AN INVESTIGATION INTO THE FIRST-ORDER FACTOR STRUCTURE OF THE PERSONALITY AND PREFERENCE INVENTORY-NORMATIVE (PAPI-N) ON A RELATIVELY LARGE SOUTH AFRICAN SAMPLE

LIZELLE WILBERS

Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce at the University of Stellenbosch

Supervisor: Prof CC Theron

Marc 2015

DECLARATION

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

L Wilbers
Date: November 2014

ABSTRACT

Organisations in a free market economy exist with the purpose to serve and provide the market with products and services that the market values while at the same time satisfying the triple bottom line of profit, people and planet. The extent to which an organisation will succeed in this aim, however, depends to a large extent on the calibre of its workforce. Human resource management represents a range of interventions with the purpose of contributing to an organisation's success, through the acquisition and maintenance of a high quality and competent work force, as well as to ensure the effective and efficient use of human talent in a manner that will add value to an organisation. Personnel selection represents one of these human resource functions and thereby constitutes a critical human resource management intervention in as far as it attempts to regulate human capital movement into, through and out of the organisation with the expectation that this will result in increased employee job performance. Industrial-organisational psychologists and human resources practitioners frequently use psychometric/psychological tests in the selection process, which provide them with objective information on complex constructs such as intellectual ability or personality, that are hypothesised to be determinants of the level of job performance that selected applicants will achieve. Accurate predictions can however, only be derived from measures of such psychometric/psychological tests if the constructs they attempt to measure are in fact determinants of job performance, if the tests provide reliable, valid and unbiased measures of these constructs and the nature of the relationship between the predictor constructs and the criterion construct is validly understood. Personality represents an influential determinant of job performance. The Personality and Preference Inventory-Normative (PAPI-N) is a personality questionnaire that is widely used in industry. This provides the essential justification for the primary objective of this research, which was to evaluate the first-order factor structure of the PAPI-N through a factor analytic investigation on a relatively large sample of the South African working population.

The data used in this study was obtained from the data archives of Cubiks (Pty) Ltd, with written permission from the intellectual property holder, to utilise the sample data for the purpose of this research. The South African PAPI-N database comprised all respondents who were assessed by Work Dynamics, the official distributor of Cubiks's products and services in South Africa, in the period 2007 to 2012. Item and dimensionality analyses were performed on the 20 subscales of the PAPI-N as well as the Social Desirability scale. This was done to assess the success with which the subscales represented the underlying personality constructs. The results in the item analysis revealed that in about 50% of the PAPI-N subscales concern arose about the extent to which the items of the subscales responded in unison to systematic differences in a single underlying latent variable. Results from the dimensionality analysis showed that 12 of the 20 personality dimension measures were compatible with the position that the items comprising these subscales measure what they are designed to measure. In contrast, eight out of the 20 subscales failed the uni-dimensionality test.

A spectrum of goodness-of-fit statistics was used to evaluate the measurement model fit. The measurement model's overall fit was acceptable. The null hypothesis of exact fit was rejected but the null hypothesis of close fit could not be rejected ($p>.05$). The fit indices reflected a close fit in the parameter and a very good model fit in the sample. Although the measurement model fitted the data closely, the factor loadings (although statistically significant) were generally of a moderate degree. Approximately twenty-eight percent (27.78%) of the completely standardised factor loadings fell below the critical cut-off value of .50. This would suggest that the individual items generally (72.22%) do represent the latent personality dimensions they were designed to reflect acceptably well, but that in a little bit more than a quarter of the items, less than 25% of the variance in the item responses was due to variance in the latent variable it was designed to reflect. Discriminant validity was also investigated. The results showed that PAPI-N, although with some difficulty, permit the successful discrimination between the unique aspects of the latent personality dimensions.

The results of the confirmatory factor analyses suggests that while the intention of the PAPI-N to have sets of items reflecting specific primary personality factors succeeded, the subscale item measures mostly hold a sizable amount of systematic and random error. Based on the above findings, this personality measure should be used with caution in personnel selection in South Africa. Nevertheless, this study serves to extend the understanding of the psychometric properties of the PAPI-N on samples different from the UK sample on which it was originally developed and standardised. Its findings should assist in eliciting the necessary further research needed to establish the psychometric credentials of the PAPI-N as a valuable assessment instrument in South Africa with confidence. Recommendations for future research are made.

OPSOMMING

Organisasies in ‘n vrye-mark ekonomie het ten doel om die samelewing te dien en om die mark met produkte en dienste te voorsien wat waarde toevoeg, terwyl hulle terselfertyd die driedubbele eis van wins, mense en die planet bevredig. Die mate waarin die organisasie in hierdie doel slaag, hang egter in ‘n groot mate af van die kwaliteit van sy werksmag. Menslike hulpbronbestuur verteenwoordig 'n verskeidenheid ingrypings met die doel om by te dra tot 'n organisasie se sukses, deur die verkryging en instandhouding van 'n hoë gehalte en bekwame arbeidsmag, sowel as om die doeltreffende en doelmatige gebruik van menslike talent te verseker op 'n wyse wat waarde tot die organisasie toevoeg. Die keuring van personeel verteenwoordig een van hierdie menslike hulpbronfunksies. As sodanig vorm dit 'n kritieke menslike hulpbronbestuuringryping insoverre dit poog om die beweging van menslike kapitaal in, deur en uit die organisasie te reguleer met die verwagting dat dit sal lei tot verhoogde werksprestasie deur werknemers. Bedryfsielkundiges en menslike hulpbronpraktisyns gebruik dikwels psigometriese/sielkundige toetse in die keuringsproses, wat hulle met objektiewe inligting oor komplekse konstrukte soos intellektuele vermoë of persoonlikheid voorsien, onder die veronderstelling dat hulle belangrike determinante is van die vlak van werkverrigting wat gekeurde aansoekers sal bereik. Akkurate voorspellings kan egter slegs uit sodanige psigometriese/sielkundige toetse afgelei word indien die konstrukte wat hulle probeer meet, in werklikheid determinante van werkprestasie is, indien die toetse betroubare, geldige en onsydige metings van hierdie konstrukte gee en indien die aard van die verwantskap tussen die voorspellerkonstrukte en die kriteriumkonstruk geldig verstaan word. Persoonlikheid is 'n invoedryke determinant van werkprestasie. Die Personality and Preference Inventory-Normative (PAPI-N) is 'n persoonlikheidsvraelys wat algemeen in die bedryf gebruik word. Daarin lê die regverdiging vir die primêre doel van hierdie navorsing, naamlik om die eerste-orde faktor struktuur van die PAPI -N deur 'n factor-analitiese ondersoek op 'n relatief groot steekproef van die Suid-Afrikaanse werkende bevolking te evalueer, geleë.

Die data wat in hierdie studie gebruik is, is verkry uit die data-argiewe van Cubiks (Pty) Ltd, met die skriftelike toestemming van die intellektuele eiendiom-eienaar, om die steekproefdata aan te wend vir die doel van hierdie navorsing. Die Suid-Afrikaanse PAPI-N databasis bestaan uit al die kandidate wat geassesseer is deur Work Dynamics, die amptelike verspreider van Cubiks se produkte en dienste in Suid-Afrika, in die tydperk 2007-2012. Item en dimensionaliteitsontledings is uitgevoer op die 20 subskale van die PAPI-N, sowel as die sosiale wenslikheidskaal. Dit is gedoen om die sukses te bepaal waarmee die subskale die onderliggende persoonlikheidskonstrukte verteenwoordig. Die resultate van die itemontleding het getoon dat ten opsigte van sowat 50 % van die PAPI-N subskale, kommer bestaan oor die mate waartoe die items van die subskale in harmonie reageer op sistematiese verskille in 'n enkele onderliggende latente veranderlike. Resultate van die dimensionaliteitontleding het getoon dat 12 van die 20 persoonlikheidsdimensiesmetings versoenbaar is met die standpunt dat die items waaruit hierdie subskale bestaan, meet wat hulle ontwerp is om te meet. In teenstelling hiermee het agt uit die 20 subskale nie die uni- dimensionaliteitstoets geslaag nie.

A verskeidenheid pasgehalte-maatstawwe is gebruik om die pasgehalte van die metingsmodel te ondersoek. Oorkoepelend was die pasgehalte van die metingsmodel aanvaarbaar. Die nulhipotese van presiese passing is verwerp maar die nulhipotse van benaderde passing is nie verwerp nie ($p>.05$). The pasgehalte-maatstawwe het gedui op ‘n benaderde passing in die parameter en baie goeie modelpassing in die steekproef. Ofskoon die metingsmodel benaderde passing getoon het was die faktorladings (alhoewel statisties beduidend) oor die algemeen matig in omvang. Ongeveer agt-en-twintig present (27.78%) van die volledig gestandaardiseerde faktorladings was kleiner as die kritieke afsnywaarde van .50. Dit suggereer dat die items oor die algemeen (72.22%) wel die latent persoonlikheidsdimensies wat hul geoormerk is om te reflekteer, bevredigend reflekteer. In ‘n klein bietjie meer as ‘n kwart van die items is minder as 25% van die variansie in die itemresponse te wyte aan variansie in die latent veranderlike wat die item ontwerp was om te reflekteer. Diskriminantgeldigheid was ook ondersoek. Die resultate dui daarop dat die PAPI-N, ofskoon nie sonder problem nie, wel die suksesvolle onderskeid tussen die unieke aspekte van die persoonlikheidsdimensies moontlik maak.

Die resultate van die bevestigende faktorontleding dui daarop dat, terwyl die bedoeling van die PAPI-N om stelle items te hê wat spesifieke primêre persoonlikheidsfaktore reflekteer geslaagd was, die subskaal-itemmetings meestal 'n aansienlike hoeveelheid sistematiese en toevallige fout bevat. Gebaseer op die bogenoemde bevindinge, moet hierdie persoonlikheidsmeting met omsigtigheid gebruik word in personeelkeuring in Suid-Afrika. Nietemin, dra hierdie studie by tot ‘n groter begrip van die psigometriese eienskappe van die PAPI-N op steekproewe wat verskil van die Verenigde Koninkryk steekproef waarop dit oorspronklik ontwikkel en gestandaardiseerd is. Die bevindinge sal help om die nodige verderde navorsing te ontlok wat nodig is om die PAPI-N met vertroue as 'n waardevolle meetinstrument in Suid-Afrika te vestig. Aanbevelings vir toekomstige navorsing word gemaak.

ACKNOWLEDGEMENTS

I want to take this opportunity to thank my Heavenly Father for blessing me with opportunity and abilities to undertake this research. I want to thank Him for blessing my life with a foundation out of which absolute love, care, strength and security blooms. Give thanks to the ways how He never forsaken me in times when I have been at my weakest. In God I have hope because in John 16:3 it states "I have told you these things, so that in me you have peace". In this world you will have trouble but take heart. I have overcome the world".

I want to thank my parents, grandparents and family for their endless love, support and their involvement in my success. I am grateful to be blessed with inspirational role models who believed in my abilities when at times I doubted myself. To my dearest mother and father words could never fully express my sincere gratitude for all your sacrifices, love and believe in me. Thank you for presenting me with the opportunities to keep building and working towards achieving my full potential and passion in life.

I would like to give special thanks to my thesis supervisor, Prof. Callie Theron for helping and supporting me in reaching my goal. I could not have done it without your wisdom. Also special thanks to Cubiks (Pty) Ltd, particularly Dr. Rob Veldtman and his staff, as well as Work Dynamics, for allowing me to use the PAPI-N as subject of my thesis and the collected data.

I also give thanks to all my friends for their love, support and believing in me..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND OBJECTIVE OF THE STUDY

This chapter presents the research objective and an explanation as to why the research objective is considered relevant and important for the discipline and practice of Industrial Psychology in South Africa.

1.1. INTRODUCTION

Any organisation exists for some reason. Organisations in a free market economy exist with the purpose to provide goods or services to society, to provide its stakeholders with maximum profits and to positively impact on the quality of life of its employees and on the physical environment. Organisations in a free market economy need to serve the market with products and services that the market values while at the same time satisfying the triple bottom line of profit, people and planet (Elkington, 1998). An organisation will only be successful in as far as it meets that aim (Hackett, 1985). However, the success of any business depends to a large extent on the calibre of its workforce (Plumbley, 1985). No business can exist or operate without the support of human resources, in fact, every organisation needs human resources for the conduct of various business activities (Akrani, 2009). In essence, organisations exist through, are managed, operated and run by people (Moyo, 2009). To an enormous degree organisational success will significantly depend on the effectiveness and quality of its employees (Twigge, 2003). This inter-related and interdependent relationship between employees and the organisation suggests the importance of human resources as an important asset in an organisation, which should be utilised and managed effectively.

Human resource management represents a range of interventions with the purpose of contributing to an organisation's success by improving employee job performance while ensuring the physical and mental well-being of its work force (Psychology Dictionary, 2012). The essential logic underlying human resource management flows from the basic premise that organisational success is significantly dependent on the quality of its human capital and the manner in which they are utilised and managed (Moyo, 2009). The primary function of human resource management is to contribute towards achieving organisational goals and objectives, through the acquisition and maintenance of a high quality and competent work force, as well as to ensure the effective and efficient use of human talent in a manner that will add value to an organisation (Akrani, 2009; Habeeb, 2009).

1.1.1 PERSONNEL SELECTION

Personnel selection represents one of these human resource functions (Twigge, 2003) and can be regarded as a primary mechanism through which the organisation's overall effectiveness can be enhanced (Habeeb, 2009). In general, organisations differentiate themselves from one another in terms of their industry, size, type, operations, and position in the market (Habeeb, 2009). However, the main element that distinguishes one organisation from another is its employees (Akrani, 2009; Habeeb, 2009). An organisation's employees provide the essential ingredient for its competitive advantage (Habeeb, 2009). The goal of personnel selection is to add value to the organisation by ensuring that an organisation hires the highest qualified and most competent individuals as well as promoting those from within the organisation (Habeeb, 2009). Having the right people in the right place at the right time, willing and able to work effectively, and at a cost that the organisation can afford, is something for which all managers and Industrial/Human Resource practitioners should strive (Hackett, 1985).

Personnel selection thereby constitutes a critical human resource management intervention in as far as it attempts to regulate human capital movement into, through and out of the organisation with the expectation that this will result in increased employee job performance (Theron, 2007). Selection, however, usually implies a situation where there are more applicants for openings than there are vacancies available or even training and developmental opportunities. Selection procedures therefore follows a methodology to collect information about individuals in order to determine the individual best suited for success in a particular job (HR-Guide, 2001), or to identify those who might benefit most from further training and development opportunities (Paterson & Uys, 2005). More specifically, selection procedures are designed to filter those prospective employees from the total group of applicants that would perform optimally on the multi-dimensional criterion construct (η) (job performance or training performance) (Cronbach & Gleser, 1965). The ultimate or final criterion is the criterion construct which selection seeks to affect (future job or training performance). Furthermore, the subgroup of applicants has to be chosen so as to ensure that the average performance on the ultimate criterion is maximised (Austin & Villanova, 1992). The utility scale/payoff and the actual outcomes or ultimate criterion should thus always be the focus of interest in selection decisions (Bartram, Baron & Kurz, 2003; Ghiselli, Campbell & Zedeck, 1981). This seemingly and too often forgotten fact has significant implications for the decision-maker in the interpretation and evaluation of selection decision information.

