,88 and α = 0,58 respectively. Only the Cronbach Alpha coefficient of the polytomous response scale adhered to the general guideline of a score of α > 0,70.

OPSOMMING

Die algemene objektief van hierdie studie was om te bepaal watter reponsskaal, polionomiese of dichotomiese, die beste is om te gebruik wanneer rekenaargebaseerde persoonlikheidstoetsing gedoen word. ‘n Gerieflikheidsteekproef is gebruik en die vraelys is afgeneem onder voorgraadse en nagraadse studente van ‘n tersiêre institusie in Suid-Afrika wat beskikbaar was en in staat was om die toets te skryf (N=724). Die meetinstrument wat gebruik was, was die aangepaste weergawe van die *South African Personality Inventory (SAPI)*. Wanneer daar na die skeefheid en kurtosis van die tellings van die items kyk was, is daar gevind dat baie min items van die rekenaargebaseerde dichotomiese responseskaal normaal verspreid was, maar dat al die items van die polionomiese responsskaal normaal verspreid was. Daar was ook gevind dat albei weergawes van die responsskale ‘n gemiddelde faktor lading bo 0,30 gehad het. Die rekenaargebaseerde polionomiese responsskaal het ‘n algemene gemiddelde faktor lading gehad (0,57) wat hoër was as dié van die

dichotomiese responsskaal (0,50). Die polionomiese respons skaal het ook ‘n hoër interne konsekwentheid gehad as die dichotomiese responsskaal, wat duidelik was met hulle Cronbach Alpha (α) ko-effisiënte van $\alpha = 0,88$ en $\alpha = 0,58$ onderskeidelik. Net die Cronbach Alpha ko-effisiënt van die polionomiese response skaal het voldoen aan die algemene riglyn van ‘n telling van $\alpha > 0,70$.

The context of personality testing can be divided into many different fields or schools of thought and can yield many different arguments and paradigms, both nationally and internationally. It is therefore important to consider “what” personality is and “how” it is defined within the context of “personality testing”. Engler (2009) defines personality as “someone’s public image” (p. 2), therefore the way a person is perceived within a public or open environment or is seen as by others. The defining and understanding of personality can be seen as one of the main concerns of many personality psychologists. Burger (2011) summarises the study of personality psychology as a study that is concerned with people and their differences. This entails looking at each individual and his/her unique attributes and behaviours. Burger (2011) further states that personality psychology makes use of many different theories which are divided into different categories. These categories can be listed as: the psychoanalytical, trait, biological, humanistic, behavioural/ social learning and cognitive approaches. He also states that these six approaches complement and assist each other in understanding, defining and conceptualising personality, however, these approaches can be contradictory to each other in terms of competing accounts of behaviour. The assessment of these different categories can differ tremendously depending on the individual theory, psychological paradigm and between different countries.

Within South Africa many years have passed since the end of the Apartheid regime in 1994. South Africa became a multi-cultural society, investing in every unique group, race and language. It is, however, startling to notice that when it comes to psychological test development very few new culturally significant personality measurement instruments have been successfully developed (Foxcroft, 2004). According to Foxcroft (2004) test developers should give great consideration on how they develop their tests. Considerations such as presentation mode, item format and response mode are of utmost importance to ensure that certain sub-groups of test-takers aren’t advantaged or disadvantaged when taking the test. Some examples of test formats are pencil-and-paper testing or computer-based testing.

Consideration must be given as to what test format to apply when practising personality testing. Tests usually involve a form of paper-based examination where the test is constructed to cover some specific content within a specific time-period and can be arranged in either a forced-choice or free-response format (Holt, Kysilka & Kysilka, 2006). Paper-and pencil testing can be defined as the traditional testing method where test-takers are provided with paper-based test and asked to complete a response scale or other form of answering sheet by using a pencil-based method (Meade, Michels & Lautenschlager, 2007). Another method that has become more popular within personality testing is computer-based testing procedures.

According to Gregory (2000, as cited in Davies, Foxcroft, Griessel and Tredoux, 2005) computer-based testing has many advantages, such as more efficient designing of tests, using feedback on test performance to allow for individual tailoring of tests (based on individual responses), making use of multimedia technology and digital imagery to allow a more interactive and dynamic interface between the test and test-takers, allowing companies access to testing instruments via the Internet to assist in legal selection and placement procedures, more accurate scoring in a shorter time span, better interpretation of results based on different hypotheses and rules, and compiling test reports to assist in providing feedback to test-takers. Chen, Ankenmann and Chang (2000) also state that with the noteworthy increase of the use of computer technology, computer-based adaptive testing (CAT) has become more popular in testing environments. They also found that the main advantage of CAT was that it enabled more proficient and accurate trait estimation than pencil-and-paper based tests.

Personality assessment, measurement and testing

Holt, Kysilka, and Kysilka (2006) describe assessment as a formal process that is utilised by practitioners to determine students' progress in terms of learning outcomes and desirable learning objectives. It can further be explained as a "test", "measurement" or "evaluation". Therefore it is evident that assessment involves some kind of "scoring" process, where the individual or individuals are scored in terms of their answers and responses and rated on a set scale. It is therefore imperative that the assessment of personality be considered a vital part of individual assessment. The systematic measurement and observation of the many different

aspects of personality are called personality assessment. Some examples of personality assessment include: motivation (e.g., goal striving), personality pathology (e.g., depression), different personality traits (e.g., dominance, self-esteem and extroversion), personality dynamics (e.g., reaction to stress), and personality development (e.g., styles of aging) (Carducci, 2009).

Carducci (2009) also identifies the three major functions of personality assessment as: assisting in obtaining information about people in a manner that is accurate and significant, helping to convey this accumulated information in a constructive way to the individuals who utilize these personality tests such as personality psychologists, psychiatrists and personnel managers, and assisting in the prediction of future behaviour by utilizing the gathered information from the personality test results. He also states that personality assessment is used in many different settings such as the military, business, psychiatric hospitals and high schools to assist in placing individuals in the most suitable settings for their unique personalities. Jooste and Foxcroft (2005) state that, as with the assessment and measurement of an individual's interests and abilities, the measurement of personality can be seen as a significant part of career counselling when looking at the person-environment-fit within organisations. They further state that the aim of personality assessment is "to identify an individual's salient personality characteristics and to match these characteristics to the requirements of occupations." (p. 171). Morgenson et al. (2007) found that although people had thought for a long time that personality was not related to occupational or job performance, studies in the early 1990s showed differently and provided evidence that it does have an influence, and as a result is important when considering recruitment and selection procedures.

Personality psychologists have also become more interested in the culture of the individual whose personality is being measured and what influence this has on the results obtained (Burger, 2011). Jooste and Foxcroft (2005) found that psychologists commonly take assessment measures that where only developed in one culture and then translate and apply the same measurement to be used within another culture. Psychologists use these measurement instruments to make cross-cultural comparisons but they tend not to take into consideration that with structured assessment measures, such as personality assessments and inventories, the items and constructs that are

measured must have the same meaning in both of the different cultures and that no mistake as to the translation of the construct meanings should be possible (González et al. 2005).

The ethical issues of personality measurement must also be considered when using personality tests. Carducci (2009) summarises these issues as:

- *Personal concerns:* such as the invasion of privacy if the test-takers unknowingly reveal information about themselves. To increase confidentiality, access to and the distribution of test results must be regulated.
- *Legal concerns:* where a group of individuals are discriminated against due to the unfair impact of testing procedures and results. To ensure that personality tests do what they say they do, Equal Rights legislation must be put in place and more attention must be given to the development of personality tests to allow fair and ethical testing procedures, and
- *Social concerns:* these include unfair labelling of test-takers because of test results, unfair treatment of individuals because of labelling and the self-fulfilling prophecy in the behaviour of the labelled individuals.

Computer-based assessment, measurement and testing

In the 1980s Standing and Keays (1986), Mervielde (1988), and Waller and Reise (1989) documented some of the advantages and possible advantages of computer-based testing for the future. Mervielde (1988) stated that it was clear that computer-based testing held advantages such as not only using inventories, questionnaires and verbal responses, which may influence and guide the test-taker into certain responses, but using computer-based tests to obtain more information and more relevant testing practices.

Davies, Foxcroft, Griessel and Tredoux (2005) describe computer-based measurement as advantageous as it is helpful in the achievement of the highest levels of standardisation of assessment instructions as it limits confusion of test-takers when administering the test and avoids discrimination between test-takers, the computer administers and scores the personality measurement in an objective and logical manner and thus eliminates the probability of the potential biasing effect of the testing practitioner, the time it takes to complete the personality

measurement is reduced, it allow immediate scoring of the testing measurement and rapid feedback can be provided to the test-takers as well as to the testing practitioner, it creates the possibility of obtaining additional information about the test-takers by computing response time to each individual question or item, it allows better measurement of test-takers' spatial and perceptual abilities as computers have the ability to utilise graphic images and multi-media based items when constructing questions and the use of voice-activated and touch screen applications helps to assist test-takers who are physically and/or neurologically disabled whereas it might have been otherwise impossible to do with pencil-and-paper based measurements. This can be a step in the right direction when considering past discrimination against minority groups and people with disabilities. It can therefore be another way of assisting each and every person in South Africa with obtaining the best position available to their specific needs and abilities.