If the decision-maker knew beforehand how well an individual would perform on a particular job, selection would not present a difficult decision problem (Twigge, 2003). The ideal situation would be if selection decisions could be based directly on the information of the multi-dimensional criterion (performance) construct (Theron, 2009b). However, the ideal situation is practically not possible, since information on actual job performance can never be obtained directly at the time of the selection decision as the individual's actual performance only discloses itself after being employed. In the absence of such desired information the only alternative to make better than chance decisions, would be to predict future criterion performance from relevant, though limited, information available at the time of the selection decision and then base the selection decision on the predicted criterion performance expected from that individual (Theron, 2007).

The decision-maker is therefore tasked with the responsibility to obtain substitute (predictor) information that is available at the time of the selection decision (Twigge, 2003), to infer future criterion performance from the substitute information and then to base the decision on these criterion-referenced inferences (Theron, 2007). However, in personnel selection the primary focus of interest should always be on the criterion and not on the predictor from which inferences about the criterion are made (Ghiselli et al., 1981). This position is formally acknowledged by the APA sanctioned interpretation of validity and especially predictive validity (Ellis & Blustein, 1991; Landy, 1986; Messick, 1989; Society of Industrial and Organisational Psychology, 2003). This position, furthermore, also underlies the generally accepted regression-based interpretations of selection fairness (Cleary, 1968; Einhorn & Bass, 1971; Huysamen, 2002; Theron, 2007). While this might not seem to be a consequential argument, the criterion-centric nature of personnel selection is critically important and failure to appreciate its importance lies at the root of a number of popular misconceptions regarding the use of psychometric tests in personnel selection. It specifically forces one to critically rethink (a) the use of construct referenced norms in personnel selection, (b) the belief that tests are the villains responsible for adverse impact, and (c) the belief that assessments techniques can be certified as Employment Equity Act (EEA) (Republic of South Africa, 1998) compliant (Theron, 2007).

An accurate (clinical or mechanical) estimate of measures of the criterion construct will be possible from predictor information if it meets the following three conditions. Firstly, the predictor needs to correlate with a valid and reliable measure of the criterion; secondly, the nature of the predictor-criterion relationship in the appropriate applicant population has to be accurately understood by the decision-maker; and lastly, construct valid measures of the predictor construct must be available (Theron, 2009b).

In the absence of direct criterion information at the time of the selection decision, only two possible options exist in terms of which relevant substitute (predictor) information can be obtained from which expected criterion performance can be inferred. Substitute information (X_i) can be considered relevant to the extent that it will permit an accurate prediction of a (construct valid) measure (Y) of the criterion construct (η). Substitute information (X_i) will permit an accurate prediction of the criterion construct (η) to the extent that it systematically correlates with a (construct valid) measure (Y) of the criterion construct (η). As with any organisation, jobs are designed and created to serve a specific purpose. Jobs consist of a defined set of inter-related behavioural tasks or demands required to accomplish some objective. A distinction can be made between task-related behaviours and contextual behaviours (Myburgh, 2013). An individual's level of performance achieved on these tasks/demands is not simply a random walk through the work place but rather is determined by a complex nomological network of person- and environmental factors (Theron, 2009b). This points to two options that could provide such correlates of performance, namely; to operationalise (via X_i) the person-centred constructs (ξ_i) required to perform successfully on the job as inferred from the job description, or to evaluate (via X_i) how well a person responds to the demands, constraints and opportunities that constitute the job and that need to be met to be considered successful, as inferred from the job description, outside the job. These two options can be referred to as the construct- and content orientated approach to selection. Both approaches obtain substitute information for the ultimate criterion by measuring latent variables through observable behaviour elicited by a stimulus set. While the fundamental measurement logic is the same, the nature of the latent variables is, however, different. In the construct orientated approach to selection person constructs (ξ_i) are measured by sampling the relevant predictor construct domains. In the content orientated approach to selection the performance construct (η) is measured outside the job by sampling the criterion construct (Binning & Barrett, 1989).

In terms of the content orientated approach, the stimuli are designed to elicit behavioural manifestations of the performance construct off the job. In other words, the predictor stimuli elicit actual behaviour that the actual components of the job would have been elicited but they do so outside the target job via a simulation of the job or via another job similar in the task and contextual demands that it imposes to the target job. Although the actual response to the sample of stimuli is determined by a nomological network of person-centred constructs, the identity of these latent variables is not always known. Thus the stimuli only recall the nature of responses elicited by facets of the job (Theron, 2009b). This requires that the job in question is systematically analysed through an appropriate job analysis technique. Job analysis determines the performance domain by identifying the job competencies or key behavioural performance areas that collectively denote job success if exhibited on the job and that translate into the outcomes for which the job exists and that collectively denote job success if achieved on the job. Predictor information would then be obtained by simulating the demands that need to be met on the job to be considered successful or by assessing performance in another job for which the target job competencies also can be considered relevant. In a selection context, these simulations necessarily occur off the job and prior to the selection decision. Such simulations would reflect behaviours that, if in future exhibited on the job after appointed, it would denote a specific level of job performance. This implies that in a content orientated approach, predictor sampling is guided by evoking a performance domain (Binning & Barrett, 1989).

In terms of the construct orientated approach, the stimuli are designed in such a way that a person's response to them is mainly a function of a specific, defined person construct (ξ_i] or it elicits historical recall of behaviour in which ξ_i expresses itself. A construct orientated approach thus involves identifying psychological construct domains that significantly correlate with the performance domain and then to develop predictors that could adequately sample these construct domains (Binning & Barrett, 1989). In terms of this approach to predictor development, a performance hypothesis is developed in the form of a tentative job performance structural model that maps job competency potential latent variables onto the job competencies and latent outcome variables that constitute job success (Moyo, 2011). The job in question will also be systematically analysed with the purpose of inferring presumed critical incumbent attributes believed to be determinants of the level of criterion performance that would be attained from the description of the job content and context (Twigge, 2003; Moyo, 2009). If the complex performance hypothesis is valid, it would in principle be possible to estimate job performance from measures of the potential latent variables. However, this is only possible if the nature of the relationship between the performance construct and its person-centred determinants are accurately understood in the appropriate applicant population and if the predictor constructs could be measured in a construct valid manner at the time of the selection decision (Moyo, 2009).

Selection procedures are thus possible in terms of the construct orientated approach only if it is based on a valid substantive performance hypothesis, if the nature of the relationship between the performance construct and its person-centred determinants are accurately understood, and if construct valid measures of the person-centred determinants are available at the time of the selection decision (Theron, 2007). The effectiveness of a selection procedure consequently depends on the extent to which the underlying performance hypothesis is valid (Twigge, 2003).

To establish the validity of the performance hypothesis, operational hypothesis are empirically derived from the overarching substantive performance hypothesis by defining the performance construct and the explanatory psychological constructs operationally. Both operational definitions of the performance construct and the explanatory psychological constructs constitutes premises in a deductive argument (Theron, 2009b). The validity of these premises determines the validity of a deductive argument, since the premises provide conclusive grounds for the truth of the conclusion. In other words, the conclusion that is derived from the deductive argument will only be true if the premises are true (Theron, 2009b). Therefore to justify that the operational performance hypothesis constitute a valid testable representation of the theoretical performance hypothesis requires evidence on the construct validity of the operational measures of the performance construct and the explanatory psychological latent variables (Moyo, 2011). Should the deductive argument be valid and empirical conformation of the operational performance hypothesis be found, the substantive performance hypothesis may be regarded as valid since it survived an opportunity to be refuted (Theron, 2009a). The claim that job performance can be predicted from a range of operational predictor measures through a construct-orientated approach, is partially justified if it can be shown that the substantive performance hypothesis is valid and if it can be shown that the operational measures of the explanatory psychological constructs and the operational measure of the criterion construct are construct valid (Theron, 2007). Evidence about the construct validity of the criterion and predictor measures, however, only constitutes a necessary, though not sufficient condition, to justify the claim that job performance can be predicted from a range of operational predictor measures and to *de facto* achieve valid criterion estimates. It also needs to be shown that the construct valid measures of the explanatory predictor constructs and the construct valid measure of the criterion construct correlate statistically significantly ($p<.05$), that the nature of the relationships between the predictor and criterion measures are accurately understood and that the criterion estimates derived from this understanding statistically significantly correlate with construct valid measures of the criterion construct. Predictive validity refers to the question whether it is permissible to derive criterion inferences from predictor measures. This is not convincingly established by only demonstrating that predictor measures correlate with the criterion (Binning & Barrett, 1989).

Practitioners frequently use psychometric tests in the selection process as measures of predictor constructs as they provide information of complex constructs such as intellectual ability, personality, knowledge and skills which are difficult to measure in a standardised, objective, reliable and valid manner with other techniques such as interviews (Paterson & Uys, 2005). Psychological tests can therefore help the decision-maker to make an informed decision about an individual's suitability for selection and developmental purposes (Paterson & Uys, 2005). In the South Africa context, however, this would only be possible if the construct of interest can validly and reliably be measured across all ethnic groups and if the target construct can be measured in the same way across these groups.

In South Africa, there exists a definite need for psychological measures that meet the standard requirements of validity and reliability which also give unbiased measures of the target construct across race, gender and cultural groups (Moyo, 2011). Given the multicultural and multilingual nature of the South African society, practitioners are faced with many challenges in psychological test use (Foxcroft, Paterson, Le Roux & Herbst, 2004). In addition to the requirements that a diverse society demands (Van Zyl & Tylor, 2012), psychological assessment has been and is currently being shaped by: strict legislation (e.g. Employment

Equity Act; Republic of South Africa, 1998) and its political dispensation; the need for appropriate measures that can be used in a fair[1] and unbiased way across all cultural groups in South Africa; the need for practitioners to take responsibility for ethical test use; and relevant practice guidelines provided by statutory (e.g. the Professional Board for Psychology) and other controlling bodies (Foxcroft & Roodt, 2005).

Consequently, this places the responsibility on practitioners, test developers and distributors, to generate sophisticated, indisputable scientific evidence that the measurements used in South Africa are psychometrically suitable for and relevant to the South African context. Moreover, this challenges the Industrial-Organisational Psychology fraternity to demonstrate that the assessment techniques used in personnel selection in South Africa succeed in measuring the intended predictor constructs as constitutively defined across different ethnic groups and that the assessment techniques measure their target constructs in the same way across different ethnic groups. It is within this context that the assessment of personality occurs in South Africa.

1.1.2 RELEVANCE OF PERSONALITY AT WORK

According to Anderson and Lewis (1998), there are many factors that influence an individual's working identity and behaviour at work. Five of these factors are presented in Figure 1.1 below.

The model presents five basic factors that contribute to a person's behaviour at work. Ability contributes to a person's behaviour at work in terms of the extent to which he/she can efficiently perform multiple processes to achieve a specified goal. Other factors that determines behaviour at work includes a person's intelligence in terms of his/her capacity for abstract and critical reasoning; demographic factors such as age, education and social class; and motivation – the driving force that provokes action – which includes interest, needs and values. Furthermore, the model emphasises personality by referring to all those characteristics of an individual that accounts for consistent patterns of responses across everyday situations[2].

---

[1] It is thereby not implied that the fairness of the criterion inferences derived from predictor information can be ensured by the judicious choice of selection instruments.

[2] It is thereby not implied that personality brings about consistency in behaviour independent of the nature of situations but rather that personality in interaction with specific (perceived) characteristics of the situation bring about consistency in behaviour (Mischel, 2004).

Views regarding the use of personality measures for personnel selection have changed over the years. In their review of research published in the *Journal of Applied Psychology and Personnel Psychology* over a 12 year period from 1952 to 1963, Guion and Gottier (1965) came to the conclusion that personality tests do not warrant use in personnel selection. This pessimistic position was generally not opposed until Barrick and Mount (1991) and Tett,

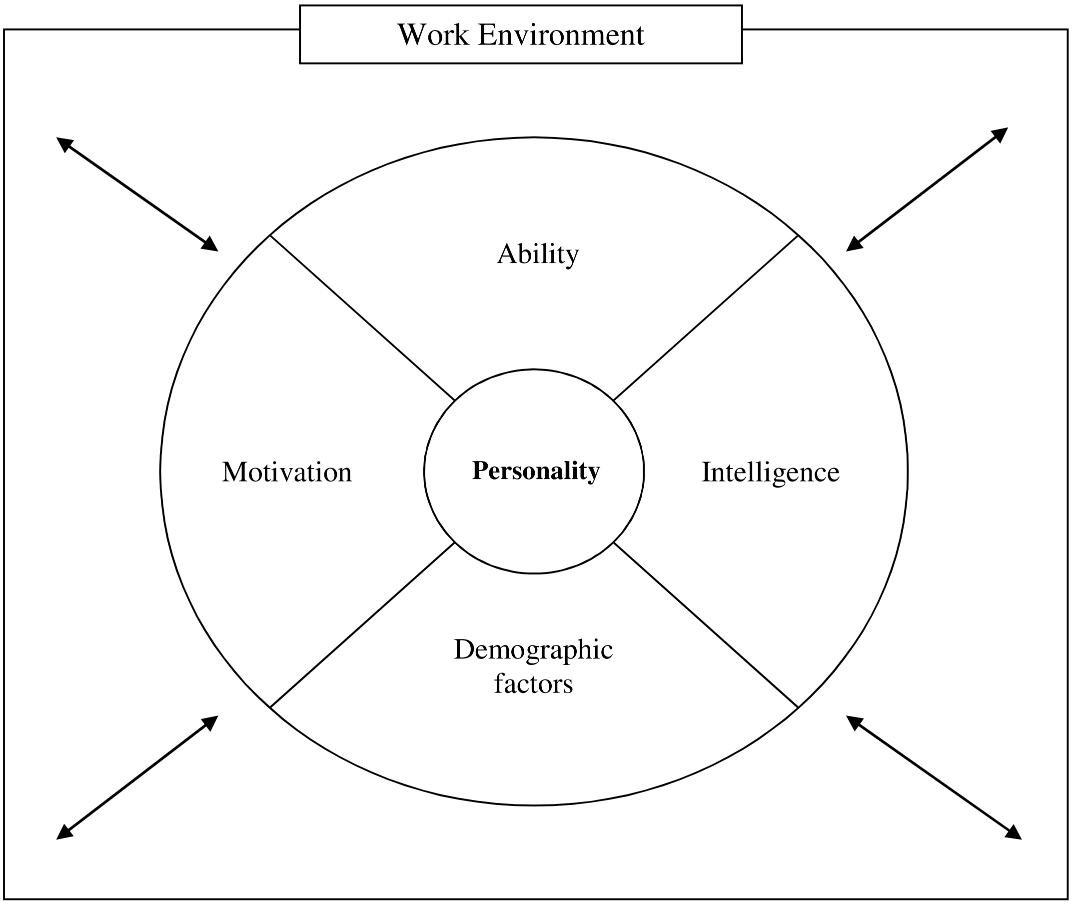

Figure 1.1: Our work identity model developed by Anderson, P., & Lewis, C (1998). *PAPI Technical Manual*. London: PA Consulting/Cubiks Copyright 1998 by PA Consulting/Cubiks.