It is further stated by Davies, Foxcroft, Griessel and Tredoux (2005) that the effects of cheating on a measurement or test are minimised as the tests can be tailored to each individual and their individual answers, the testing practitioner is provided with more control of the testing procedure and measurement, for example the response time of a question or item can be limited and will only show on the computer screen for a limited time, the use of computer-based measurement is less labour-intensive and more cost-effective as fewer testing practitioners and assistants are needed during each testing activity as well as during the administration of the computer-based test, when scoring of the computer-based measurement takes place the possibility of errors and inaccurate scoring is lessened as the scoring is done by the computer and not by a testing practitioner who can make human errors, and test security is higher as tests and test-materials can not easily be taken out of the testing venue and be distributed to others.

According to Mills (2010) computer-based tests can assist in helping measurement professionals in understanding test-taker performances and abilities much better, allow for much better interpretation of test-taker responses by being able to copy and store test-taker responses better and for longer than when using pencil-and-paper tests, and if computer-based tests start to supplement or replace pencil-and-paper tests many expensive measurements can become available to more testing practitioners as they will be more affordable and economical. However, Mills (2010) also noted that as the use of computer-based testing measurements increases many

questions arise, such as access to computer labs where the computer-based tests are performed, the security of computer-based assessments, test-delivery system reliability, the psychometric quality of the computer-based tests, and theoretical model sufficiency that underline these computer-based tests.

Disadvantages of computer-based testing procedures and measurements as listed by Davies, Foxcroft, Griessel and Tredoux (2005) are illegal copying and distribution of computer-based personality tests when they are made available on the Internet, security concerns when computer-based personality tests are available on the Internet, confidentiality problems, when large quantities of information stored on the computers are accessed by unauthorised people, with interpretation of computer-generated assessment reports the practitioner still has to use his/her qualified professional experience to evaluate the results. Inexperienced testing practitioners may not know this and fail to adhere to the standards needed to provide correct reporting, weakly validated scoring routines and/or errors in programming may exist and are difficult to detect if they form part of the software programming, computer anxiety can exist in older test-takers or disadvantaged groups that aren't familiar with the computer-human interface and can have a negative effect on their test performance, and computer illiteracy can impact test-takers negatively as they aren't familiar with the manner in which computers operate. It is nonetheless clear that many of these listed disadvantages are due to the human-element in the computer-person relationship and can not be contributed only to the physical computer.

It is therefore important to consider the ethical issues, rules and regulations surrounding computer-based tests when using these to conduct personality assessment or any psychological measurements, so as not to discriminate against any test-taker or group. The International Test Commission (ITC) (2005) has certain regulations that any assessment practitioner should follow. These include ensuring that all test practitioners or persons using Internet-delivered or computer-based tests are themselves competent to do so, establishing the computer-based test's potential utility, using and applying only technically and psychometrically valid computer-based tests that have already been validated, evaluated and classified by the Psychometrics Committee of the Professional Board for Psychology, checking equivalence if the computer-based tests are applied when they have previously been in paper-and-pencil form, using a computer-based test taking

into consideration aspects such as fairness issues and human factors, providing an alternative paper-and-pencil based option to those test-takers who are not familiar with computer-based testing or have neurological or physical disabilities that make computer-based testing impossible, using tutorials or pre-tests to prepare test-takers, verifying the identity of the individual taking the test with Internet-delivered testing and providing appropriate support and supervision when administration of the computer-based test is done as to provide the needed support to the test-taker.

In South Africa it is important for testing practitioners to enhance their assessment abilities to be able to do the best, most valid and reliable data collection, especially with computer-based or Internet-delivered testing. Appropriate training in information technology, the development, considerations, advantages and disadvantages of Internet-delivered and computer-based tests, Internet-delivered and computer-based test practise standards, and gaining experience in using Internet-delivered testing, computer-based testing and computer-based test interpretation (CBTI) systems is very important (Davies, Foxcroft, Griessel & Tredoux, 2005). When developing and using computer-based tests it is also essential to discuss the response scale that is applied, whether dichotomous or polytomous.

Dichotomous and polytomous response scales

Dichotomous response scales use forced-choice items as the test-taker is forced to choose the one alternative over the other and usually involve “true/false”, “supported/not supported” or “cause/effect” items (Holt et al, 2006). Colton and Covert (2007) found that information gathered from using a dichotomous response scale is concise, unambiguous and easy to understand. Dichotomous response scales are also great for rating instruments where the test-takers have to indicate the presence or absence of a specific behaviour or action. An advantage of a dichotomous response scale is that it is efficient and clear, although care must be taken to ensure that bias is not present. Bias can exist when one option is emphasized over another.

Spector (1992) also stated that strategies have been developed in the past to try and counter bias in testing scenarios. Scales like forced-choice response formats, also known as dichotomous

response scales, have been utilised to try and prevent bias by eliminating social desirability from the equation. Negative consequences of using a dichotomous response scale, according to Hoijtink, Rooks and Wilmink (1999), include conflicting parameter estimates, incorrect standard errors of the estimates, and incorrect distributions for goodness-of-fit statistics.

Deutschlander (2009) states that dichotomous items are fairly easy to write, but that a "yes/no" response is not always enough to assist in gathering the needed data. When this is the case, adding more response options is necessary and when this happens it is no longer a dichotomous response scale but a multiple-response close-ended scale or rating scale. A polytomous response scale can be categorized as a type of rating scale. Deutschlander (2009) also describes a rating scale as a question that "requires the participant to assign a value to something, but unlike ranking-order scales, the participant s not required to compare an option with other options in another question." (p. 41). She further gives examples of polytomous response rating scales, as summarised in Table 1.

Table 1
Examples of polyotomous response scales

1	**2**	**3**	**4**
strongly disagree	disagree	neither agree nor disagree	agree
excellent	good	average	fair
very important	important	average importance	slightly important
extremely likely	likely	not sure	unlikely
1	**2**	**3**	**4**
strongly disagree	disagree	neither agree nor disagree	agree
excellent	good	average	fair
very important	important	average importance	slightly important
extremely likely	likely	not sure	unlikely
1	**2**	**3**	**4**

Deutschlander (2009) also states that polytomous rating scales are helpful when determining user satisfaction, determining frequency of an event or when asking test-takers to self-assess their own individual competencies or levels of knowledge and skill. It is also important to know that

when using a polytomous response scale instead of a dichotomous response scale, there should always be a middle or average response item that represents a neutral feeling or attitude (De Vaus, 2002; Holt et al, 2006).

The South African Personality Inventory (SAPI) project

The general aim of the SAPI[1] project is to provide theoretical insight into the universal and cultural specificity of personality and personality measurement as well as the practical development of a personality measuring instrument within South Africa that complies with the current legislation laws. A specific aim of the SAPI project is the development of a comprehensive personality questionnaire that includes all the 11 official language groups within South Africa. As a result this project covers all the major aspects of personality and personality testing reasoned significant within the South African milieu (Nel et al., in press).

As part of the SAPI project this article aims to identify the most reliable and valid response scale to be used in conjunction with the development of the SAPI questionnaire. The use of the chosen response scale will benefit the SAPI project in the sense that all versions were tested and considered and the best option was chosen.

RESEARCH OBJECTIVES

The research objectives of this study are divided into the general objective and the specific objectives.

General objective

The general objective of this research is to identify the optimal response scale, dichotomous or polytomous, to be used during computer-based personality measurement.

[1 1] The SAPI, an acronym for the South African Personality Inventory, is a project that aims to develop an indigenous personality measure for all eleven official languages in South Africa. Participants are Byron Adams (University of Johannesburg), Deon de Bruin (University of Johannesburg), Karina de Bruin (University of Johannesburg), Leon Jackson (North-West University), Carin Hill (University of Johannesburg), Deon Meiring (University of Pretoria and University of Stellenbosch), Alewyn Nel (North-West University), Ian Rothmann (North-West University), Michael Temane (North-West University), Velichko Valchev (Tilburg University, the Netherlands), and Fons van de Vijver (North-West University and Tilburg University, the Netherlands).

Specific objectives

The specific objectives of this research are:

- To conceptualise personality and computer-based measurement according to literature.
- To identify to what degree the response scale influences the reliability and validity of the personality test.
- To identify recommendations for future research and practice.