Jackson and Rothstein (cited in Morgeson *et al*., 2007a) in 1991 challenged it with their meta analyses. The pendulum has now swung back and personality is now generally appreciated as an influential causal antecedent of job performance (Moyo & Theron, 2011) and especially contextual performance (Borman & Motowidlo, 1993; Van Scotter & Motowidlo, 1996). Although cognitive ability remains the best predictor of job performance (Schmidt & Hunter, 1998), numerous studies indicated that the use of tests specifically designed to measure personality within the context of work increases criterion-related validity, thereby increasing the utility of personality constructs in industrial and organisational psychology (Sanz, Gil, Barras, & García-Vera, 2006). The importance and relevance of personality as a contributing factor to work behaviour cannot go unstated, and has been researched from many perspectives (Anderson & Lewis, 1998). For example, Goodstein and Lanyon (1999) demonstrated the usefulness of personality traits in predicting job performance, job satisfaction and leadership, while other researchers emphasised the notion of a person-job-organisational fit (Anderson & Lewis, 1998). Ultimately, understanding personality is critical as it can help with predicting how an individual will behave when placed in a specific situation characterised by specific features, for example, being exposed to new conditions or when put under stress of being unable to rely on acquired expertise or previous experiences. Employee performance is complexly determined by a nomological network of latent variables characterising the employee[3] and the nature of the environment that the employee is operating in. Personality is embedded in this nomological network. If the Industrial-Organisational Psychology discipline can obtain a valid understanding of the manner in which this complex nomological network of latent variables, characterising the employee and the nature of the environment that the employee is operating in, affects his/her work performance then the discipline can begin the process of understanding how an employee will perform against the demands of a specific job. This will also increase the understanding of how the profession can help improve their performance given an understanding of what a person is really like, what motivates them, what particular qualities they have and the nature of the environment that they are working in (Anderson & Lewis, 1998).

Demonstrating the usefulness of personality traits in predicting job performance has fostered the use of personality measures in personnel selection. However, emphasis on traits does not preclude the study of other personality elements. Along with traits, needs are among the leading candidates to be useful units for personality research (Sanz et al., 2006). The *Personality and Preference Inventory* (PAPI) is a self-report questionnaire based on Murray's need-press theory and is designed to assess needs (i.e. drivers or motivators) and roles (or behaviours) as experienced or displayed in the workplace. The questionnaire explores a broad range of personality dimensions, which are split between *role* and *need scales*. The *role scales* measure the individual's perception of themselves in the work environment and looks at areas

[3] The latent variables characterising the employee included relatively non-malleable dispositions like personality traits, intelligence and interests but also more malleable attainments like knowledge and more transient states like psychological ownership, psychological empowerment and engagement.

such as *leadership*, *work style* and *planning*. The *need scales* probe the deeper inherent tendencies such as the *need of an individual to belong to a group*, *the need to be noticed* and *the need to achieve*. There are two different versions of the PAPI available, the ipsative format (PAPI-I) and the normative format (PAPI-N). For this research, only the normative version of the Personality and Preference Inventory (PAPI-N) was used.

Since its inception, PAPI-N has become a leading work-related personality questionnaire used by more than 5000 professionals across the globe (Cubiks, 2012). It is also widely used by psychologists and psychometrists in South Africa. The confident use of the PAPI-N in selection in South Africa requires (a) that a convincing argument be developed as to why and how personality (as interpreted by PAPI-N) should be related to job performance, (b) that a structural model derived from the foregoing argument fits empirical data, i.e. there is support for the performance hypothesis, (c) that evidence be available that the predictor and criterion constructs are validly and reliably measured in the various sub-groups typically comprising applicant groups in South Africa, (d) that evidence be available that (at least) race and gender group membership do not systematically affect the manner in which the predictor and criterion constructs express themselves in observed measures, (e) that evidence be available that the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures, (f) that evidence be available that criterion predictions derived (clinical or mechanical) from the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures and (g) that evidence be available whether (at least) race and gender group membership does explain variance in the criterion (either as a main effect or in-interaction with the criterion estimates derived from the PAPI-N) that is not explained by the criterion estimates derived from the PAPI-N.

The objective of this research is therefore to contribute to the available psychometric evidence with regards to the third aspect (c) mentioned above. The confident utilisation of the PAPI-N in specific personnel selection procedures aimed at filling posts in specific positions in specific organisations would, however, in addition to the above also require credible evidence on the fairness and utility (Guion, 1998) of the selection procedure.

1.2 OBJECTIVES OF THE STUDY

The PAPI-N is based on a specific interpretation of personality. Specific personality dimensions which are structured around the needs and roles as displayed or experienced in the work environment are distinguished in terms of this interpretation. The architecture of the PAPI-N reflects a specific design intention. The design of the PAPI-N questionnaire reflects the intention to construct twenty essentially one-dimensional sets of six items each to reflect variance in each of the twenty latent personality domains collectively comprising the personality construct. The PAPI-N items are designed to function as homogenous stimulus sets to which applicants respond with behaviour that is a relatively uncontaminated expression primarily of a specific underlying latent personality dimension. Specific items were therefore selected for each scale because they are believed to reflect and correlate with a specific first-order personality dimension.

The scoring key of the PAPI-N reflects the expectation that all items comprising a specific subscale should load on a single dominant factor. This implies that the items can be used to obtain an observed score for that specific personality dimension, and that dimension only. When computing a subscale score for a specific personality dimension, only those items comprising that specific subscale are combined. It does not imply that the twenty first-order personality dimensions do not to a certain degree share variance. The PAPI-N assumes that the needs and roles are interrelated and could be interpreted in terms of seven second-order factors (Anderson & Lewis, 1998). A very specific measurement model is thereby implied in which each specific latent personality dimension comprising the PAPI-N's interpretation of personality reflects itself primarily in the specific items written for the specific subscale. The first-order measurement model could also be expanded into a second-order measurement model also reflecting the manner in which second-order personality factors express themselves in first-order personality dimensions.

The objective of this research study is to evaluate the fit of the first-order measurement model of the PAPI-N, as implied by the architecture of the instrument and the constitutive definition underlying its constructs, on a relatively large sample of the South African working population.

1.3 STRUCTURE OF THE THESIS

An overview of the development of the PAPI-N will be reported in Chapter 2. This chapter will also present the definition of personality underlying the PAPI-N. Available international and South African psychometric evidence on the reliability and validity of the PAPI-N as a measure of personality within the work environment (given its specific constitutive definition) will also be reviewed. In Chapter 3 the methodology used to evaluate the PAPI-N measurement model fit will be described. Chapter 4 will present the research results and Chapter 5 will present the conclusions and implications for future research.

CHAPTER 2

AN OVERVIEW OF THE PAPI-N AS A MEASURE OF PERSONALITY IN THE WORK ENVIRONMENT

2.1 INTRODUCTION

This section of the thesis will explain the process followed by the developers of the PAPI-N in the construction of this personality questionnaire specifically designed to assess behaviours and preferences at work. The introductory section emphasised the need for a close psychometric inspection of the PAPI-N as a measure of personality as the new PAPI-N English version is widely used in South Africa. The purpose of the study is to focus on the psychometric credentials of the PAPI-N as to justify its use as a valuable assessment tool in the context of South Africa. The objective of the research is to evaluate the first-order factor structure through a factor analytic investigation of the PAPI-N. Thus a confirmatory factor analysis will be undertaken into the first-order factor structure to determine whether all the items in the questionnaire reflect the latent personality dimensions (according to PAPI-N's scoring key) for which they were designed in the group to be studied. Evaluation of the fit of a measurement model essentially is an evaluation of the success with which a latent variable carrying a specific constitutive definition has been operationalised by means of an instrument developed with a specific design intention. Should the measurement model implied by the design intention and scoring key of the PAPI-N, fit the sample data well, it would, however, still constitute insufficient evidence to justify its use within the South African multi-cultural setting. The fact that measurement model fit has been shown on a large South African sample would still beg the question whether (a) the measurement model fit holds across the various gender-racioethnic sub-groups, and if so, (b) whether the model parameters are the same across such groups. Therefore a critical question is whether the measurement model underlying the PAPI-N succeeds in measuring the construct across different gender-racioethnic groups as it was constitutively defined and whether the inference that can be made about the state/level of the measured construct given a specific observed score is the same across groups. However, the objective of this study is to only evaluate the fit of the first-order measurement model underlying the PAPI-N on a large South African sample.

This chapter will further discuss the constitutive definition underlying the PAPI-N, followed by an overview of the development of PAPI-N and its development in South Africa, including the structure of the instrument. Finally, current available reliability and validity findings on the PAPI-N will be discussed.

2.2 DEFINITION OF PERSONALITY UNDERLYING THE PAPI

The term 'personality' derives from the Latin word 'persona' which refers to the 'mask' that actors used in Greek theatre to portray various stage roles. Over time, however, the word has evolved to display the character being portrayed rather than the mask alone (Anderson & Lewis, 1998). Although personality is such a commonly known word today, it still remains a complex and dynamic concept (Van der Merwe, 2005). Personality generally refers to the different ways in which people behave, and usually refers to an individual's normal behaviour (Van der Merwe, 2005). In psychological terms, however, personality could be defined in many different ways (Anderson & Lewis, 1998). Consider the following for example:

Allport (as cited in Anderson & Lewis, 1998, p. 1) defines personality as:

> the dynamic organisation within the individual of those psychophysical systems that determine his unique adjustment to his environment.

Cattell (as cited in Anderson & Lewis, 1998, p. 2) defines personality as:

> that which permits a prediction of what a person will do in a given situation.

Atkinson, Atkinson, Smith and Bem (as cited in Smith & Smith, 2005, p. 32) provide the following definition:

> The characteristics of thought, emotion and behaviour that define an individual's personal style and influence his or her interactions with the environment.

Block, Weiss and Thorne (as cited in Anderson & Lewis, 1998, p. 2) define personality as:

> more or less stable, internal factors that makes one person's behaviour consistent from one time to another and different from the behaviour other people would manifest in comparable situations.

Meyer, Moore and Viljoen describe personality in Van der Merwe (2005, p. 23) as:

> the continuous changing, but relative stable organisation of all physical, mental and spiritual characteristics of the individual, that determines behaviour. These characteristics are interacting in the context in which the individual finds himself.

Anderson and Lewis (1998, p. 2) concluded and described personality as:

> a complex set of unique psychological qualities that influence an individual's characteristic patterns of behaviour across different situations and over time.

The foregoing definitions tend to suggest that personality determines behavioural consistency across different situations. This stance has been critically challenged over an extended period of time by people such as Mischel (2004). Mischel's (2004) criticism has frequently been misunderstood that he claims that no such thing as personality exists (Smith & Smith, 2005). This is, however, not the case. Mischel (2004) argued that variability in the behaviour of individuals with a stable personality structure across situations is due to the interaction between personality and situational characteristics. Characteristics of the situation along with personality traits both need to be treated as necessary and integral components of personality theory. It is not so much objective characteristics of the situation that are important, but rather the individual's subjective interpretation of the situation. An individual with a stable personality structure can therefore only be expected to behave consistently across situations if the individual perceives the salient characteristics of the situation to be similar (Mischel, 2004). Mischel's stance does, however, raise concern about how personality can be measured if the manner in which a specific personality structure manifests itself is not consistent over a variety of situations (Murphy & Davidshofer, 2001). This line of reasoning points towards the need for domain-specific personality questionnaires (e.g. work-related personality questionnaires). Van der Merwe (2005) indicated that the aspect of behavioural changes over time and across situations still remains an issue amongst theorists. Smith and Smith (2005) also mentioned in their overview of the literature that researchers believe that individuals behave very differently across different situations and that although there is less regularity in their behavioural style across situations there is more consistency within situations. For example, according to Hartsorne and May's (1928) example on cheating (as sited in Smith & Smith, 2005), who indicated that different individuals cheated in different situations and therefore a low correlation was found between cheating in examinations and cheating in other situations, which suggested that neither personality or the situation independently affect behaviour but rather the interaction between the two. Other researchers, however studied personality over a long period of time and found adolescents to be consistent in their behaviour across situations. In addition, considerable evidence suggested personality to be very much stable after the age of 30, while weaker evidence indicated towards a 4% change (per year) in personality between the ages of 18 and 21 (Smith & Smith, 2005).

Given the complexities associated with personality, Murphy and Davidshofer (2005) proposed three important facts when trying to explain personality:

- Personality, like a fingerprint (Anderson & Lewis, 1998), is unique in the sense that no one is identical in terms of their behaviour, temperament or preferences.
- Individuals may not behave across all situations in the same manner, because as situations vary so will a person's behaviour.
- While people are unique, there still exists some commonality in their behaviour.

In the next section, theories that emerged from different paradigms to try and make sense of personality are discussed. There are many theories of personality that exist within the literature, which this section divided into two groups, namely implicit and explicit personality theories.

2.3 THEORIES OF PERSONALITY

According to Bruner and Tagiuri (as cited in Smith & Smith, 2005), implicit personality theories refers to those ideas that individuals develop about the manner in which people's personality characteristics fit together. Individuals hold a network of assumptions, which they base on relationships among various traits and behaviours (Implicit Personality Theory, n.d.). Furthermore, Asch (1946) found that the presence of one trait is often associated with the existence of other traits or characteristics. When individuals associate a particular trait with someone, they will therefore assume that the person also possesses other additional traits as well (Implicit Personality Theory, n.d.). For example, people who are good looking are at the same time also assumed to be vain, or people who make friends easily do not have deep relationships (Smith & Smith, 2005). It can be seen that most implicit personality theories tend to describe people either in terms of physical appearance or group membership (Smith & Smith, 2005). These naive personality assessments can often be accurate, but are also open to many sources of error such as prejudice, bias and social identity issues (Anderson & Lewis, 1998), and it would be risky to base important decisions upon them (Smith & Smith, 2005).

Explicit theories of personality, on the other hand, attempt to understand and explain the roots, structure and correlates of personality as well as to predict behavioural trends on the basis of personality. These explicit theories of personality include the Psychoanalytic, Humanistic, Social learning theories, as well as the Psychometric approach to personality (Anderson & Lewis, 1998).