METHOD

Research design

A quantitative research approach is followed. It is a form of conclusive research and involves great representative samples as well as moderately structured procedures of data collection (Struwig & Stead, 2001). For the purpose of this study a cross-sectional design is employed which entails a study that takes place at a single point in time (Du Plooy, 2002). The data used consists of primary data and the approach is factor analysis and the examination of alpha coefficients.

Participants and procedure

The target population consisted of undergraduate and post-graduate University students (N = 724) from a South African tertiary institution. The participants differed in terms of age, race, language and gender. The sampling method that was used is convenience non-probability sampling since participants are easily accessible and available. The total participants were split into two groups where different types of response scales (dichotomous or polytomous scale) were administered. The characteristics of the participants can be viewed in Table 1.

Table 2

Characteristics of the participations for the computerized testing mode: dichotomous response scale (n = 380) and polytomous response scale (n = 344); (N = 724)

Item	Category	Frequency	Percentage	Frequency	Percentage
		Dichotomous		Polytomous	
Race	White	140	36.80	140	39.00
	Black	171	45.00	154	42.90
	Coloured	18	4.70	48	13.40
	Indian	51	13.40	17	4.70
	Missing Values	0	0	0	0
Gender	Male	88	23.20	78	21.70
	Female	292	76.80	281	78.30
	Missing Values	0	0	0	0
Age	17 – 19 years	63	16.60	61	17.10
	20 – 21 years	200	52.70	190	52.90
	22 – 23 years	85	22.40	77	21.40
	24 – 25 years	32	8.40	31	8.60
	Other	0	0	0	0
	Missing Values	0	0	0	0
First Language	Afrikaans	39	10.30	37	10.30
	English	184	48.40	178	49.60
	isiNdebele	4	1.10	4	1.10
	isiXhosa	16	4.20	12	3.30
	isiZulu	50	13.20	47	13.10
	Siswati	19	5.00	29	8.10
	Sepedi	18	4.70	17	4.70
	Sesotho	30	7.90	18	5.00
	Setswana	5	1.30	3	0.80
	Tshivenda	6	1.60	5	1.40
	Xitsonga	4	1.10	4	1.10
	Other	5	1.30	5	1.40
	Missing Values	0	0	0	0
Education	Grade 12	315	82.90	301	83.80
	Certificate	12	3.20	11	3.10
	Diploma	12	3.20	12	3.30
	Bachelors	37	9.70	34	9.50
	Honours	1	0.30	1	0.30
	Masters'	3	0.80	0	0
	Other	0	0	0	0
	Missing Values	0	0	0	0
English Proficiency	Very poor	2	0.50	2	0.60
	Poor	3	0.80	1	0.30
	Good	131	34.50	140	39.00
	Very good	244	64.20	216	60.20
	Missing Values	0	0	0	0

With the dichotomous response scale the total number of test-takers was *n* = 380, whereas with the polytomous response scale the total number of test-takers was *n* = 344, which totals the number of respondents at 724. Most of the test-takers were Black students with the dichotomous (45%) and with the polytomous (42,90%) response scale. The rest of the dichotomous version's respondents were White (36,8%), Indian (13,40%) and Coloured (4,70%). With the polytomous version the respondents were White (39,00%), Coloured (13,40%) and Indian (4,70%). There were no missing values. More than 75% of the test-takers were female, both with the dichotomous and polytomous response scale, at 76,80% and 78,30% respectively. The most prominent age group of the test-takers was between 20-21 years at 53,70% and 53,90%

respectively for the dichotomous and polytomous response scale versions. Predominantly most of the test-takers were English-speaking at 48,40% for the dichotomous version and 49,60% for the polytomous version. Regarding the educational levels, most of the participants had Grade 12 (82,90% of the dichotomous response scale; 83,80% of the polytomous response scale). Most of the test-takers rated their English proficiency as very good with the dichotomous response scale test-takers' "very good" rating being 64,20% of the test-takers and the polytomous response scale test-takers' "very good" rating being 60,20%. It is eminent from the collected characteristics of the participants that with the dichotomous response scale version and the polytomous response scale version the different characteristics of the test-takers were very close to being evenly divided and that there were no major differences in the make-up of the respondents between the two versions.

In order to collect the data from the participants, authorisation from the tertiary institution was obtained, thus allowing the researcher to use students to gather the data. The lecturers of the students were approached and asked for assistance in gathering the data. As agreed upon by their lecturers, students were approached during a class session. The students were asked to sign a consent letter allowing the use of their information and inputs in the study. It was made clear that they have certain rights including: anonymity, the right to withdraw at any time and confidentiality. During the taking of the test the appropriate instructions are given for the answering of the questionnaire. The questionnaire was taken electronically by means of computers provided by the universities. The students were required to login and complete the questionnaire.

Measuring instrument

An adapted version of the South African Personality Inventory (SAPI)'s Soft-Heartedness cluster is used in this study. Nine personality clusters were identified in the first phase of the SAPI project (refer to Nel et al., in press) namely: Extraversion, Soft-Heartedness, Conscientiousness, Emotional Stability, Intellect, Openness, Integrity, Relationship Harmony and Facilitating. It further consists of 37 sub-clusters and 191 facets. For the purpose of this study, which is to identify the optimal response scale to be used three facets from the Soft-Heartedness cluster,

namely *generosity*, *compassion* and *appreciation* were used (Nel et al., in press). *Generosity* can be defined as "Liking to spend money on others and sharing things with them.", *compassion* can be defined as: "Being concerned with the well-being of others and showing mercy. Being sensitive to the needs of others, and doing good for others. Being compassionate and empathetic.", and finally, *appreciation* can be defined as: "Expressing thankfulness and gratitude. Commending others on work well done. Enjoying, being grateful and satisfied with what you have". These three facets performed well during the initial pilot studies of the SAPI project (Flattery, 2010). A 24 item questionnaire was used to measure these facets.

The rationale for not using the complete instrument is that the optimal response scale for a computer-based personality test is the main concern and not how the facets, sub-clusters or clusters combine or coincide. Therefore only the three facets that performed in terms of their psychometric properties of the Soft-Heartedness cluster were utilised as a basic method of gathering data in terms of the response scales.

The dichotomous response scale is in the form of 0 = "disagree" and 1 = "agree", where the polytomous response scale consists of 1 = "strongly disagree", 2 = "disagree", 3 = "somewhat disagree/somewhat agree", 4 = "agree" and 5 = "strongly agree". The test booklets contain identical questions and items, only the response scale is changed between a dichotomous response scale and a polytomous response scale. Examples of statements that are answered under each sub-cluster includes: *generosity* - "I give things to people without expecting anything in return", *compassion* – "I feel sympathy for people who have problems", and *appreciation* – "I am happy when I see good things happening in other people's lives".

Statistical analysis

To identify the optimal response scale statistical analysis was conducted by means of the SPSS program (SPSS, 2008). Analysis of the data was done by descriptive statistics, factor analysis and Cronbach Alpha coefficient analysis. It determines which of the response scales have the highest functionality, validity, and reliability. Within descriptive statistics the data was analysed by looking at the distribution of the items (means, standard deviation, skewness and kurtosis)

with both the dichotomous and polytomous scale. With exploratory factor analysis the inter-item correlations, and item loadings on the respective facets (*generosity*, *compassion* and *appreciation*) were investigated, as well as the variance explained. The reliability was determined by identifying the Cronbach Alpha coefficients of each facet. A cut-off point of 0,70 for the Cronbach Alpha was utilised (Nunnally & Bernstein, 1994).

RESULTS

In this section the results of this immediate study will be reported. Firstly the results of the dichotomous response scale will be reported followed by the results of the polytomous response scale. Descriptive statistics of the items will firstly be provided, then the exploratory factor analysis and lastly the reliability of the respective constructs.