2.3.1 PSYCHOANALYTIC THEORY

Psychoanalytic theory was originally proposed by Sigmund Freud (Moyo, 2009), who developed the very concept of human personality (Anderson & Lewis, 1998). Shared by all psychodynamic personality theories, is the assumption that personality is shaped, and behaviour is motivated by three interdependent psychological forces (Anderson & Lewis, 1998; Moyo, 2009). According to Freud, human behaviour and/or actions are prominently motivated and determined by motives and/or desires, through which these can, consciously or unconsciously, affect one's thoughts and behaviour (Anderson & Lewis, 1998). The psychodynamic approach to personality therefore suggests that differences in personality arise from the manner in which the three interdependent psychological forces collectively work together or come into conflict. Freud referred to these forces as the id, superego and ego. The id is conceptualised as the unconscious part of personality which is governed by irrational uncontrolled pleasure-seeking impulses, especially sexual, physical and emotional pleasures, which demand immediate gratification without considering whether it is socially desirable or morally acceptable (Anderson & Lewis, 1998). The superego represents the second part of personality which develops through socialisation during childhood and includes a person's values and moral ideas of right and wrong. The superego also contains the ego ideal, as it directs the individual's view of the kind of person he or she would like to become. The superego is often in conflict with the id, as the id wants to do what feels good while the superego operates according to what is right. The ego represents the last part of personality which contains the reality-based aspect of personality and refers to an individual's conscious beliefs about the causes and consequences of behaviour. It arbitrates the conflict between the urges of the id and the demands of the superego as the ego chooses an action that gratifies id impulses without undesirable consequences. The ego operates according to the reality principle and would therefore put reasonable choices before pleasurable demands. When the id and superego come into conflict, the ego resolves this by working out a compromise that would partially satisfies both. However, it becomes more difficult for the ego to meet an optimal compromise when these pressures between the id and superego intensify. Accordingly, the ego often resorts to defence mechanisms such as repression, projection and regression that could resolve conflicts and maintain a favourable self-image (Anderson & Lewis, 1998). Based on this theory, the psychoanalytic approach emphasises that personality development is dependent on early childhood experiences and largely determined by unconscious motives and conflicts between primitive impulses/wishes/desires and learned social morals/values (Moyo, 2009). However, the psychoanalytic theory has been criticised for being good history but bad science as it does not reliably predict what will occur in present and future events. Furthermore, the fact that Freud developed a theory of 'normal' personality from clinical observations and in-depth studies of patients with mental disorders that has little to say about healthy lifestyles, offers the pessimistic view that human natures develops out of conflicts, traumas and anxieties (Anderson & Lewis, 1998).

2.3.2 HUMANISTIC THEORIES

The phenomenological approach to personality consists of a number of theories that differ in some respect but share a common emphasis on subjective experiences. Phenomenological theories primary focus on how the individual perceives and interprets events rather than on a person's motivational history, and tends to emphasise understanding behaviour rather than predicting behaviour (Anderson & Lewis, 1998). The phenomenological approach to the study of personality includes theories such as the humanistic (because they focus on those qualities such as self-direction and freedom of choice that differentiate humans from animals) and self-theories (because they deal with internal, subjective self-experiences that constitute a human's being) (Anderson & Lewis, 1998). The humanistic approach to personality stresses the positive side of human nature and emphasises the importance of self-actualisation, the basic tendency to develop and realise one's inherent potential and growth. This theory is based on the notion that an individual's unique biological and learned tendencies serves as a basic motivator that will direct a person's behaviour towards the goal of self-actualisation. The term self-actualisation can best be understood as the constant striving to realise one's potential in terms of fully developing one's personal abilities and talents. Therefore, the experiences that are perceived to maintain or enhance the person are evaluated positively and pursued, while those experiences that oppose the positive growth of the self are evaluated negatively and avoided. However, the humanistic approach to personality has been criticised for ignoring the role of unconscious motives and situational determinants of behaviour, oversimplifying the complexities of personality by reducing it to a given trend of self-actualisation, and fails to predict how an individual will respond in a given situation (Anderson & Lewis, 1998).

2.3.3 SOCIAL LEARNING THEORIES

The social learning theory was developed by Albert Bandura (1977) and Walter Mischel (1968). This theory of personality emphasises the importance of situational and/or environmental influences as determinants of behaviour and personality. According to this school of thought, individual differences in behaviour results from variations in learning experiences in the course of growing up. Most behaviour patterns are learned through observational learning while some are learned through direct experience or reinforcement, that is, when an individual is rewarded or punished for behaving in a certain way. Furthermore, social learning theorists assumes that an individual's actions in a given situation depends on the specific characteristics of the situation, their own personal appraisal of the situation and past reinforcement or observations in similar situations. This theory stresses the importance of personality differences in terms of cognitive development and social learning experiences rather than motivational traits in predicting how someone will behave in a specific situation. From this perspective, behaviour and personality are largely influenced and shaped by situational or environmental conditions. Social learning theories, however, have been subjected to criticism in that they mainly focus on environmental influences and previously learned experiences while ignoring the origin of new behaviour such as creative ideas, achievements and inventions. Finally, many theorists are unwilling to concede that personality has little stability as this school of thought implies (Anderson & Lewis, 1998).

2.3.4 THE PSYCHOMETRIC APPROACH TO PERSONALITY

The psychometric approach assumes that personality has a defined structure. It consists of identifiable traits, which results from either inherited or learning factors that are relatively stable, can be measured precisely and objectively, and can be used to predict future behaviour. Their focus is on comparing different individuals on specific aspects of personality, rather than on the uniqueness of each individual. Theorists differ about the number of traits, as some emphasises many narrow concepts (traits) while others puts more emphasise on a smaller number of broader constructs (types) (Anderson & Lewis, 1998). The different trait and type theories are next discussed.

2.3.4.1 Type Theories of Personality

This approach to describing personality involves classifying people into a limited number of personality types. More specifically, distinct patterns of personality characteristics are used to assign people to discrete categories. Type theories also attempts to predict future behaviour on the basis of a person's personality type (Anderson & Lewis, 1998). The type theory was first postulated by Hippocrates, who believed that personality was determined by four humours (blood, phlegm, black bile and yellow bile), each associated with a specific temperament. He believed that an individual's personality is dependent on which humour was predominant, for example, if an individual's predominant humour was blood, that person would have a sanguine type of personality (Smith & Smith, 2005). Other type theories include the works of Sheldon who developed his types of personality on the basis of their somatotypes or body builds (endomorphic, mesomorphic, and ectomorphic) (Smith & Smith, 2005), and Jung (1921, 1971) who developed a theory of psychological types designed to categorise people into several personality patterns. Jung's theory proposed that each person has a psychological type which is determined by a combination of four (bipolar) dimensions, namely extraversion-introversion (E-I), thinking-feeling (T-F), sensing-intuition (S-N), and judgment-perception (J-P) (Foxcroft & Roodt, 2005).

Type theories have been subjected to a number of criticisms. Firstly, type theories ignore a great amount of empirical evidence that personality is normally distributed and that people are not restricted to just a few groups (Smith & Smith, 2005). Secondly, they are criticised for oversimplifying the complexities of personality because types are all-or-nothing phenomena and not matters of degree to which they can be described. Thirdly, they do not explain how an individual's behaviour is caused, or how personality develops, they merely identify and describe characteristics that correlates with behaviour. Finally, some type theories have proved to be of little value in predicting future behaviour (Anderson & Lewis, 1998).

2.3.4.2 Trait Theories of Personality

While type theories suggest that people can be fitted into separate, discontinuous categories, trait psychology on the other hand propose hypothetical, continuous dimensions that vary in degree and quality (Anderson & Lewis, 1998). Trait theories are fundamentally based on the idea that individuals consists of a number of dispositions that cause them to respond in a consistent way across situations (Moyo, 2009). One of the most influential trait theories was proposed by Gordon Allport in 1937 (Anderson & Lewis, 1998). Allport perceived traits as the fundamental building blocks of personality and the source of individuality that are consistent in human behaviour. They are the enduring qualities which connect and unite a person's reactions to people and events (Anderson & Lewis, 1998). In 1936 Allport and Odbert identified sixteen thousand traits, and in 1937 Allport tried to reduce these traits into a more manageable number by dividing them into cardinal, central and secondary traits (Smith & Smith. 2005). He referred to cardinal traits as those general and enduring dispositions that affect most of human behaviour. They are moreover the dominant and pervasive factors in a person's personality. Central traits are more common dispositions with a more selective but still influence effect on an individual's behaviour in certain situations, while secondary traits are much more specific and express themselves in a few selected situations (Pervin, Cervone & John, 2005).

Another trait theory of personality was proposed by Eysenck, who developed a trait model that links types, traits and behaviours into a hierarchical structure of personality (Smith & Smith, 2005). Eysenck was one of the first factor analysts of personality that used exploratory factor analysis to reduce traits into a more manageable number (Smith & Smith, 2005). At first, he distinguished between two independent second-order factors of personality namely Extroversion-Introversion and Neuroticism-Stability, and later added a third factor, Psychoticism – Tough-Tender Mindedness to complete his three-order personality factors (Anderson & Lewis, 1998).

One of the most commonly known trait theorists is Raymond Cattell, who has proposed that personality can be structured according to sixteen traits (Anderson & Lewis, 1998). Cattell's aim was to construct a common taxonomy of personality traits within which the large numbers of qualities that make human beings individual and unique could be understood and managed in a simplified way (1943; 1965; 1979; 1990). Cattell analysed Allport and Odbert's list of traits, which he later narrowed down through the lexicon approach to obtain a multidimensional personality structure (John & Srivastava, 1999). In his aim to construct a common taxonomy of traits, Cattell used factor analysis to identify basic building blocks of personality that would provide such taxonomy (Pervin *et al.*, 2005). He found that human personality traits can be structured according to sixteen primary factors (Smith & Smith, 2005), which he perceived as the source traits of normal personality, because they are stable and determines an individual's consistent behaviour as well as the differences in surface traits[4] (Cattell, 1965). According to Cattell, each of the sixteen traits falls on a continuum, which means that each person contains a certain degree of every trait. Thus, an individual might be high in some traits but low in others (Heffner, 2011). To determine where on the continuum an individual falls, Cattell developed the self-descriptive Sixteen Personality Factor Questionnaire (16PF) as a measurement of the sixteen personality factors (Cherry, n.d.). Cattell further performed a second-order factor analysis on the sixteen personality traits and discovered five over-arching (second-order/global) personality factors that underlie the first-order factors (Sixteen Personality Factors, n.d.). These second-order traits are known as Extraversion, Anxiety, Tough-mindedness, Independence and Self-Control (Moyo, 2009). Moreover, the second-order factors represent common themes shared by specific primary factors that are derived from the relationships between the first-order factors (Foxcroft & Roodt, 2005). For example, the second-order factor, Extraversion contained primary factors such as Warmth (A+), Liveliness (F+), Social Boldness (H+), Privateness (N-), and Self-reliance (Q2-) (Moyo, 2009).

---

[4] Cattell (1950) distinguished between source and surface traits. Moreover, he describes them in Smith and Smith (2005, p. 38) as follow: "Source traits are the basic, fundamental aspects of personality that direct the style of behaviour that an individual chooses. They are often given strange names, which may have little direct connection with work behaviour, because the way in which traits are manifested depends on other factors, such as ability and motivation. Surface traits are produced by the interactions of source traits and other factors such motivations. They are the forms of behaviour that are observed. They are often given common-sense names, but they explain only a narrow band of behaviours. Surface traits may not be particularly stable."

Costa and McCrae (1985b; 1989; 1992; 1995) carried out similar factor analyses and proposed the Big Five personality dimensions (Moyo, 2009), which are labelled Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness to Experience (Smith & Smith, 2005). Since the Big Five were originally identified through the factor analysis of Cattell's first-order 16PF model, considerable similarity were found between Cattell's global factors and the Big Five dimensions identified by Costa and McCrae (1985; 1989; 1992; 1995). In the Fourth and Fifth Editions of the 16PF, Conn and Rieke (1994) also indicated that the second-order factors of the 16PF were closely related to the Big Five dimensions of personality.

As with type theories, trait theories have been criticised for not explaining how personality develops, or suggesting how behaviour is caused (Anderson & Lewis, 1998). Trait theories have also been criticised for being atheoretical, in that it merely labels and quantifies personality and does not help to psychologically understand personality dynamics (Smith & Smith, 2005). Another criticism of trait theories is that it portrays a rather stable image of personality, while failing to demonstrate consistency across situations (Anderson & Lewis, 1998). The theory assumes that personality traits are broad dispositions that regularly and persistently determine an individual's behaviour, feelings and thoughts in a variety of everyday situations (Moyo, 2009). However, personality traits do not express themselves across all situations in the same way (regardless of the nature of the situation). An individual's behaviour, thoughts and feelings will vary to some degree and quality across situations as a function of the demands perceived from the situation. Irrespective of these situational differences, personality traits generally express themselves in a consistent behavioural way across many situations (Pervin *et al.*, 2005). Anderson and Lewis (1998) conclude that it is necessary to distinguish personality traits from transient psychological states, for the reason that states will fluctuate more than traits. The PAPI is based on a trait interpretation of personality. These criticisms of the trait theory consequently need to be considered and explored with caution, when critically reflecting on the construction and use of the PAPI (Anderson & Lewis, 1998).

Furthermore, it is important to note that an emphasis on traits does not preclude the mobilisation of other personality elements (such as needs). Along with traits, Schmidt and Hunter (cited in Sanz et al, 2006) considered psychological needs to be leading factors that have to be taken into account in personality research attempting to psychologically explain behaviour. Winter, John, Stewart, Klohnen, and Duncan (1998, p. 232) distinguish between needs and traits, by referring to needs as 'people's goals and desires', while traits are defined as 'people's habitual patterns of cognition, affect and behaviour'. More specifically, traits are concerned with the question of *how* a person will behave, while needs answer the question of *why*, thereby reflecting two fundamental but different constructs of personality that both should be taken into account when considering the manner in which personality affects behaviour within the workplace (Winter et al., 1998). One of the most influential motivational theories in personality psychology was Murray's (1938) need-press theory, which provided the basis for the development of several measures such as the widely known Personality Research Form (PRF; developed by Jackson, 1984). The PAPI-N was also grounded in and heavily influenced by Murray's work (Anderson & Lewis, 1998).