Results for the dichotomous scale

Table 3
Descriptive Statistics of Results of Computer-based Dichotomous Response Scale (*n*=380)

Item		Mean	*SD*	Skewness	Kurtosis
i001	**I share what I have with others**	0.94	0.23	**-3.91**	**13.34**
i002	**I give to people without expecting anything in return**	0.89	0.32	-2.45	**4.03**
i003	**I share my knowledge with others**	0.98	0.15	**-6.29**	**37.76**
i004	I buy things for others	0.86	0.35	-2.09	2.38
i005	I give money to the poor	0.71	0.45	-0.95	-1.11
i006	I treat others with gifts	0.69	0.46	-0.82	-1.33
i007	I give food to people who do not have any	0.80	0.40	-1.51	0.27
i008	I provide for those in need	0.80	0.40	-1.53	0.33
i009	**I am sensitive to others' needs**	0.95	0.22	**-4.02**	**14.26**
i010	**I feel sympathy for others who have problems**	0.96	0.19	**-4.94**	**22.49**
i011	**I can share in someone's emotions**	0.91	0.29	-2.89	**6.37**

i012	When someone cries I also feel like crying	0.63	0.48	-0.56	-1.70
i013	**I get sad when I see someone suffering**	0.95	0.22	**-4.02**	**14.21**
i014	**I get sad when someone I care about is sad**	0.97	0.17	**-5.64**	**29.98**
i015	I feel other people's problems as my problems	0.48	0.50	0.07	-2.01
i016	**I am sensitive to other people's feelings**	0.94	0.23	**-3.91**	**13.34**
i017	**I like pleasant things**	0.99	0.07	**-13.73**	**187.48**
i018	**I value the company of people close to me**	1.00	0.05	**-19.49**	**380.00**
i019	**I enjoy delicious food**	0.99	0.11	**-8.58**	**71.97**
i020	**I value life as it is**	0.91	0.29	-2.88	**6.35**
i021	**I am happy when I see good things happening in other people's lives**	0.98	0.14	**-6.70**	**43.10**
i022	**I value the little things in life**	0.95	0.22	**-4.15**	**15.27**
i023	**I value pleasant experiences**	1.00	0.05	**-19.47**	**379.00**
i024	**I value other people's work**	0.94	0.23	**-3.91**	**13.34**

Table 3 includes the means, standard deviations, skewness and kurtosis of the 24 items used. When taking into consideration each of the items' skewness and kurtosis values, inspection of Table 3 shows that only a few items were normally distributed. These items were i004 ("I buy things for others"), i005 ("I give money to the poor"), i006 ("I treat others with gifts"), i007 ("I give food to people who do not have any"), i008 ("I provide for those in need") as well as items i012 ("When someone cries I also feel like crying") and i015 ("I feel other people's problems as my problems"). Items i002 ("I give to people without expecting anything in return"), i011 ("I can share in someone's emotions") and i020 ("I value life as it is") showed a skewness < 2,00 and a kurtosis that was >4,00 and will be included in future analysis. The remaining items had values that were >2,00 on skewness as well as >4,00 kurtosis scale. As a result these were excluded from further data analysis. Only nine items of the initial 24 items were retained after this process.When looking at the mean scores of the dichotomous response scale it was found that it had an average mean of 0,84 which indicates that with most items the test-takers had a tendency

to answer towards the “agree” or “1” option. This means that most of the items had positive responses which had a mean of 0,90 or higher and the respondents answered “Agreed” most of the time.

Table 4
Eigenvalues of the Sample Correlation Matrix of Dichotomous Response Scale (*n* = 380)

Component	Eigenvalue	% of variance	Cumulative % variance explained
1	**2.27**	**22.69**	**22.69**
2	**1.42**	**14.23**	**36.93**
3	**1.09**	**10.94**	**47.87**
4	**1.01**	**10.07**	**57.94**
5	0.92	9.22	67.16
6	0.82	8.15	75.31
7	0.76	7.59	82.90
8	0.61	6.08	88.97
9	0.59	5.91	94.88
10	0.51	5.12	100.00

Table 4 shows that 4 eigenvalues were >1. As a result it is proposed that four factors be extracted. The amount of variance explained is also indicated by Table 3. The first four constructs accounted for more than half of the variance explained which is 57,94%. Since only three facets’ items were included in the study and not four, it shows that the data explains more facets that what was originally indicated. When three factors are extracted, it will explain 47,87% of variance.

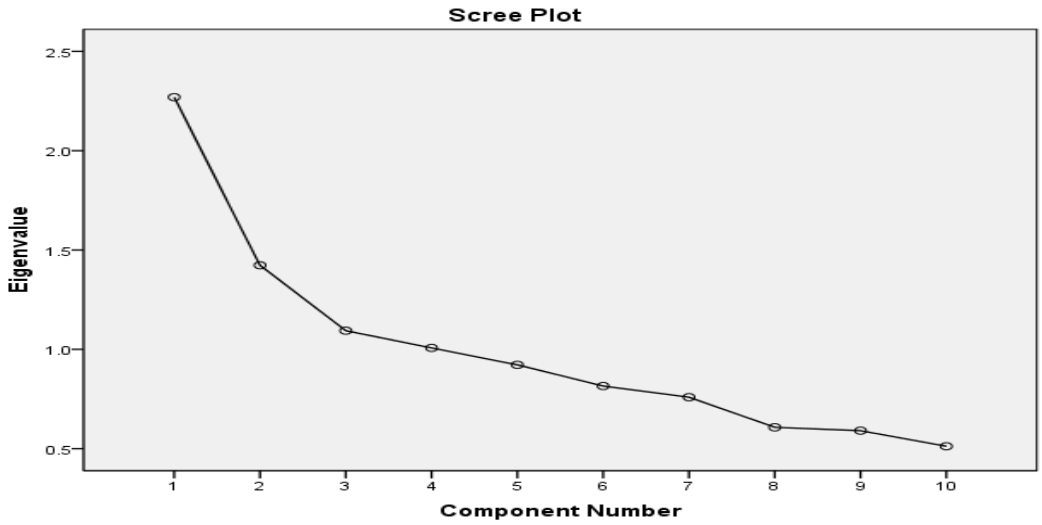

Figure 1: Scree Plot of eigenvalues of the computer-based dichotomous response scale.

Figure 1 suggests that three factors be extracted as the dent is made on the third factor. These three factors account for more variance explained than the remaining factors. This finding differs from the results portrayed in Table 3, where four factors were extracted. However, three factors will be utilised since only three facets were measured by the questionnaire, and as this is theoretically supported (DeKoster, 1998).

Table 5
Factor Loadings and Communalities of the Dichotomous Response Scale (*n*=380)

Item	Factor loading	h^2
i002		0.08
i004	0.60	0.36
i005	0.60	0.36
i006	0.55	0.30
i007	0.52	0.27
i008	0.68	0.46
i011	0.42	0.17
i012	0.32	0.10
i015	0.31	0.10
i020		0.08

According to Table 5 communalities (h^2) ranged from 0,08 to 0,46 with an average of 0,23. All items had communalities (h^2) that were > 0,20 except for items i002, i011, i012, i015 and i020. Factor loadings on the overall cluster of Soft-Heartedness ranged from 0,31 to 0,68 resulting in an average factor loading of 0,50. All items, excluding items i002 ("I give to people without expecting anything in return") and i020 ("I value life as it is") had factor loadings of >0,30. As a result these two items will not be used in further analysis. From the nine items that showed adequate distribution from Table 3, only seven items were retained after this process.

When considering the dichotomous response scale it was found that only one general factor could be extracted due to the other items having inconsistent and/or low loadings and values or having high skewness and/or kurtosis values. The Cronbach Alpha coefficient of this factor was $\alpha = 0,58$ which was lower than the accepted Cronbach Alpha value of $\alpha = 0,70$ (Nunnally & Bernstein, 1994). This means that the dichotomous response scale has a low internal consistency and is generally seen as an unreliable version of a response scale.

Results of the polytomous scale

Table 6
Descriptive Statistics of Results of Computer-based Polytomous Response Scale (*n*=344)

Item		Mean	*SD*	Skewness	Kurtosis
i001	I share what I have with others	3.98	0.70	-0.77	2.17
i002	I give to people without expecting anything in return	3.94	0.78	-0.57	0.77
i003	I share my knowledge with others	4.16	0.61	-0.17	-0.10
i004	I buy things for others	3.82	0.77	-0.52	0.56
i005	I give money to the poor	3.46	0.90	-0.41	0.24
i006	I treat others with gifts	3.58	0.86	-0.47	0.13
i007	I give food to people who do not have any	3.73	0.82	-0.38	0.03
i008	I provide for those in need	3.70	0.76	-0.21	-0.03
i009	I am sensitive to others' needs	4.18	0.66	-0.40	-0.05
i010	I feel sympathy for others who have problems	4.28	0.69	-0.63	0.08
i011	I can share in someone's emotions	4.12	0.76	-0.85	1.30
i012	When someone cries I also feel like crying	3.49	1.08	-0.34	-0.58
i013	I get sad when I see someone suffering	4.24	0.71	-0.81	0.90
i014	I get sad when someone I care about is sad	4.47	0.65	-1.08	1.14
i015	I feel other people's problems as my problems	3.28	0.99	-0.12	-0.44
i016	I am sensitive to other people's feelings	4.13	0.65	-0.26	-0.20
i017	I like pleasant things	4.58	0.54	-0.86	0.29
i018	I value the company of people close to me	4.68	0.49	-0.99	-0.47
i019	I enjoy delicious food	4.74	0.50	-1.88	3.55
i020	I value life as it is	4.25	0.73	-0.68	0.32
i021	I am happy when I see good things happening in other people's lives	4.33	0.63	-0.45	-0.31
i022	I value the little things in life	4.39	0.67	-0.80	0.27
i023	I value pleasant experiences	4.60	0.50	-0.56	-1.36
i024	I value other people's work	4.12	0.61	-0.15	-0.10

Table 6 represents the means, standard deviations, skewness and kurtosis of the 24 items used. On closer inspection of the items' skewness and kurtosis values it was clear that there were no items whose skewness was > 2,00 and kurtosis was >4,00. These items are all of the listed items ranging from i001 to i024, which means that all the items used were normally distributed and

will be included in further analysis. The polytomous response scale had an average mean of 4,09 which means that most of the test-takers had a tendency to answer towards the "agree" or "4" option. There was a higher number of positive responses, Strongly Agree or Agree being the choices made most often. This means that the respondents might have seen the items as socially desirable.

Table 7
Eigenvalues of the Sample Correlation Matrix of Polytomous Response Scale (*n* = 344)

Component	Eigenvalue	% of variance	Cumulative % variance explained
1	**7.41**	**30.87**	**30.87**
2	**2.35**	**9.78**	**40.65**
3	**1.84**	**7.65**	**48.30**
4	**1.38**	**5.75**	**54.05**
5	**1.16**	**4.83**	**58.88**
6	**1.02**	**4.27**	**63.15**
7	0.78	3.25	66.40
8	0.74	3.10	69.50
9	0.69	2.89	72.39
10	0.65	2.72	75.11
11	0.63	2.62	77.73
12	0.60	2.51	80.24
13	0.56	2.32	82.55
14	0.53	2.19	84.75
15	0.50	2.09	86.84
16	0.46	1.89	88.73
17	0.43	1.80	90.53
18	0.42	1.73	92.26
19	0.36	1.49	93.75
20	0.34	1.42	95.17
21	0.33	1.39	96.56
22	0.31	1.27	97.84
23	0.28	1.17	99.00
24	0.24	1.00	100.00

According to Table 7 there were 6 eigenvalues that were >1, which results in six factors being extracted. Table 7 also indicates the amount of variance explained. The six factors that were extracted accounted for 63,15% of the variance explained. In the measuring instrument, only three facets were measured which deviate from this result. As can be seen from Table 7, with three factors 48,30% of the variance is explained.

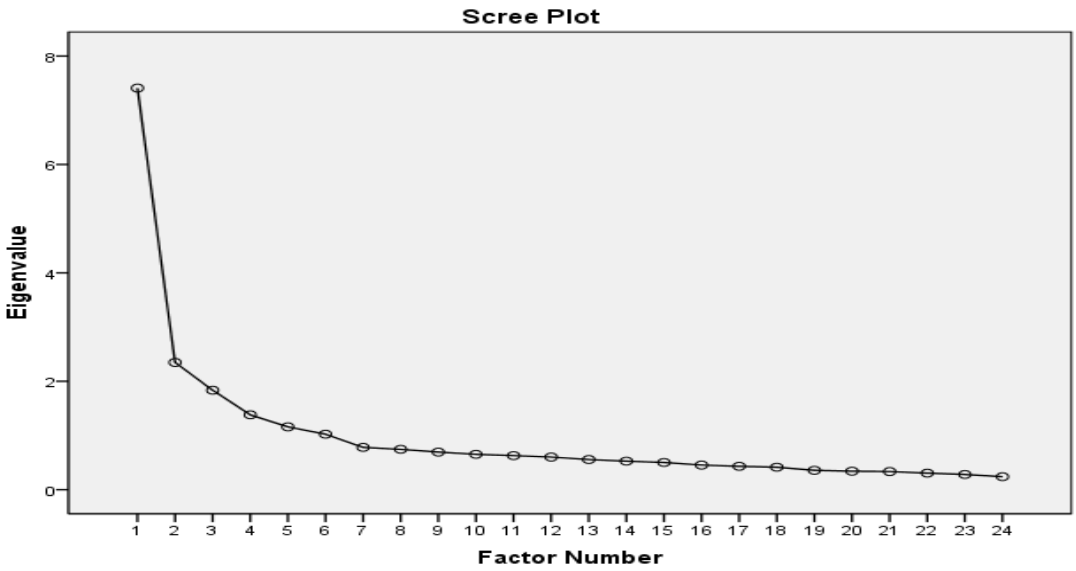

Figure 2: Scree Plot of eigenvalues of the computer-based polytomous response scale.

Figure 2 suggests that two factors be extracted as the dent is made on the second factor. These two factors account for more variance explained than the rest of the factors. Theory suggests three factors (as three facets were included from the Soft-Heartedness cluster) to be extracted which deviate from what was found in both Table 6 and Figure 2. The optimal would have been for the results to be consistent, as indicated by DeKoster (1998).

Table 8
Factor Loadings and Communalities of the Polytomous Response Scale (*n* = 344)

Item	Factor loading	h^2
i001	0.44	0.26
i002	0.37	0.22
i003	0.42	0.27
i004	0.59	0.32

i005	0.75	0.49
i006	0.45	0.24
i007	0.72	0.52
i008	0.73	0.56
i009	-0.55	0.51
i010	-0.69	0.60
i011	-0.62	0.47
i012	-0.62	0.37
i013	-0.56	0.50
i014	-0.51	0.41
i015	-0.68	0.40
i016	-0.67	0.58
i017	-0.63	0.48
i018	-0.67	0.47
i019	-0.61	0.33
i020	-0.38	0.23
i021	-0.37	0.36
i022	-0.57	0.42
i023	-0.69	0.55
i024		0.29

According to Table 8 the communalities (h^2) ranged from values 0,22 to 0,60 with a general average of 0,41. All of the items' communalities (h^2) were > 0,20. Only the factor loadings on the overall cluster of Soft-Heartedness was explored, and the loadings of the items ranged from 0,37 to 0,75 with an average factor loading of 0,58. All the listed items, except item i024 ("I value other people's work"), had factor loadings > 0,30. This item will be excluded from further analysis. After this process, 23 items were retained from the 24 items, which still yielded better results as observed from the dichotomous scale findings.

The alpha coefficients of the all three facets were further explored. These three factors had Cronbach Alpha coefficients that were higher than general guideline of $\alpha > 0,70$. These three factors are represented by three facets namely the *generosity* facet, the *compassion* facet and the *appreciation* facet. The *generosity* facet had an alpha coefficient of 0,80, the *compassion* facet had an alpha coefficient of 0,85 and the *appreciation* facet had an alpha coefficient of 0,80,

which means that these three individual factors confidently measure the Soft-Heartedness cluster as a whole.

DISCUSSION

The general objective of this study was to determine which optimal response scale to use in computerised personality testing, namely dichotomous or polytomous. These results will be crucial to the development of the South African Personality Inventory (SAPI) questionnaire in order to determine which response scale is optimal for administration. This discussion will firstly focus on the results of the dichotomous scale followed by a discussion of the polytomous scale. A conclusion will also be made at the end of the discussion.

In reviewing the descriptive statistics of the dichotomous scale, 58% of the items had a skewness score that was higher than the accepted score of 2 and 71% of the items had a kurtosis score that was higher than the accepted score of 4. As a result the following items were removed from the equation: i001 “I share what I have with others”, i002 “I give to people without expecting anything in return”, i003 “I share my knowledge with others", i009 “I am sensitive to others’ needs”, i010 “I feel sympathy for others who have problems”, i011 “I can share in someone’s emotions”, i013 “I get sad when I see someone suffering”, i014 “I get sad when someone I care about is sad”, i016 “I am sensitive to other people’s feelings”, i017 “I like pleasant things”, i018 “I value the company of people close to me”, i019 “I enjoy delicious food”, i020 “I value life as it is”, i021 “I am happy when I see good things happening in other people’s lives”, i022 “I value the little things in life”, i023 “I value pleasant experiences”, and i024 “I value other people’s work”. As a result very few of these items were normally distributed. A reason for all of these items having such high skewness and/or kurtosis values can be due to the fact that test-takers may have answered randomly or may not have wanted to answer differently due to the impression a different answer might make.

A common cause of high skewness and/or kurtosis values can be due to bias in the sample selection procedure which favours certain groups, such as the higher number of females that were used in obtaining the results in this study (Thornley & France, 2010). This statement is

supported by findings from Colton and Covert (2007) who also found that bias had a large influence on test-taker responses. When considering the mean scores of the dichotomous response scale it was found that it had an average mean of 0,84 which shows that the test-takers tended to answer similarly on the same items, mainly the "agree" or "1" option. This means that most of the items had positive responses which had a mean of 0,90 or higher and the respondents answered "Agreed" most of the time. This can mean that the respondents might have seen the items as socially desirable. The tendency to answer towards a specific response can be caused by cultural differences in test-takers, the individual perceived meanings of items in the test by the test takers or the test-takers' social concerns (Carducci, 2009; González et al., 2005).