Murray's Need-press Theory

Murray developed a model of personality which comprises of two main elements namely *needs* and *presses*, which he found to be a real source of conflict for the individual (Murray, 1938). To Murray human action was largely motivated by the desire to satisfy tension-provoking needs (Anderson & Lewis, 1998). He referred to these needs as hypothetical forces that influence and drive an individual's feelings, thoughts and behaviour. Murray further distinguished between primary needs, which are based upon biological demands such as the need for air, and secondary needs, which are psychological needs such as the need for affiliation (Anderson & Lewis, 1998). According to Murray, primary needs are essential in personality development because through the interaction with others and in the process of satisfying needs, the child learns to develop a conception of the world (Anderson & Lewis, 1998). Although such early childhood experiences determine how these primary needs are satisfied, it is the secondary needs or personality that develops as a result of these experiences. Furthermore, he believed that if a particular need is strong, much of an individual's energy and behaviour will be directed towards satisfying that need, while if a need is weak, it will give little energy and direction to a person's behaviour. For Murray (Anderson & Lewis, 1998), needs are therefore central to personality and can be described in terms of the following dimensions:

- ***Overt versus covert***: some needs are observable to all while some needs are repressed because they are perceived as threatening
- ***Focal versus diffuse***: some needs are capable of satisfaction by achieving a specific single goal while other needs are capable of satisfaction through the achievement of a variety of goals
- ***Proactive versus reactive***: some needs drive from within while other needs develop as a result of or response to an external event
- ***Modal versus effect***: some needs are expressed as activities that provide their own intrinsic pleasures while other needs are directed towards the achievement of a specific goal that has valence.
- ***Mutually supportive of each other versus in conflict with each other***: some needs may help in satisfying other needs, while some needs may be in conflict with other needs.

Murray recognised the difficulty inherent in identifying and organising an individual's needs, and in essence his model reflects the complexity of personality (Anderson & Lewis, 1998).

While Murray (1938) described *needs* as the 'internal determinants of behaviour', he referred to the term *press* as the external determinants of a person's behaviour originating from a person's perception of his/her environment (Anderson & Lewis, 1998, p. 51). He believed that people continuously operate in an environment which they evaluate in terms of the threats or promises that could either facilitate or obstruct their efforts to reach or avoid a specific need-satisfying goal. Presses could therefore either reinforce or weaken a person's need-based behaviour, and are thus situational characteristics that influence people to behave or act in certain ways. Murray further distinguished between those environmental forces that are merely perceived to be significant (beta press) but do not actually are consequential, and those that are very real in the effects they hold for the individual (alpha press) (Anderson & Lewis, 1998).

Further to this, Murray (1938, p. 750) also developed the concept of *thema*, which he referred to as a 'combination of a particular press and a particular need (press-need interaction)'. This helped him to describe behavior in a more global, less segmented way (Anderson & Lewis, 1998). To demonstrate the complexity of *thema*, consider the following example: an experience of *rejection* (p Rejection) might have reinforced a person's need for *abasement*[5] (n Abasement), which in turn caused him/her to exhibit increasingly *passive behaviour* or the desire to blame themselves. In addition, constant experiences of *rejection* may produce the need for *abasement* or activate an individual's need for *aggression*. On the other hand, *thema* could also be initiated by a need, for example a person's excessive need for *affiliation* (n Affiliation) causes him/her to behave inappropriately and as a result incites *contempt*, which acts as a catalyst for *rejection* (p Rejection) (Anderson & Lewis, 1998).

Ultimately, presses are able to facilitate, prevent or interfere with the occurrence and satisfaction of needs. Understanding this dynamic interaction between needs and presses can be of great value in a psychological sense as it can help with the identification of motivational trends and emotional issues such as conflict, anxiety and emotional change (Anderson & Lewis, 1998).

Murray's concept of *thema* in essence could be interpreted in terms of an *interactionist* approach as it suggests that to predict behaviour in any given situation, one needs to know about an individual's general personality traits and the objective and/or subjective meaning of the situation for the person concerned (Anderson & Lewis, 1998). Put another way, from an *interactionist* perspective, Murray's concept of *thema* can be reflected in this dynamic interaction of:

$$\text{Behaviour} = \text{Person} \times \text{Situation}$$

Thus in the context of Murray's model, an individual's personality (as manifested in their behaviour) is a function of the dynamic interaction between a person's *needs* and the *presses* associated with any given situation (Anderson & Lewis, 1998).

2.4 OVERVIEW OF THE DEVELOPMENT OF THE PAPI

The original PAPI was developed by Dr Max Kostick, a professor at the Department of Industrial Psychology at Boston State College, USA in the early 1960's (Anderson & Lewis, 1998). Kostick's primary aim was to design an instrument which is based on a sound

[5] Anderson and Lewis (1998, p. 49) describes n Abasement as: *To submit, surrender, admit inferiority*.

theoretical model, that comprehensively covers all aspects of personality relevant to the workplace but without the need to mobilise clinical terminology when interpreting the results, and which could be easy to use by non-psychologists[6]. In developing PAPI, Kostick recognised the work of numerous personality theorists, but considered the work of Murray to offer the most compelling theoretical base for personality measurement (Anderson & Lewis, 1998). The theoretical model on which the PAPI is based is rooted in Murray's need-press theory of personality. Murray's theory also provided the rationale for constructing PAPI with two distinct subtests. One scale measuring preferences, which Kostick named needs, and the other scale measuring perceptions, which he referred to as roles (Anderson & Lewis, 1998). The conceptualisation of Kostick's needs and roles can be mapped onto Murray's needs and presses respectively. Table 2.1 below demonstrates how the PAPI need scales reflect Murray's need classification system.

Table 2.1: *PAPI needs and Murray's need classifications system*

PAPI Needs	Murray's Needs
Need for rules and supervision (W)	n Passivity
Need to be noticed (X)	n Exhibition
Need to belong to groups (B)	n Affiliation
Need to relate closely to individuals (O)	n Succorance
Need to control others (P)	n Dominance
Need to be forceful (K)	n Aggression
Need to achieve (A)	n Achievement
Need to be supportive (F)	n Deference
Need to finish a task (N)	n Order
Need for change (Z)	-

Adapted from Anderson & Lewis (1998)

In accordance with Murray's model of personality the PAPI was thus designed to elicit the complex interaction between the two elements of needs and roles (Anderson & Lewis, 1998). Kostick, however, wanted to design a personality measure which specifically focused on the workplace and therefore some refinements were necessary in order to make the PAPI more relevant to this context. For example, the need to avoid pain, injury, and death (n Harm avoidance) would be invasive and have low face validity in the workplace. It was thus necessary for Kostick to identify which needs and significant environmental influences were important and relevant for a work-style measure such as the PAPI. However, it was also

[6] The latter aspiration would, however, have brought him in conflict with the Health Professions Council of South Africa (HPCSA) in as far as Act 56 of 1975 interprets personality assessment as a psychological act and reserves it to individuals registered with the HPCSA.

necessary for him to justify the use of a reduced number of needs and to provide some criteria for the inclusion into or exclusion from his measure. In his aim of identifying the relevant variables that would be meaningful to assess within an organisational context, Kostick conducted a survey of executives and managers. Ultimately the results of Kostick's survey provided the rationale for the inclusion or exclusion of needs and the identification of significant environmental influences (roles) (Anderson & Lewis, 1998).

Since Kostick's needs and roles were identified from a sample of the work population, the interaction between the two dimensions was considered to have a meaningful basis. The dynamic interaction between needs and roles on which the PAPI is based, represents the uniqueness of such a measure of personality in the workplace. This unique feature probably differentiates the PAPI from most self-report inventories, since many of them only focus on the dispositional factors that are consistent over time, while the PAPI seeks to understand human behaviour in terms of the interaction between the person and the situation, and consequently attempts to translate personality traits and situational or contextual information into actual behaviour. Finally, the PAPI adopts a fundamentally situational specific framework which embraces both needs and roles in order to understand the complexities of the dynamic interaction between personality x situation. In summary, the key features of the PAPI need and role scales are listed in Table 2.2 (Anderson & Lewis, 1998, p. 54).

Table 2.2: *Key features of the PAPI's need and role scales*

PAPI needs	**PAPI roles**
Ten need scales Expressed as preference statements (e.g. I like to do new things) Measure an individual's preference for behaving in a particular way based on what has gone before (i.e. an estimate of general tendency)	Ten role scales Expressed as perception statements (e.g. I always focus on the steps ahead) Measure an individual's self-perception of their behaviour in the work environment
Aim of needs is to achieve need satisfaction and avoid need frustration	Include situational or contextual characteristics that presses an individual to behave in a certain way (e.g. organisational culture, job demands, supervisor's management style, and life experiences outside work)
Widely held to be a measure of stable personality traits Those needs that are not or only partially satisfied may result in dysfunctional behaviour PAPI needs are closely related to Murray's needs on the Edwards Personal Preference Schedule	Do not measure transient psychological states, roles are reliably measured on PAPI

2.5 DEVELOPMENT OF PAPI-N

The PAPI-N questionnaire was developed from the original PAPI version designed by Kostick in the 1960's (Chong & Hughes, 2004). Anderson and Lewis (1998) indicated that Kostick originally developed the PAPI as an ipsative measure which was primarily designed for use as a counselling and discussion tool. In 1996, following on major revisions of Kostick's original instrument, two versions of PAPI were launched. The current versions of PAPI are, PAPI-I, which is an ipsative version, and PAPI-N, which is a normative version. The rationale for the revisions of Kostick's original instrument in 1996 arose from the following concerns.

Firstly, social changes have occurred since the time PAPI was first published in the 1960's, and as a consequence the scales and items may no longer have been measuring work domains in a manner that is meaningful to the current nature and context of employment. Further to this, changes in legislation (especially in the United States) placed new demands on ethical and responsible test usage, as well as demands for promoting equal opportunities in the workplace (especially amongst members of protected groups). Concern over the probable social desirability imbalance between PAPI statements existed, which could have serious consequences for the instrument's reliability and validity (Anderson & Lewis, 1998).

Secondly, the impetus for the revision programme arose from the acknowledgement that ipsative data required the application of statistical methods that were very different from those techniques traditionally used. There has been a growing feeling within the field of psychometrics that the data provided as statistical support for ipsative measures were inappropriate, because the underlying assumptions of these traditional methods did not fit an ipsative model. Principally, the argument was that scores on an ipsative measure cannot be considered in insolation from all other scores as they are all inter-correlated. In other words, it would be incorrect to compare scores with each other or with scales from other measures, rendering traditional methods of determining stability and validity meaningless. Ultimately this means that norming procedures and inter-individual correlations would be insignificant and inappropriate. Ipsative instruments should therefore not be used in situations where comparisons between individuals are required, i.e. selection.

These social and statistical issues were therefore considered important and provided the rationale for implementing a revision programme, which would include correcting social desirability imbalances and producing appropriate reliability and validity data [7]for the revised PAPI-I. Furthermore, a separate version of PAPI was designed as a response to overcome the psychometric criticisms of using an ipsative instrument in situations where individual comparisons are required (i.e. selection). Such applications were considered to be appropriately served by the development of a normative based measure. In summary, the objectives of the 1996 PAPI development programme, as cited by Anderson and Lewis (1998, p. 58), were thus to:

- Identify and correct social desirability issues impacting on the PAPI scales and items, including the adverse effect against members of previously disadvantaged groups and the impact of equity in the workplace.
- Evaluate the reliability and validity of PAPI using appropriate statistical techniques.
- Address concerns of current users that PAPI may be developed into a different and unrecognisable instrument.

2.5.1 FEATURES OF PAPI-N

PAPI-N is a normative questionnaire specifically designed to assess the most relevant needs in the workplace and individual's perceptions of themselves in terms of their behaviour at work (Sanz et al., 2006). The PAPI-N questionnaire consists of 126 items, comprising of 126 single statements, each accompanied by a rating scale. Respondents are asked to give ratings on a seven-point Likert type scale, ranging from 'absolutely agree' to 'absolutely disagree' (Anderson & Lewis, 1998). The PAPI-N measures 20 personality scales of which each scale comprises of 6 statements, with one additional scale added to measure social desirability, contributing 6 further items to the questionnaire (Anderson & Lewis, 1998). The 20 personality scales consist of the 10 need and 10 role scales described in Table 2.3. In the normative version of the PAPI the observed raw score on each of the 10 need and 10 role

[7] The intention of the development programme was to leave PAPI as close as to the original version as possible. Alterations were only made to the content where there was evidence to justify them. Prior to the development programme, an independent review of the English language version was undertaken in 1994. The review was conducted to evaluate PAPI from a statistical and theoretical point of view, as well as in terms of equal opportunities and potential for adverse impact. The qualitative research findings had a significant impact on the development of the revised PAPI. The most significant finding was the differences in views, attitudes and feelings between managers and support staff. The quality and nature of those differences were primarily responsible for the decision to develop PAPI as an instrument only for certain groups of the working population. Strengths and limitations of the PAPI were therefore identified, which provided a sound basis from which to plan the revision programme (Anderson & Lewis, 1998).

scales are transformed to a sten scale where 1-3 represents *low* scores, 4-7 represents *average* or *normal* scores, and 8-10 represents *high* scores (Anderson & Lewis, 1998).