When reviewing the exploratory factor analysis results, it was decided that with the dichotomous response scale three factors can be extracted, explaining only 47,87% of variance. Initially the eigenvalue showed that four factors can be extracted, accounting for 57,94% of the variance explained, which means that the remaining items accounted for and represented 57,94% of the utilised data. However, it was decided that due to a large number of extractions of items and low inter-item correlation values that only one general factor be utilised as it were representative of all the items remaining (Field, 2000).

The dichotomous response scale's communality values were all higher than the accepted 0,20 cut-off point except for items: i002 ("I give to people without expecting anything in return"), i011 ("I can share in someone's emotions"), i012 ("When someone cries I also feel like crying"), i015 ("I feel other people's problems as my problems"), and i020 ("I value life as it is"). This means that these items have no inter-item correlation with the other items listed and do not measure the general cluster of Soft-Heartedness. Items i002 ("I give to people without expecting anything in return") and i020 ("I value life as it is") had factor loadings < 0,30 and were excluded from further analysis. Items that did have factor loadings higher than those of 0,30 where items: i004 ("I buy things for others"), i005 ("I give money to the poor") and i008 ("I provide for those in need"). This means that very few of the items remained and that the validity of this response scale's items can be questioned due to the limited acceptable items. It can also be concluded that when considering the inter-item correlations there were no significant

correlations between these items and the other items listed, this was also experienced in studies by Hoijtink, Rooks and Wilmink (1999) when they utilised dichotomous response scales.

The Cronbach Alpha coefficient of the one general factor of the dichotomous response scale did not adhere to the general guideline of $\alpha > 0{,}70$ (Nunnally & Bernstein, 1994). The Cronbach Alpha coefficient of the dichotomous response scale was 0,58. Which, in turn, means that this scale cannot be considered a reliable response scale and cannot be said with confidence measures what it is intended to measure, namely the cluster of Soft-Heartedness.

In reassessing the above results it can be deduced that test-takers tended to answer similarly and in a specific manner on most items listed, that most test-takers had an inclination to answer towards the "agree" option, that only one general factor be extracted as this is representative of the remaining items and that this version of a response scale had low inter-item correlations and low internal consistency and cannot be considered a reliable predictor of the constructs being measured. Deutschlander (2009) also supports this conclusion in her statement that a dichotomous or forced-choice response format is not always enough to assist in gathering the needed data and cannot always discriminate between the needed factors or identify the needed data to get the wanted results.

When investigating the results of the polytomous response scale, the results appear more favourable. All the items of the polytomous response scale were normally distributed as all the items had a skewness value < 2 and a kurtosis value <4. This means that test-takers answered consistently and in a non-random manner when faced with the questions (Thornley & France, 2010). None of the items were eliminated at this point. When considering the average mean score (4,09) of the polytomous response scale it can be assumed that most of the test-takers had an inclination to answer towards the "agree" or "4" option. This means that most of the items had positive responses which had a mean of 4 -5 and the respondents answered "Agree" or "Strongly Agree" most of the time. This can mean that the respondents might have seen the items as socially desirable.

Exploratory factor analysis indicated that six factors be extracted from the data as 63,15% of the variance is explained herewith. However, it was decided that only three main factors be extracted as these account for 48,30% of the variance explained (DeKoster, 1998). It can therefore be said that the remaining items account for and represent 48,30% of the measured data (Brace, Kemp & Snelgar, 2006). These three main factors account for and respresent 48,90% of the data obtained and can be used with confidence in measuring the main cluster, namely Soft-Heartedness.

When considering the communalities of the data collected from the polytomous response scale it was found that all the items had communality values higher than the accepted cut-off point of 0,20 and was included for further analysis. However, i024 ("I value other people's work") did not adhere to this standard and was excluded from further analysis. Thus, this item's inter-correlation was low and did not show any association with the other items listed and did not have a loading on any of the other items or with the general Soft-Heartedness cluster. As a result it can be said that most of the items listed were well-founded and can be seen as valid items. The other items adhered to this standard. On further inspection of the data three general factors were extracted. These three factors' Cronbach Alpha coefficients adhered to the general guideline of $\alpha > 0,70$. The *generosity* facet had an alpha of 0,80, the *compassion* facet had an alpha of 0,85 and the *appreciation* facet had an alpha of 0,80, which means that these three individual factors confidently measure the intended Soft-Heartedness cluster as originally intended. These impressive reliability coefficients can be seen as satisfactory when considering their usage, namely measuring the main construct (Thornley & France, 2010).

When reassessing the above results of the polytomous response scale it can be concluded that test-takers answered in a consistent and non-random manner and tended to answer towards the "agree" or "4" option and three main factors were extracted which represent the *generosity*, *compassion* and *appreciation* facets of the Soft-Heartedness cluster. This response scale had high inter-item correlations and high internal consistency and can be considered a reliable predictor of the cluster being measured.

From the above-mentioned results the general objective can be discussed. The general objective was to determine which response scale (dichotomous or polytomous) was the most optimal when

using a computer-based personality measurement. It can be concluded that all of the items from the polytomous response scale were normally distributed and very few of the items from the dichotomous response scale had normal distribution scores, which means that the polytomous response scale tends to counter random answering and can be seen as the optimal response scale, unlike the dichotomous response scale. This is also evident from previous studies (Bijttebier, Vertommen & Vander Steene, 2001; Litzinger, Lee, Wise & Felder, 2007). The polytomous response scale had a much higher validity than that of the dichotomous response scale when considering the average factor loadings of each version, namely 0,58 and 0,50 respectively. The polytomous response scale can also be seen as the more advanced response scale, due to the polytomous response scale's average communality value of $h^2 = 0,41$ being much higher than the dichotomous response scale's average communality value of $h^2 = 0,23$. The average Cronbach Alpha coefficients of each response scale, dichotomous and polytomous, were $\alpha = 0,58$ and $\alpha = 0,88$, respectively. In review, there is no assurance that the dichotomous response scale is an adequate predictor of the Soft-Heartedness cluster it was supposed to measure. Only the polytomous response scale had a Cronbach Alpha coefficient that was acceptable as $\alpha > 0,70$ (Nunnally & Bernstein, 1994). As a result is can be said that the polytomous response scale is the more valid and reliable response scale. This means that it can be concluded with confidence that the polytomous response scale is a reliable and consistent measure, with three facets, namely *generosity*, *compassion* and *appreciation*, that it is an adequate measurement of the cluster Soft-Heartedness.

LIMITATIONS

This study did not include all representative individuals of the population and only made use of students from tertiary institutions. The results can therefore not be generalised to the whole public and is limited to students of tertiary institutions. The different cultures of the test-takers were also not taken into consideration and can influence the validity and reliability of the results obtained (Burger, 2011). Computer anxiety and computer familiarity were not determined before computer-based testing took place and as a result some individuals may have been negatively influenced or discriminated against (Foxcroft, Watson & Seymour, 2004; Mills, 2010).

RECOMMENDATIONS

In accordance with the previously listed limitations some recommendations for future research and practice can involve the use of the test procedure on a wider variety of population groups and the in-depth study and identification of the different individual cultures of each test-taker and what influence this can have on the testing procedure and test results. It is also important that when duplicating this study in the future test-takers be familiarised with computer-based testing procedures and that a practice course in computer-based testing be provided. To ensure that there are no miss-interpretations and human errors when administering and reporting of the test, test practitioners should also be familiarised with computer-based testing procedures. In future the Item Response Theory can be utilised to identify the optimal scale.

REFERENCE LIST

Bartram, D. (2006). Measurement on the internet: Issues, challenges and opportunities in the field of occupational assessment. In D. Bartram, & R. K. Hambleton (Eds.), *Computer-based measurement and the internet: Issues and advances* (pp. 13-38). Thames Ditton, UK: John Wiley & Sons, Ltd.

Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologist: A guide to data analysis using SPSS for Windows* (3rd ed.). London, UK: Palgrave Macmillan.

Bull, J. (1999). Computer-assisted assessment: Impact on higher education institutions. *Educational Technology & Society, 2*(3), 123-126.

Burger, J. M. (2011). (2011). *Personality* (8th ed.). Wadsworth: Cengage Learning.

Carducci, B. J. (2009). *The psychology of personality: Viewpoints, research and applications* (2nded.) West Sussex, UK: John Wiley & Sons Ltd.

Chen, S., Ankenmann, R. D., & Chang, H. (2000). *A comparison of item-selection rules at the early stages of computerized adaptive testing, 24*(3), 241-255.

Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation* (1st ed.). San Francisco, CA: John Wiley & Sons, Inc.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407-422.

Davies, C., Foxcroft, C., Griessel, L., & Tredoux, N. (2005). Computer-based and internet-delivered assessment. In C. Foxcroft, & G. Roodt (Eds.), *Introduction to psychological assessment in the South African context* (2nd ed.). Cape Town, South Africa: Oxford University Press.

De Vaus, D. A. (2002). *Surveys in social research* (5th ed.). St. Leonards, NSW: Allen & Unwin.

DeKoster, J. (1998). *Overview of factor analysis.* Retrieved November, 10, 2011 from http://www.stat-help.com/notes.html.

Deutschlander, C. (2009). *Surveying fundamentals for business analysts.* Vienna, VA: Management Concepts, Inc.

Du Plooy, G. M. (2002). *Communication research: techniques, methods and applications.* Lansdowne: Juta Publishers.

Engler, B. (2009). *Personality theories: An introduction* (8th ed.). Boston, MA: Houghton Mifflin Harcourt Publishing Company.

Field, A. P. (2000). *Discovering statistics using SPSS for Windows.* London, UK: Sage Publications Ltd.

Flattery, A. (2010). *Developing and validating a hostility, gratefulness and active support measuring instrument*. Unpublished masters' dissertation, North West University, Potchefstroom, South Africa.

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

Goffin, R. D., Jang, I., & Skinner, E. (2011). Forced choice and conventional personality assessment: Each may have unique value in pre-employment testing. *Personality and Individual Differences, 51,* 840-844.

González, N., Quintana, J. M., Aróstegui, I., Padierna, A., Martinez, E., Crespo, I., & Vesga, M. A. (2005). Translation and psychometric testing of the Basque version of the SF-36 health survey. *Qual Life Survey, 14,* 549-554.

Gregory, R. J. (2000). *Psychological testing: History, principals, and applications* (3rd ed.) Boston: Allyn & Bacon. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context(2nd ed.),* (pp 153-166) Cape Town: Oxford University Press.

Hoijtink, H., Rooks, G., & Wilmink, F. W. (1999). Confirmatory factor analysis of items with a dichotomous response format using the multidimensional Rasch model. *Psychological methods, 4*(3), 300-314.

Holt, L. C., Kysilka, M., & Kysilka, M. L. (2006). *Instructional patterns: Strategies for maximizing student learning.* Thousand Oaks, CA: SAGE Publications, Inc.

International Test Commission, ITC. (2005).*Guidelines on Computer-based and Internet Delivered Measurement*: ITC.

Jooste, M., & Foxcroft, C. (2005). Measures of affective behaviour, adjustment, and well-being. In Foxcroft, C., & Roodt, G. (Eds.), *Introduction to psychological assessment in the South African context* (2nd ed.). Cape Town, South Africa: Oxford University Press.

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly compatible?. *Organizational Research Methods, 10*(2), 322-345.

Mervielde, I. (1988). Cognitive processes and computerized personality assessment. *European journal of personality, 2,* 97-111.

Mills, C. N. (2010). Computer-based testing: Building the foundation for future assessments. Mahwah, NJ: *Lawrence Erlbaum Associates, Inc.*

Morgenson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729.

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

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

Spector, P. E., (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA:SAGE Publications, Inc.

SPSS Inc. (2008). *SPSS 16.0 for Windows*. Chicago, IL: SPSS Inc.

Standing, L., & Keays, G. (1986). Computer assessment of personality: Demonstration of gullibility. *Social behavior and personality, 14*(2), 197-202.

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

Thornley, J. H. M., & France, J. (2010). *Mathematical models in agriculture: Quantitative measures for the plant.* Oxfordshire, UK: Clive R, Ireland.

Van der Linden, D., Bakker, A. B., & Serlie, A. W. (2011). The general factor of personality in selection and assessment samples. *Personality and Individual Differences, 51,* 641-645.

Waller, N. G., & Reise, S. P. (1989). Computerized adaptive personality assessment: An illustration with the absorption scale. *Journal of personality and social psychology, 57*(6), 1051-1058.

CHAPTER 3

CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

This chapter contains information regarding the previously stated general and specific objectives of this study. It also contains reflections on the study's limitations as well as recommendations for future general use and use within the South African Personality Inventory (SAPI).

3.1 CONCLUSION

In this section the results will be discussed in accordance to the set specific objectives as provided in Chapters 1 and 2.

According to the first research objective, the aim was to determine how personality and computer-based measurements are conceptualised according to previous literature. Personality measurements have been used widely throughout many different fields. Carducci (2009:40) defines personality measurement as "the systematic measurement of the many aspects of personality". According to Goffin, Jang and Skinner (2011) personality measurement is the use of personality tests to help assist in the selection of individuals to be placed within a specific context or position. It includes the assessment of behaviour, statements and repetitive actions.

Personality measurement is foremost a testing procedure used for assessment purposes and selection procedures (Van der Linden, Bakker & Serlie, 2011). Methods of personality assessment are firstly, objective self-report techniques which measure personality by means of single-dimension and multiple-dimension assessments containing a standardized set of measurement and scoring procedures; secondly, behavioural techniques of assessment which observes behaviour, environmental cues and cognitive processes, and lastly, psycho-physiological measurement where bodily responses associated with stimuli are assessed (Carducci, 2009).

In South Africa personality measurement tools are also used to aid in selection, placement, determination of job satisfaction and development (Bester, 2008). Cheung (2004) conceptualises personality assessment as the instrument used in clinical settings to aid in diagnostic and treatment decisions. Examples of these personality measurements are the Minnesota Multiphasic Personality Inventory (MMPI) and the Eysenck Personality Questionnaire (EPQ). These are used to assess psychopathology within many different clinical and organisational psychology samples which range from psychiatric patients, drug users, juvenile delinquents and medical patients to areas such as adjustment difficulties of family members or students and personnel placement procedures. Burger (2011) defines personality psychology as a study that is concerned with the difference among individual people. Personality, on the other hand is defined as "consistent behaviour patterns and intrapersonal processes originating within the individual" (Burger, 2011:16).

The second part of the first objective was to conceptualise computer-based testing. A computer-based test refers to a test that is scored and administered by using a computer (Struwig & Stead, 2001). Computer-based assessments can be categorized into different kinds of assessments, which include stand-alone applications that use only one computer, private network applications, and online assessments where applications are provided across public networks (Wainer & Dorans, 2000).

More specifically these categories are: computer-based testing (CBT), which is when a paper-based format is copied in a digital format and administered via computer interface, computer-based adaptive testing (CAT), where a set of statistical procedures are utilized to make it possible for the assessor to tailor the test to respond to each individual's unique responses as they answer, computer-based test interpretation (CBTI) where the assessor is assisted in interpreting the data and the test-takers' test performance, and lastly, internet delivered testing (IDT) where computer-based tests are administered via the World Wide Web which allows for faster distribution of these tests to many different assessors (Bull, 1999; Bartram, 2006; Foxcroft & Roodt, 2009).

The second objective of this study was to conceptualise and to differentiate between a dichotomous response scale and a polytomous response scale. A dichotomous response scale can

be defined as a two-format or forced-choice, response scale, where a "yes/no" or "agree/disagree" answer is required (Cox, 1980; Goffin, Jang & Skinner, 2011). Dichotomous scales are seen as easy to apply and aid in avoiding fence-sitting by test takers. This response scale can also be considered as easy to understand, easy to react to and easy to answer as it provides a simple choice format. Goffin, Jang and Skinner (2011) further state that, when applied within personnel selection practices and fields, a forced-choice or dichotomous personality response scale has been found to predict criteria such as counterproductive work behaviour much better than normally used standardized conventional personality measures.

A polytomous response scale can be categorized as a Likert scale item, which means that it is a specific type of rating scale that is closed-ended and used to assess attitudes and opinions. Polytomous response items offer five different choices that a test-taker can choose from, an example is 1= strongly disagree, 2= disagree, 3= neither agree nor disagree, 4= agree, and 5= strongly agree (Deutschlander, 2009). According to De Vaus (2002) a rating scale "involves a set of responses where the alternative answers are ordered from low to high" (p. 102). This allows test-takers to choose between the higher and lower options in order to convey where their attitude lies. He further states that this type of rating scale requires that the test-taker should only give one single response to each item presented, variables are produced where the responses can be ordered from high to low, and the manner in which each item is answered is not affected or constrained by the manner in which the other items in the same set is answered.

The third objective of this study was to identify to what degree the type of response scale influenced the reliability and validity of the personality test. With the dichotomous response scale very few of the 24 items were normally distributed as is pertinent from the fact that 58% of the items had a skewness score that was higher than the accepted score of 2 and 71% of the items had a kurtosis score that was higher than the accepted score of 4. The items of the polytomous response scale, however, were all normally distributed as all of the items listed had skewness scores lower than 2 and kurtosis scores lower than 4. From this it can be concluded that the test-takers answered randomly and arbitrarily with the dichotomous response scale and consistently and regularly with the polytomous response scale.