Table 2.3: *PAPI-N Scales*

PAPI scale		***Low***	***High***
P	**Need to control others**	Prefers to let others take the lead, co-operative	Likes to influence, persuade and control others
L	**Leadership role**	Less confident in formal managerial role	Confident in ability to manage and lead others
C	**Organised type**	Disorganised, untidy. May be more flexible and adaptable	High degree of personal organisation, neat and tidy
H	**Integrative planner**	Spontaneous and reactive to external events	Tends to plan ahead and schedule well in advance
D	**Attention to detail**	Low attention to detail	High attention to detail
W	**Need for rules and supervision**	Non-conformist, prefers to use own initiative	High need for frameworks, instructions and guidance
R	**Conceptual thinker**	Practically orientated - focuses on day to day activities	Tends to be creative and theoretical in style of thinking
Z	**Need for change**	Low desire for change, tolerant of routine	High need for change and variety at work, restless
N	**Need to finish a task**	Low need to personally complete tasks, may prefer to delegate	High need to personally complete and see through tasks
X	**Need to be noticed**	Low need for recognition, prefers to stay in the background	High desire for personal recognition, enjoys the spotlight
B	**Need to belong to groups**	Self-reliant, independent	Strong need to work in, and be accepted by, teams
S	**Social harmoniser**	Unconcerned with being sociable at work, possibly task-orientated	Highly sociable and outgoing at work, enjoys networking
O	**Need to relate closely to individuals**	Distant, objective - low need to get close to others	Affectionate, trusting - likes to get close to others
I	**Ease in decision making**	Cautious, takes time when making decisions. May be indecisive	Tends to make very quick decisions, may be impulsive
T	**Work pace**	Careful, unhurried work style	Fast, enthusiastic worker, strong sense of urgency
K	**Need to be forceful**	Avoids conflict, dislikes confrontation	Forceful and direct, may tend to be confrontational
E	**Emotional restraint**	Tends to be open with both positive and negative emotions	Presents a calm and controlled exterior at work
A	**Need to achieve**	Low need for achievement at work, uncompetitive	High need for personal achievement at work, ambitious and competitive
F	**Need to be supportive**	Little need for praise from superiors, may question authority	Strong sense of loyalty towards immediate boss and organisation
G	**Role of the hard worker**	Not motivated by hard work in its own right	Motivated by and values hard work in itself

Adopted from PAPI-N Supplement User Guide (Cubiks, 2007)

Anderson and Lewis (1998) report that the following revision process was followed in the development of PAPI-N, as indicated in the technical manual:

- For the development of PAPI-N, the original PAPI items were analysed as part of the revision process. The results confirmed the need to revise the PAPI items. The two major criteria for including or rejecting a statement were if it was within the domain of the latent personality dimension that it was meant to reflect, and one that discriminates efficiently between scales by reflecting a genuine preference of the individual, who is completing the inventory, for one domain over another.
- Following the revision of the original PAPI scales, three of the scales were adapted for the new PAPI versions. The V-scale that assessed vigorous physical activities was excluded from the revised PAPI version, due to the fact that a (high) preference on this scale may well have reflected the ability to be physically active and this could have seriously distorted the profiles of those with physical impairments. Furthermore, in the light of heightened sensitivity to the need for equal work opportunities for the physically impaired, and the introduction of legislation in the UK to this effect, the scale was removed (Anderson & Lewis, 1998). The R-scale which assesses theoretical thinking also contained statements relating to the activity of thoughtful, integrative planning. Myers and McCaulley (1985) suggested that these are likely to be two separate domains. The scale was therefore refined specifically to working with creative, conceptual issues, and a new scale – assessing integrative planning (H) was developed. The other scales were also evaluated for domain homogeneity and the extent to which scale titles were for conceptually relevant given the nature of the domain.
- In the development of the PAPI-N statements, 120 of the 126 statements were taken from the original PAPI-I's 180 statements. The requirement was that the conceptual domain of each PAPI scale was represented by a sufficient number of statements to ensure stability for each of the scales. It was found that this number could be reduced to six whilst maintaining an acceptable level of stability. Thus the 120 statements taken from the 180 statements of PAPI-I were the six that had the highest factor loadings on each of the 20 conceptual domains.
- A Social Desirability scale was also included in the new PAPI-N, which makes up the remaining six items of the questionnaire. This allows for the psychometric verification

of an individual's responses, as it measures the extent to which an individual has attempted to present him/herself in a more favourable light.

- The seven-point Likert scale, ranging from 'absolutely agree' to 'absolutely disagree' was selected for the following reasons. Firstly, an odd number of categories allow respondents to choose a middle category, if that is a candidate's genuine response. Secondly, research by Matell and Jacoby (1972) and Potter (1995) as cited in Anderson and Lewis (1998) found that respondents are more likely to use the full range of scores available on a 7-point gradation when compared to 3 or 5 points; and lastly evidence by Rotter (1972) and Potter (1995) suggests that these range headings constitute the range of possible responses that are necessary to properly assess reactions to statements introduced to people regarding their work (Anderson & Lewis, 1998).

2.6 OVERVIEW OF THE DEVELOPMENT OF THE PAPI-N FOR SOUTH AFRICA

Since the PAPI-N was originally developed in the United Kingdom (UK), Cubiks[8] in collaboration with Work Dynamics[9] embarked upon a rigorous process of translation, standardisation, and validation of the PAPI-N English version to make it more relevant to the South African assessment context (Cubiks, 2007). The adapted South African English version of the PAPI-N was then translated into the main South African language groups: including Xhosa, Zulu, Sepedi, Sesotho and Afrikaans, to ensure that the items worked effectively for a South African working population, as indicated in the PAPI-N Supplement User Guide for South Africa (Cubiks, 2007). Subsequently, the PAPI-N items were reviewed by five South African subject matter experts (including one psychometrist and two psychologists) in terms of their suitability for the South African population, taking into account the many languages and different ethnic groups. Fifty eight items were adapted in the finalised South African version of the PAPI-N, whilst ensuring that the meaning and integrity of the items was maintained as close as possible to the original standardised PAPI UK version (Cubiks, 2007). Although a quantitative review of the items of the South African PAPI-N is necessary and

---

[8] Cubiks Group Limited is an international assessment and development consultancy. Cubiks was formerly the assessment and development practice of PA Consulting Group. In June 2007, Cubiks completed a management buy-out which granted them full independence and subsequently became the new owners of PAPI.

[9] Work Dynamics is a national human resource and management consultant company, and also the official distributor of Cubiks's products and services in South Africa.

critically important to achieve a construct valid measure of the personality construct as it is constitutively defined by the instrument developers it is not sufficient. Quantitative evidence is required that shows that the design intentions of the test developers succeeded. As described above the design intention was that specific items should reflect specific latent need or role dimensions of personality. The constitutive definition of personality along with the design intention implies a specific measurement model that maps latent personality dimensions onto individual items. The critical question is whether these single-group measurement models fit data obtained for each of the language groups. If at least a close fit is obtained the question in addition is whether the measurement model parameters can be considered satisfactory. If a single-group measurement model fit is achieved in the various language groups the question on the measurement invariance and equivalence of the different language versions of the South African PAPI-N also has to be examined.

Norms have been developed for the final South African English version by administering the instrument to a sample (N=555) of managerial and non-managerial staff, across a wide range of job functions and corporate sectors and transforming the obtained raw score distribution to a sten scale. The sample represented respondents in the following industries as indicated in the PAPI-N Supplement User Guide (Cubiks, 2007): Retail, Medical and Pharmaceutical, Financial, Education, Manufacturing, Government, Transport, Part-time students in service industries, Engineering and Telecommunications.

2.7 AVAILABLE FORMATS OF THE PAPI-N

The PAPI-N Questionnaire is available for both traditional paper and pencil and computerised administration through the internet-based online version. For the paper and pencil version, question booklets and self-scoring answer sheets are available. Both versions can be administered individually or in a group context by a HPCSA registered psychologist or psychometrist or a trained person under the supervision of a registered psychologist (Cubiks, 2007). The administration of both the online and paper and pencil versions is guided by the detailed, standardised instructions in the technical manual of PAPI (p. 107-110 and 115, 1998). Both formats should be administered under standardised conditions and it is essential that respondents are guided by clear and consistent instructions with regards to the purpose of the assessment. When an online assessment has been completed, administrators can immediately generate a participant's report. If the participant has completed a paper and pencil version of PAPI-N, the administrator can manually enter the participant's scores into Cubiks Online, or alternatively, follow the self-scoring instructions in the Technical Manual (p. 123 – 128). Furthermore, other assessment materials such as profile charts, norms and interpretative guides are available for the PAPI-N questionnaire. For feedback on a respondent's PAPI Profile the following computerised reports are available: Narrative Report, Hypothesis Report, Job Profiler Interview Guide, Respondent Feedback Report, Wheel report and a Group Report (Cubiks, 2007). Detailed information regarding the interpretation, feedback, and report writing is available in the technical manual (p133-160, 2007).

2.8 RELIABILITY OF THE PAPI-N MEASURES

Foxcroft and Roodt (2005) define the reliability of an instrument as the consistency with which it measures what it is intended to measure. Moreover, Anderson and Lewis (1998) refer to reliability of a psychological questionnaire as "the measurement error within the instrument, including its administration and scoring" (p.85). Reliability refers to the proportion of systematic variance in the observed scores (Nunnally, 1978). There are many factors that could influence the reliability of a measure, which thereby produces unwanted random measurement error (Anderson & Lewis, 1998). Thus when a test produces consistent measures under consistent conditions (Foxcroft and Roodt, 2005) or when observed measures are relatively free of random measurement error those measures can be considered reliable (Kaplan & Saccuzzo, 2001). Anderson and Lewis (1998) indicated that in the assessment of the PAPI-N's reliability, it is important that the PAPI-N should not be considered as two different instruments but rather as an instrument which comprises of 20 scales. Each of the 10 need scales and each of the 10 role scales should be treated as separate measures that only come together in the interpretation of the results because of the interdependency between the latent personality dimensions measured by the scales. This section will discuss the reliability findings for the PAPI-N measure as reported in literature by Cubiks, the technical manual and other research studies. Cronbach's alpha coefficients were mostly used as a measure of internal consistency reliability.

The PAPI Technical Manual reported good internal consistency for each of the 20 PAPI-N scales for a UK sample, with alpha coefficients exceeding .70 (Anderson & Lewis, 1998).

These reliability findings for PAPI-N were obtained on a general management sample (N=164), across a wide range of corporate sectors in the UK (Anderson & Lewis, 1998). Additionally, each of the scales has been tested in order to establish PAPI-N's stability over time. Anderson and Lewis (1998) used the test-retest method, with a time interval of two to three weeks to indicate stability levels. This was also based on a general management sample (N=100), across a range of corporate sectors. Table 2.4 presents the reliability findings on each of the 21 scales (including the social desirability scale) as reported by Anderson and Lewis (1998).

Table 2.4: *PAPI-N Internal consistency and stability reliability coefficients: UK General management sample*

Scale	Internal consistency Cronbach's Alphas (N=164)	Scale	Test-retest Reliability coefficients (N=100)
G	.82	G	.83
P	.72	P	.62
L	.86	L	.81
C	.86	C	.89
H	.82	H	.80
D	.85	D	.86
W	.85	W	.88
R	.77	R	.86
Z	.82	Z	.84
N	.88	N	.91
X	.85	X	.68
B	.83	B	.84
S	.81	S	.83
O	.76	O	.88
I	.90	I	.88
T	.89	T	.87
K	.71	K	.62
E	.87	E	.91
A	.79	A	.85
F	.78	F	.83
SD	.80	SD	.80

Adopted from Technical Manual (Anderson & Lewis, 1998)

As indicated above, the internal consistency of all 20 scales exceeded .70 but none so high that they would cause concern with regards to statement-repetition or that items are too similar to each other. Excessively high internal consistency reliability coefficients raise the concern that essentially the same question was asked a number of times (Chong & Hughes, 2004). It can be seen from the test-retest reliability results that PAPI-N indicated an acceptable level of stability (Anderson & Lewis, 1998). Furthermore, Cubiks (2007) reported reasonably acceptable coefficient alphas for the total/general group, including managerial and non-managerial groups, with five exceptions (see below). Table 2.5 presents the reliability coefficients of each of the scales for the different groups.

Table 2.5: *Reliability coefficients for the PAPI-N scales on a South African sample (N=555)*

Scale	General Cronbach's Alphas	Managerial Cronbach's Alphas	Non-Managerial Cronbach's Alphas
G	.685	.608	.715
P	.639	.614	.601
L	.722	.680	.699
C	.758	.797	.744
H	.749	.819	.731
D	.723	.799	.689
W	.746	.820	.712
R	**.299**	**.125**	**.379**
Z	.698	.809	.685
N	.745	.820	.731
X	.732	.686	.743
B	.767	.757	.766
S	.618	.650	.581
O	**.594**	.681	**.561**
I	.750	.730	.755
T	.764	.752	.775
K	**.391**	**.429**	**.375**
E	**.530**	.639	**.495**
A	**.560**	**.591**	**.576**
F	.657	.720	.640

The results on the Social Desirability scale was also reported in the PAPI-N Supplement User Guide by Cubiks (2007) for the three groups: general (α = .80), managerial (α = .810) and non-managerial (α = .806). The results presented in Table 2.5 indicated that the following five scales (highlighted in bold) needed to be reviewed as the aim (cut-off) for each scale was set at a minimum Cronbach alpha of ≥ .6: Conceptual thinker (R); Need to relate closely to individuals (O); Need to be forceful (K); Emotional restraint (E); and Need to achieve (A) (Cubiks, 2007). Another study by Sanz et al. (2006), reported that alphas for 12 of the Spanish PAPI-N scales were higher than .80, while the remaining scales had alphas that were above the standard level of .70, but with two exceptions: R (alpha=.66) and K (alpha=.41).

There is still some controversy and confusion on the question of what an acceptable level is or how high a reliability coefficient of a measure should be (Moyo, 2009). Foxcroft and Roodt (2005) argued that the answer to this question depends largely on what the measure intends to be used for. Clark and Watson (1995) indicate that an acceptable level of reliability should be above .60, while other researchers considers .70 to be an acceptable level of reliability for the reason that in the case of reliability coefficients below .70 the standard error of measurement is over a half (.55) of a standard deviation of the test score (Moyo, 2009). Nunnally (1978) believes that alpha coefficients should be quite high and not be below .70. Test developers usually set high standards for reliability coefficients that falls within the .70 to .90 range, neither too low nor too high (Nunnally, 1978). According to Anastasi and Urbina (cited in Foxcroft & Roodt, 2005), standardised measures should have reliability coefficients between .80 and .90. Murphy and Davidshofer (2005, p. 149) pointed out that reliability of tests may be crucial in some situations but less important in other settings. When a test is used for preliminary rather than for making final decisions and/or used to sort individuals into small categories on the basis of gross individual differences, lower reliabilities (estimates around .70 are usually regarded as low) would be acceptable according to Murphy and Davidshofer (2005). Higher reliabilities are required when an instrument is used for making critical decisions and when individuals are grouped into many different categories based upon fairly small individual differences (e.g. intelligence measures) (Murphy & Davidshofer, 2005). According to Foxcroft and Roodt (2005), reliabilities for standardised personality and interest questionnaires should be .80 to .85 while measures of intelligence or aptitude should be .90 or higher. With reference to the PAPI-N, Anderson and Lewis (1998) argued that if alpha coefficients are too low, the scale would most likely contain very diverse and ambiguous items. However, in striving to achieve high alpha coefficients some researchers believe that this could lead to the measurement of rather narrow and psychological trivial variables. Thus if a measure focuses on a very narrow trait breath, the instrument can lack validity for assessing broader personality traits (Anderson & Lewis, 1998).

Consequently, a generally acceptable rule of thumb for describing reliability coefficient was introduced by George and Mallery (2003) who interpreted Cronbach's alpha coefficients as shown in Table 2.6.

Table 2.6: *General rule of thumb for describing reliability coefficients*

Cronbach's alpha	**Reliability coefficient**
$\alpha \geq .90$	Excellent
$.90 > \alpha \geq .80$	Good
$.80 > \alpha \geq .70$	Acceptable
$.70 > \alpha \geq .60$	Questionable
$.60 > \alpha \geq .50$	Poor
$.50 > \alpha$	Unacceptable

The present study will set an alpha coefficient of .70 as the critical cut-off value when interpreting the results of the item analysis of the scales in the study of PAPI-N's reliability.

2.9 VALIDITY OF THE PAPI-N MEASURES

All psychometric tests are designed to make inferences about individuals' standing on constructs or latent variables (Van Der Merwe, 2005). When making inferences, it could either be on the construct being measured or on another construct that is systematically related to the construct being measured. Decisions are made about individuals based on the inferences on the constructs derived from the measures. Validity refers to the permissibility and accuracy of these inferences. Validity is a critical aspect of the psychometric evaluation of a psychological measuring instrument. Murphy and Davidshofer (2005) indicated that should the measures of an instrument fail to demonstrate reliability, it would have implications for both the validity of the inferences derived from the measures and the decisions based on the inferences. A test that provides reliable measures, does, however, not necessarily mean that valid inferences can be derived from the measures. It is therefore essential that a psychometric instrument such as the PAPI-N must be able to demonstrate that its measures are reliable and that the inferences derived from the observed scores on the personality construct, as it is constitutively defined by the PAPI-N, are valid as well as inferences on criterion constructs deemed relevant to decision-making (Anderson & Lewis, 1998). This section will discuss available evidence supporting the validity of construct and criterion inferences derived from the PAPI-N measures as reported in literature.

2.9.1 CRITERION-RELATED VALIDITY

The technical manual of the PAPI-N (Anderson & Lewis, 1998) indicated that the criterion-related validity of PAPI-N was established on a sample of 40 managers as to determine whether the test scores were able to discriminate between the various types of behaviour (adequate versus inadequate work performance) either in terms of the present (i.e. concurrent criterion validity) or in terms of the future (i.e. predictive validity). The predictive validity results showed correlations up to $r = .42$ ($p < .05$) (Anderson & Lewis, 1998). In addition, Cubiks (2003) conducted a criterion-related concurrent validity study to establish the PAPI-N's ability in predicting sales performance for sales advisors within a Netherlands insurance company. This was a validity study only, no reliability results were reported in the study. The PAPI-N scales that were identified as potentially being related to sales performance were based on a sale manager's expertise and research literature. The job performance criterion for this research study was the respondents' sales performance figures for the year 2002. The results are indicated in Table 2.7 below.

Table 2.7: *Relationship between selected PAPI-N scales and sales performance (n=70)*

PAPI scale	**Correlation with sales performance**
G Role of the hard worker	.42**
A Need to achieve	.37**
T Work pace	.32**
P Need to control others	.31**
X Need to be noticed	.29*
N Need to finish a task	.27*
K Need to be forceful	.24*
B Need to belong to groups	-.18
S Social Harmoniser	.14
O Need to relate closely to individuals	.00

**p<.01, *p<.05

Adopted from Cubiks (2003)

The relationship between the PAPI-N scales and sales performance was determined with Pearson's correlation coefficient. The correlations in Table 2.7 indicate that 7 out of the 10 selected PAPI-N scales significantly (p<.05) relate to sales performance, which suggests that there is a relationship between PAPI and sales performance. A stepwise multiple regression analysis was conducted to investigate the predictive relationship between the specific PAPI-N scales and sales performance. The results obtained on the optimum regression model was significant (F=14.441, p<.05), with only 3 of the 7 scales (G, X, and B) emerging as predictor variables that significantly (p<.05) explain unique variance in sales performance. The weighted combination of these three PAPI-N scales explained more than 25% of the variance in insurance sales performance. This research study indicates that PAPI-N can be a useful assessment tool for predicting performance in a sales context, and further provides supportive evidence on how personality traits can have an impact on job performance (particularly in sales performance) (Cubiks, 2003). Although not many criterion-related validity studies have been conducted on the PAPI-N, the few studies that have been reported in literature indicated adequate to reasonable levels of criterion-related validity for the PAPI-N (Felthan & Hughes, 1999).

Paragraph 2.9.1 discussed validity from the perspective of the extent to which inferences about a criterion construct are permissible from measures on the PAPI-N scales. Paragraph 2.9.2 will discuss the validity from the perspective of the extent to which inferences about the personality construct as constitutively defined by the PAPI (Anderson & Lewis, 1998) are permissible from measures on the PAPI-N scales.

2.9.2 CONSTRUCT VALIDITY

Psychologists have a great interest in measuring abstract attributes, however because constructs themselves are abstract in nature, it makes the process of determining whether the test provides an accurate measure of a specific construct difficult (Murhpy & Davidshofer, 2005). A construct holds two essential properties, they are an abstract summary of some regularity in nature and they are related to or connected with concrete, observable entities or events (Murphy & Davidshofer, 2005). A test provides a measure of a specific construct by translating the abstract construct into concrete, behavioural terms. This process refers to construct explication which provides a definition of how a construct relates to a number of behaviours. These behavioural denotations are used to measure the construct. In the case of personality measures the behavioural denotations of the various personality dimensions serve as stimuli to which test takers have to respond by indicating to what extent the behaviours are applicable to them based on historical recall of their own recent behaviour. Construct explication thus plays a vital role in determining a test's construct validity since it depends on a detailed description of the relationship between a specific construct and a number of behavioural denotations of the construct used as items (Murphy & Davidshofer, 2005).

According to Murphy and Davidshofer (2005), the aim of construct validity is to determine whether the test scores on an instrument provides a good indication of a specific construct. To validate inferences about a construct calls for a demonstration that a test measures the specific construct on which information is required for decision-making. According to Cascio (1998) there are various ways to gather evidence relevant to a measure's construct validity, namely (a) an analysis of the internal consistency of the measurement procedures, (b) convergent and discriminant validation, (c) correlations with established measures of the same construct, (d) factor analysis of a group of procedures indicating which of them share a common variance and thus measure the same construct, and (e) co-variance structure modelling.

Neuman (2000) indicated that when conducting an exploratory factor analysis the results will indicate how well the items relate to the underlying factors comprising the construct and thus indicate whether the items load on one or more factors. Anderson and Lewis (1998) performed a principal component factor analysis on the 20 aggregate scales (ten roles and ten needs) of the UK English PAPI-N version and obtained seven factors, which they labelled Active Dominance (PAPI-N scales: P and L), Conscientious Persistence (C, H, D and W), Openness to Experience (N, Z, and R), Sociability (B, O, X, and S), Work Tempo (T and I), Agreeableness (K and E), and Seeking to Achieve (A, F, and G). Cubiks (2007) also performed a principal axis factor analysis to establish PAPI-N's construct validity in South Africa and found 5 out of the 7 second-order factors with eigenvalues greater than 1. They are Conscientious Persistence, Seeking to Achieve, Active Dominance, Work Tempo, and Openness to Experience. Sociability and Agreeableness showed to combine with two factors namely Seeking to Achieve and Agreeableness. The relatively low reliability index on these scales indicated the need for further item analysis as mentioned in the PAPI-N Supplement User Guide (Cubiks, 2007).

As an essential part of construct validity, it is important to determine whether the hypothesised latent dimensions of the personality construct relate to a similar and meaningful structure of personality as empirically identified by other researchers (Anderson & Lewis, 1998). The Technical Manual compiled by Anderson and Lewis (1998) reported on PAPI-N's construct validity and found strong correlations between the PAPI scales and the Five Factor Theory dimensions, as four of its factors supported the five structure model of personality (Chong & Hughes, 2004). Cubiks also reported on relationships between the PAPI-N and the Fundamental Interpersonal Relations Orientation – Element B (FIRO-EB) (Cubiks, 2004), as well as a measure of the Big Five known as the Trait Descriptive Adjectives (TDA-100) (Cubiks, 2006). Correlations between the PAPI-N scales and the TDA scales are presented in Table 2.8, and the relationships between the PAPI-N scales and the FIRO-EB scales are reported in Table 2.9.

Table 2.8: *Correlations between PAPI-N scales and TDA scales on a sample of 65 employees within a telecommunication and business services sector.*

	TDA Scale				
PAPI-N Scale	**Surgency**	**Agreeableness**	**Conscientiousness**	**Emotional Stability**	**Intellect**
P	.29*	-.23	.06	.05	.40**
L	.41**	-.04	.08	.08	.45**
C	-.32**	.09	.70**	.11	.11
H	-..09	.08	.54**	.08	.44**
D	-.28*	-.20	.57**	.18	.25
W	-.15	.08	.26*	-.10	-.37**
R	.34	.05	.09	.20	.75**
Z	.38**	.03	-.09	.17	.49**
N	-.23	-.02	.50**	.12	.28*
X	.66**	.10	-.37**	.00	.37**
B	.24	.03	-.08	-.23	-.09
S	.49**	.48**	-.21	-.05	.11
O	.31*	.32*	-.19	-.26*	.14
I	.37**	.03	-.18	.16	.25*
T	.17	-.012	.15	.11	.23
K	.42**	-.18	-.07	.02	.43**
E	-.18	.07	.14	.36**	-0,10
A	.32**	-.26*	-.05	-.14	.38**
F	.13	-.01	.09	-.13	-.02
G	-.12	.03	.42**	.26*	.25*

** Correlation is significant at the .01 level (2-tailed)

Adopted from Cubiks (2006)

As indicated above, there were a considerable number of statistically significant ($p<.05$) relationships found between the PAPI-N and the TDA scales. Based on the hypothesised links (refer Anderson and Lewis (1998) as reported on page 41 in this thesis) between the PAPI-N and the TDA scales, 10 out of the 13 showed significant ($p<.05$) correlations with 7 correlating at or above the .01 level of significance. Further to this study, an additional 24 correlations were found between the PAPI-N scales and the TDA scales. As a second part of this study, the relationship between the PAPI and an additional measure of the Big Five, namely the NEO-FFI was also researched. From this study, 44 significant correlations were found between the PAPI-N and the NEO-FFI scales, where 22 of the relationships overlapped with a single PAPI-N scale for both the TDA and the NEO-FFI measures of a specific Big Five factor. These findings indicated that there were a relationship between the PAPI and the well-established factors of the Big Five (Cubiks, 2006).

Table 2.9: *Relationship between PAPI-N scales and FIRO-EB scales (N=47)*

PAPI Scale	Major FIRO-EB correlations
P Need to control others	I want to control people (.62), I control people (.57)
L Leadership role	I control people (.55), I want to control people (.47)
W Need for rules and supervision	I include people (.45), I want to include people (.39)
X Need to be noticed	I want to control people (.58), I control people (.42), I am open with people (.39), I include people (.36), I want to be open with people (.36)
B Need to belong to groups	I include people (.50), I want to include people (.47)
S Social harmonizer	People include me (.62), I include people (.61), People are open with me (.55), I am open with people (.49), I want to be open with people (.35)
O Need to relate closely to individuals	I include people (.46), I want to include people (.45), People include me (.38)
I Ease in decision making	People control me (-.39), I want people to control me (-.35)
K Need to be forceful	I control people (.43)
A Need to achieve	I want to include people (.46), I want people to include me (.46), I control people (.39), I want to control people (.36)

All correlations statistically significant at p<.05 level or lower

Adopted from Cubiks (2004)

A further study was conducted to investigate the relationship between the PAPI-N and the FIRO-EB, a well-established measure of personality which specifically focuses on interpersonal behaviour essential in social interactions. The 12 FIRO-EB scales that most prominently correlated (r>.35) with the PAPI-N scales are showed in Table 2.9. Since the PAPI-N scales comprise of task-based items, it was not expected to correlate with the FIRO-EB's people- and internally-oriented scales. This was especially true, as the task-based scales showed little correlations on the internal- and people-oriented scales. A number of correlations were found between PAPI-N's sociability scales (X, B, S and O) and the interpersonal people-oriented scales of the FIRO-EB, as well as between the FIRO-EB scales that are concerned with controlling other people and the PAPI-N's dominance scales, P and L. Ultimately, the results of this study found evidence supporting the construct validity of PAPI-N, specifically for the scales that are related to interpersonal relationships (Cubiks, 2004).

The PAPI-N attaches a specific constitutive definition to the personality construct. The design of the PAPI-N earmarked specific items to reflect specific latent personality dimensions. The constitutive definition taken in conjunction with the design of the PAPI-N defined a specific measurement model. Construct validity would be indicated if a measurement model describing the manner in which items are hypothesised to load on the latent construct dimensions fits data on the instrument at least closely, the factor loadings are statistically significant and large and the measurement error variances are statistically significant and small. In terms of the connotative meaning that the PAPI-N attaches to the personality construct the personality construct is embedded in a larger nomological network of constructs in a specific manner. Construct validity would be indicated if the structural model reflecting the manner in which personality is embedded in the nomological net fits data on the instrument at least closely, and the freed γ_{ij} and β_{ij} paths in the model are statistically significant ($p<.05$).

Although there is reasonable evidence internationally that supports the PAPI-N's construct validity sophisticated psychometric evidence is still lacking. The fit of the measurement model implied by the constitutive definition of the personality construct in conjunction with the architecture of the PAPI-N has not been evaluated by means of confirmatory factor analysis (CFA). Neither has the fit of a fully-fledged structural model been evaluated that maps the first-order personality factors onto latent variables they are conceptually meant to be related to.

In addition, there is little empirical construct validity evidence available for the PAPI-N in South Africa. The tentative conclusion that the PAPI-N provides a construct valid measure of personality will have to be tested to determine whether it is psychometrically appropriate for, and relevant to, the South African work context.

2.10 SUMMARY

This chapter explained the process followed in the construction of the PAPI-N, clarified the constitutive definition underlying the PAPI and evaluated the success with which it measures the personality construct within a working environment. The following chapter will discuss the research methodology which includes the following: the research problem and research hypothesis, the research design, statistical hypothesis and analysis, sample design, and the measuring instrument.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 INTRODUCTION

The first chapter of this study formulated and motivated the research objective by means of a detailed, systematic and reasoned argument as to why there is a need for a close psychometric inspection of the PAPI-N as a measure of personality in the workplace. The intention of this study is to evaluate the construct validity of the PAPI-N through a factor analytic investigation. More specifically a confirmatory factor analysis was undertaken into the first-order factor structure of the PAPI-N to determine whether all the items reflect the latent personality dimensions they were designed to reflect within the group to be studied. The end result of this process is to substantiate the use of this instrument as a measure of personality in the South African workplace.

3.2 RESEARCH METHOD

The PAPI-N is based on a specific interpretation of personality. The architecture of the instrument reflects a specific design intention. In conjunction with its design and the connotative meaning the PAPI-N attaches to the personality construct, the scoring key denotes a specific measurement model which suggests that responses to specific items of the PAPI-N are a function of a specific underlying latent personality dimension. The PAPI-N items are designed to function as stimulus sets to which applicants respond by describing the degree to which they agree/disagree that the behavioural description are typical of them. The assumption is that these behavioural responses on the 7-point Likert scale express their standing on a specific underlying latent personality dimension. The measurement model maps those specific items believed to reflect a specific first-order personality dimension onto that latent personality dimension. To determine whether these claims made by PAPI-N are valid, a confirmatory factor analysis in which the fit of the implied measurement model is evaluated is required.