When considering the mean scores of the two versions of response scales it was found that the dichotomous response scale had an average mean of 0,84 which means that with most items the test-takers tended to answer towards the “agree” or “1” option. With items i018 (“I value the company of people close to me”) and i023 (“I value pleasant experiences”) the mean scores were 1,00 and from this it can be concluded that the average participant answered the same when faced with this item. The polytomous response scale had an average mean of 4,09 which means that most of the test-takers tended to answer towards the “agree” or “5” option. As a result it is safe to say that the measuring instrument yielded the same manner of responding from the test-takers whether the dichotomous or polytomous response scale was used. The tendency to answer towards a specific response can be caused by cultural differences, perceptions and/or social concerns (Carducci, 2009; González et al., 2005).

The factor loadings of the polytomous response scale were much higher than those of the dichotomous response scale, ranging from 0,37 to 0,75 and from 0,31 to 0,68 respectively. As a result the average factor loading of the polytomous response scale (0,58) was higher than that of the dichotomous response scale (0,50), which means that the polytomous response scale had a higher validity than the dichotomous response scale. According to Hair, Black, Babin, Anderson and Tathum (2006) the accepted range of factor loadings is above ±0,30. Both of the response scale versions had acceptable scores although the polytomous version had obtained a higher score than the dichotomous version which means that the polytomous response scale yielded more valid results.

The polytomous response scale’s communality values were generally much higher than those of the dichotomous response scale, the highest communality value being h^2= 0,60 for i010 (“I feel sympathy for others who have problems”) and the lowest communality value being h^2= 0,22 for i022 (“I value the little things in life”), and the dichotomous response scale’s highest and lowest communalities being h^2= 0,46 for i008 (“I provide for those in need”) and h^2= 0,08 for i002 (“I give to people without expecting anything in return”) respectively. This means that the polytomous response scale version can be seen as the more reliable and advanced response scale.

According to Nunnally and Berstein (1994) the best cut-off point for determining the reliability of a measure is a Cronbach Alpha score of $\alpha > 0,70$. The dichotomous response scale generally had a lower reliability score than the polytomous response scale and only one general factor was identified; this factor had a Cronbach Alpha coefficient of $\alpha = 0,58$, thus the average Cronbach Alpha coefficient also had a score of $\alpha = 0,58$. The polytomous response scale had a higher average Cronbach Alpha coefficient than the dichotomous response scale, namely $\alpha = 0,88$. Three main factors were identified from the data, each with a reliability score > 0,70. These are 0,80 for the *generosity* facet, 0,85 for the *compassion* facet and 0,80 for the *appreciation* facet respectively. These three facets also reliably measure the overall Soft-Heartedness cluster it was intended to measure. It can thus safely be concluded that the polytomous response scale is the more reliable response scale to be used within computer-based personality measurement.

The last objective of this study was to give recommendations for future research and practice and these will be discussed in the next few paragraphs.

3.2 LIMITATIONS

This study did not primarily focus on the individual test-taker's culture and as a result the culture of the test-taker may have influenced the manner in which the items were answered. Burger (2011) states that personality psychologists have to become aware of the need to include the test-taker's culture into the equation, as most personality measurements are based on individualistic cultures and as a result can not be generalised to all cultural groups. He also states that the manner in which each individual responds to personality assessment must be taken into consideration, as this will also influence the results of the personality measurement. This study also did not include the whole of the general population, as only university students were used to obtain the data.

The way individuals and their responses are handled must also be taken into consideration as this can also influence the testing procedure. Mills (2010) states that most test legislation, test disclosure legislation and disclosure practices were developed for the paper-and-pencil testing context and not specifically for use in computer-based testing contexts. This can result in

unconscious discrimination against some of the test-taker groups that took part in this study as consideration was not given to all aspects of computerised or computer-based studies.

Computer familiarity might also be a problem when using computerised or computer-based personality assessments as this can influence the test-taker's attitude towards the testing procedure. Within South Africa the familiarity and exposure to computers are not evenly divided between all the diverse groupings and groups, and people from disadvantaged backgrounds may not be free from the negative impact that computer illiteracy has on their test performance (Foxcroft, Watson & Seymour, 2004, as cited in Foxcroft & Roodt, 2005).

Foxcroft and Roodt (2005) also state that it is not only test-takers who are influenced by computer illiteracy in South Africa but that test practitioners also experience a level of computer anxiety when they too are not familiar with the use of computer-based measurement procedures.

3.3 RECOMMENDATIONS

3.3.1 Recommendations for future research

By considering the previously stated limitations of this study the following recommendations can be made: Due to the small number of diverse participants it would be beneficial to replicate this study by including all groups of the population, which should be made up of all relevant ages, races, genders and cultures in order to get reliable and valid data which can be generalised over the whole of South Africa and all its diverse groups. Also, when conducting future research it would be beneficial to examine and apply appropriate legislation and regulations designed specifically for use within computer-based testing as not to discriminate against any party involved (Bartram, 2006).

Due to the fact that computer-based testing familiarity and test-anxiety can influence the testing procedure, as a result it would be best if, in future, test-takers be familiarised with computer-based testing procedures and receive a practice course in the use of computers and personality tests. It must also be seen to that test practitioners are educated and trained in the use of computer-based testing procedures so as to ensure that human error or a lack of knowledge does

not influence the testing process. The type of response format used with computer-based testing must also be considered in terms of reliability and validity.

3.3.2 Recommendations for the South African Personality Inventory (SAPI)-project

In the light of the results obtained in this study and the previously stated discussion, it should be recommended that computer-based personality assessment measurements best be accompanied by a polytomous response scale when higher validity and reliability are required. It would thus be recommended that the South African Personality Inventory (SAPI) questionnaire make use of a 5-point or polytomous response scale in the following format: 1 = “strongly disagree”, 2 = “disagree”, 3 = “somewhat disagree/somewhat agree”, 4 = “agree” and 5 = “strongly agree”, to yield better results than with the forced-choice or dichotomous response scale: 0 = “disagree” and 1 = “agree”.

REFERENCE LIST

Bartram, D. (2006). Measurement on the internet: Issues, challenges and opportunities in the field of occupational assessment. In Bartram, D., & Hambleton, R. K. (Eds.), *Computer-based measurement and the internet: Issues and advances* (pp. 13-38). UK: John Wiley & Sons, Ltd.

Bester, A. (2008). *The establishment of implicit perspectives of personality among Afrikaans speaking people in South Africa*. Unpublished masters' dissertation, North-West University, Potchefstroom, South Africa.

Bijttebier, P., Vertommen, H., & Vander Steene, G. (2001). Assessment of cognitive coping styles: A closer look at situation-response inventories. *Clinical Psychology Review, 21*(1), 85-201.

Bull, J. (1999). Computer-assisted assessment: Impact on higher education institutions. *Educational Technology & Society, 2*(3), 123-126.

Burger, J. M. (2011). (2011). *Personality* (8th ed.). Wadsworth: Cengage Learning.

Carducci, B. J. (2009). *The psychology of personality: Viewpoints, research and applications* (2nd ed.) West Sussex, UK: John Wiley & Sons Ltd.

Cheung, F. M. (2004). Use of western and indigenously developed personality tests in Asia. *Applied Psychology: An International Review, 53*(2), 173-191.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407-422.

De Vaus, D. A. (2002). *Surveys in social research* (5th ed.). St. Leonards, NSW: Allen & Unwin.

Deutschlander, C. (2009). *Surveying fundamentals for business analysts.* Vienna, VA: Management Concepts, Inc.

Foxcroft, C., & Roodt, G. (2005). *Introduction to psychological assessment in the South African context* (2nd ed.). 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.

Goffin, R. D., Jang, I., & Skinner, E. (2011). Forced choice and conventional personality assessment: Each may have unique value in pre-employment testing. *Personality and Individual Differences, 51,* 840-844.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tathum, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Litzinger, T. A., Lee, S. H., Wise, J. C., & Felder, R. M.(2007). A psychometric study of the index of learning styles. *Journal of Engineering Education, 96*(4), 309-319.

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

Struwig, F. W., & Stead, G. B. (2001). *Planning, designing and reporting research* (1st ed.) Cape Town, South Africa: Pearson Education South Africa.

Van der Linden, D., Bakker, A. B., & Serlie, A. W. (2011). The general factor of personality in selection and assessment samples. *Personality and Individual Differences, 51,* 641-645.

Wainer, H. & Dorans, N. J. (2000). *Computerized adaptive testing* (2nd ed.) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.