Evaluating the fit of the measurement model implied by the constitutive definition of the personality construct and the PAPI-N scoring key requires objective specific methodology. It is important to note that the credibility of the findings on the validity of these claims rest on the methodology being used. The methodology is therefore meant to serve the epistemic ideal of science (Babbie & Mouton, 2001). If the methodology would be flawed the chances of the researcher to arrive at a valid verdict on the merit of the measurement model as a hypothesis on the nature of how the construct is measured by the PAPI-N would be jeopardised. Consequently, the conclusions derived on PAPI-N's ability to measure the personality construct amongst South African employees through its intended design, could be flawed and seriously harm the credibility of the verdict on the merits of the PAPI-N as a measure of personality within the work environment. It is further important to note that one cannot expect the interested parties/scientists to simply accept the verdict concluded in the study at face value without any insight in the methodology used to reach the verdict. Credible and valid claims will more likely be achieved if the methodology is described in great detail and is made open to the evaluation and scrutiny of the scientific community where methodological flaws can be detected and corrected (Babbie & Mouton, 2001). If this is not done, the rationality of science will suffer, as will eventually also the epistemic ideal of science (Babbie & Mouton, 2001). The rationality of science can, however, only serve the epistemic ideal of science if the methodology used is carefully described and methodically motivated. The next section will therefore systematically discuss the research methodology, which includes the research problem, research hypothesis, research design, statistical hypothesis, sampling, the measuring instrument and the statistical analysis techniques.

3.3 RESEARCH PROBLEM

Several international studies (e.g. Sanz et al, 2006) have found psychometric evidence that supports PAPI-N as a reliable and valid instrument across various industries and settings. In terms of South African studies, little research has been conducted to confirm the reliability and construct validity of the PAPI-N. Moreover, none of the studies on the psychometric integrity of the PAPI-N evaluated the fit of the measurement model on a relatively large sample of the South African working population. Nevertheless, the instrument is still used to assess personality for various positions across different industries and occupations in South

Africa. This lack of research therefore necessitates an investigation into the construct validity of PAPI-N as a measure of personality within the South African workplace context.

The research problem is the question whether the PAPI-N provides a reliable and construct valid measure of personality as constitutively defined by the instrument within the multi-cultural South African work context.

3.4 SUBSTANTIVE RESEARCH HYPOTHESIS

The substantive hypothesis tested in this study is that the PAPI-N provides a construct valid and reliable measure of personality, as constitutively defined by the instrument, within a multi-cultural South African work context.

The substantive hypothesis translates into the following specific operational hypotheses:

- The measurement model implied by the scoring key and the design intention of the PAPI-N can closely reproduce the co-variances observed between the items comprising each of the PAPI-N scales[10];
- The factor loadings of the items on their designated latent personality dimensions are statistically significant ($p<.05$) and large ($\lambda_{ij}\leq.50$);
- The measurement error variances associated with each item are statistically significant ($p<.05$) but small,
- The latent personality dimensions explain large proportions of the variance in the items that represent them ($\lambda^2_{ij}\geq.25$); and
- The latent personality dimensions correlate low to moderate ($\phi_{ij}<.90$) with each other (i.e., the PAPI-N latent personality dimensions display discriminant validity).

3.5 RESEARCH DESIGN

The objective of this study is to empirically investigate the first-order factor structure of the PAPI-N as a psychological measure used within the South African workplace via confirmatory factor analysis. Further to this, the objective of the proposed research is to

[10] The social desirability subscale is included in the analysis.

contribute to the psychometric credentials of PAPI-N's ability to measure personality within a South African work context. More specifically, the research objective is to contribute to the investigation of the extent to which it is permissible to use the PAPI-N as a measure of personality amongst South African employees across various industries and occupational settings.

In order to pursue the research objective, the operational research hypothesis as mentioned in the previous paragraph needs to be tested. Therefore the validity of the hypothesised relationships between latent personality dimensions and observed indicator variables are to be investigated empirically. It is, however, not suggested that a single study of this nature will allow for a conclusive verdict on the construct validity of the PAPI-N as a measure of personality within the South African workplace context. To achieve a comprehensive investigation into the construct validity of the PAPI-N will also require the explication of the nomological network in which the personality construct is embedded and confronting the resultant structural model with empirical data. It is therefore not implied that if satisfactory measurement model fit would be obtained in this study that the PAPI-N would be indisputably cleared for use as a valuable assessment tool in South Africa. Neither is it claimed that convincing evidence on the construct validity would be sufficient evidence to justify the use of the PAPI as a selection instrument. Lack of measurement model fit would, however, seriously corrode confidence in the construct validity of the instrument and would raise questions on the use of this instrument as a predictor in personnel selection in South Africa.

To empirically investigate the merits of the stated operational research hypothesis requires a plan or strategy that will guide the gathering of empirical evidence to test validity of the hypothesised relationships. This empirical evidence providing strategy is known as the research design (Kerlinger, 1973; Theron, 2009a). The function of the research design is to try and ensure empirical evidence that can be interpreted unambiguously for or against the stated operational hypotheses including the control of variance (Kerlinger, 1973; Theron, 2009a).

The measurement model implied by the scoring key of the PAPI-N hypothesises specific measurement relations between the items comprising the instrument and the personality dimensions they were earmarked to represent. More specifically the measurement model assumes that the slope of the regression of specific items on the specific latent personality dimension the item is meant to represent is positive and statistically significant. In addition the measurement model assumes that the 20 latent personality dimensions are correlated[11]. The measurement model finally assumes that the measurement error terms are uncorrelated. To empirically test the merits of these assumptions made by the measurement model requires a strategy.

An *ex post facto* correlational design will be used. To empirically test the assumptions made by the measurement model, in terms of the logic of the *ex post facto* research design, the researcher obtains measures on the 126 PAPI items and calculates the inter-item co-variance matrix. Estimates for the freed measurement model parameters are obtained in an iterative fashion with the purpose of reproducing the observed inter-item co-variance matrix as accurately as possible (Diamantopoulos & Siguaw, 2000). If the fitted model fails to reproduce the observed co-variance matrix accurately (Byrne, 1989; Kelloway, 1998) the conclusion unavoidably follows that the measurement model implied by the PAPI-N scoring key does not provide an acceptable explanation for the observed inter-item co-variance matrix. This would than mean that the PAPI-N does not measure the personality construct as intended in the South African sample. The converse, however, is not true. If the reproduced/fitted co-variance matrix derived from estimated measurement model parameters closely corresponds to the observed inter-item co-variance matrix it does not necessarily imply that the processes postulated by the measurement model must have produced the observed co-variance matrix and that the PAPI-N therefore necessarily measures the personality construct as intended. A high degree of fit between the observed and estimated inter-item co-variance matrices would only suggest that the processes described in the measurement model provide one plausible explanation for the observed co-variance matrix. The claim that the PAPI-N provides construct valid measures of the personality construct as this instrument defines the construct would thereby have survived an opportunity to be refuted (Popper, 1972).

[11] It could not be determined from the PAPI-N manual how the PAPI-N expect the SD scores to correlate with the 20 latent personality dimensions.

3.6 STATISTICAL HYPOTHESES

The nature of the envisaged statistical analyses that will be used to test the operational hypotheses will inevitably affect the decision as to whether statistical hypotheses should be formulated and the format in which they will be formulated. For example, if an unrestricted exploratory factor analytic approach would have been used no statistical hypotheses would have been formulated, since there is no *a priori* stance on the number of factors underlying the observed co-variance matrix, their identity or the pattern with which the items load on the factors (Ferrando & Lorenzo-Seva, 2000). This option would, however, have been inappropriate in that it ignores the fact that the developers of the PAPI-N worked from a specific constitutive definition of personality and had specific design intentions on how the PAPI-N items should reflect the latent personality dimensions comprising personality.

In the case of the PAPI-N, a very specific stance is taken on the number and identity of latent personality dimensions underlying the observed inter-item co-variance matrix and the manner in which the items load on the personality factors. Operational denotations were explicitly and intentionally developed to reflect specific dimensions of the personality construct. Specific PAPI-N items were written to function as stimulus sets to which respondents would respond with behaviour which would be a relatively uncontaminated behavioural expression of a specific latent personality dimension. The PAPI-N scoring key reflects these design intentions.

It seems more reasonable towards the developers of the instrument to first evaluate the question whether their intended operational design succeeded in providing a comprehensive and relatively uncontaminated empirical grasp on the personality construct as defined. A hypothesis testing, restricted, confirmatory factor analytic approach should therefore rather be followed. In terms of this approach specific structural assumptions are made with regards to the number of latent variables underlying the PAPI-N, the relations among the latent variables and the specific pattern of loadings of indicator variables on these latent variables (Ferrando & Lorenzo-Seva, 2000; Jöreskog & Sörbom, 1993). The measurement model reflecting these structural assumptions is shown in Equation 1 (p. 56).

To the extent to which a measurement model reflecting these assumptions would fit empirical data poorly, the measurement intention of the test developers would have failed. If the verdict would go against the measurement claims of the test developers, given that they have been given a fair hearing, it would seem more justified to use an unrestricted, exploratory factor analytic approach to estimate the number of factors underlying the observed co-variance matrix, speculate on their identity and the manner in which the items load on the factors.

Structural equation modelling utilising LISREL (Jöreskog & Sörbom, 1996b) will be used to test the hypothesis that the measurement model implied by the PAPI-N scoring key can closely reproduce the observed co-variance matrix. Two overarching model fit hypotheses will be tested. More specifically the exact fit null hypothesis (H_{01}) will be tested which represents a rather ambitious stance that the measurement model accurately reflects the measurement model in the parameter (Browne & Cudeck, 1993):

H_{01}: RMSEA=0

H_{a1}: RMSEA>0

The exact fit null hypothesis represents the somewhat unrealistic position that the first-order measurement model is able to reproduce the observed co-variance matrix to a degree of accuracy that could be explained in terms of sampling error only. Browne and Cudeck (1993, p. 137) consequently argue:

> In applications of the analysis of co-variance structures in the social sciences it is implausible that any model that we use is anything more than an approximation to reality. Since a null hypothesis that a model fits exactly in some population is known a priori to be false, it seems pointless even to try to test whether it is true.

Assuming that the measurement model underlying the PAPI-N only approximates the processes that operated in reality to create the observed co-variance matrix, the following close fit null hypothesis (H_{02}) will also be tested (Browne & Cudeck, 1993):

H_{02}: RMSEA≤.05

H_{a2}: RMSEA>.05

If the exact or close fit would be found (i.e. H_{01} or H_{02} would not be rejected), or alternatively if the measurement model would at least demonstrate reasonable model fit (as indicated by the basket of fit indices produced by LISREL), the following 126 null hypotheses on the slope of the regression of item j on latent personality dimension k will be tested:

H_{0i}: $\lambda_{jk}=0$; i=3, 4, …, 128; j=1, 2, …, 126; k=1, 2, …, 21

H_{ai}: $\lambda_{jk}\neq0$; i=3, 4, …, 128; j=1, 2, …, 126; k=1, 2, …, 21

If the exact or close fit would be found (i.e. H_{01} or H_{02} would not be rejected), or alternatively if the measurement model would at least demonstrate reasonable model fit, the following null 126 hypotheses would be tested with regards to the freed elements in the variance-co-variance matrix $\boldsymbol{\Theta_\delta}$:

H_{0i}: $\Theta_{\delta jj}=0$; i =129, 130,..., 254; j=1, 2.....126

H_{ai}: $\Theta_{\delta jj} > 0$; i =129, 130,..., 254; j=1, 2.....126

If the exact or close fit would be found (i.e. H_{01} or H_{02} would not be rejected), or alternatively if the measurement model would at least demonstrate reasonable model fit, the following 210 null hypotheses would be tested with regards to the freed elements in the variance-co-variance matrix $\boldsymbol{\Phi}$:

H_{0i}: $\phi_{jk}=0$; i =255, 257,..., 464; j=1, 2.....21; k=1, 2.....21

H_{ai}: $\phi_{jk} > 0$; i =255, 257,..., 464; j=1, 2.....21; k=1, 2.....21

These 464 hypotheses will form the basis for examining the merits of the claim made by the developers that the PAPI-N successfully measures the 20 primary latent personality dimensions it intends to measure and in the manner that it intends to do according to the scoring key.

3.7 STATISTICAL ANALYSIS

Structural equation modelling (SEM) utilising LISREL 9.1 was used as the statistical analysis technique to test the proposed relationships amongst the latent personality dimensions and their item indicator variables as postulated by the PAPI-N.

Ullman (1996) describes structural equation modelling (SEM) in Davidson (2000) as “a collection of statistical techniques that allow for the examination of a set of relationships between one or more independent variables, either continuously or discretely, and one or more dependent variables, either continuously or discretely” (p. 709).

Bollen and Long (1993) argue in support of SEM:

> Structural equation models (SEMs) are a well-known component of the methodological arsenal of social sciences. Much of their attractiveness stems from their generality. Like econometric methods, SEMs allow consideration of simultaneous equations with many endogenous variables. Unlike most econometric methods, SEMs allow measurement error in the exogenous and endogenous variables. As with factor analysis developed in psychometrics and related procedures in sociometrics, SEMs permit multiple indicators of latent constructs and estimation of reliability and validity. In addition, SEMs allow more general measurement models than traditional factor-analytic structures and enable the researcher to specify structural relationships among the latent variables. Thus structural equation models are a synthesis of procedures developed in econometrics, sociometrics, and psychometrics (p.1).

This study will look at five distinct but interrelated steps, which characterise most applications of SEM (Bollen and Long, 1993; Diamantopoulos & Siguaw , 2000):

- Model specification
- Evaluation of model identification
- Estimation of model parameters
- Testing of model fit, and
- Model re-specification

Bollen and Long (1993) refer to model specification as the process of describing the nature and number of model parameters that needs to be estimated in the initial model. This phase also includes the construction of a comprehensive path diagram as a depiction of the overarching substantive hypotheses (i.e., in the case of this study, the measurement model). In the second phase, model identification needs to be evaluated and involves an examination of the data to determine whether unique values for the freed parameters of the specified model can be found. Once the model has been identified the researcher can proceed to the third phase, where an estimation technique needs to be selected. This is often determined by the nature of the variables that are being analysed. After the model parameters have been estimated, the researcher can test whether the model fits the data. Finally, re-specification and re-analysis of the model would be required in the case of poor model fit.

The design and architecture of the PAPI-N denotes a specific factor structure or measurement model. The strength of structural equation modelling (SEM) derives from the ability of this analytical technique to evaluate the fit of theoretically derived predictions on the nature of the relationships existing between indicator variables and latent variables and on the nature of the correlational relationships existing between latent variables (in the form of a measurement model) to the data. The following section aims to describe and motivate the procedures undertaken before conducting the SEM analyses. This section starts by specifying the model on which confirmatory factor analyses will be performed. Thereafter, the identification of the model needs to be evaluated. The necessity of performing item and dimensionality analyses will be explained as well as the procedures involved in these analyses. Finally, the method through which estimation of model parameters will occur is discussed and the manner in which model fit will be evaluated, described.

3.7.1 Model Specification

The architecture and the scoring key of the PAPI-N implies a hypothesis on the manner in which the individual test item scores are expected to be influenced by the dimensions of the personality construct as constitutively defined by the PAPI-N. The manner in which the responses of respondents to the PAPI-N items are hypothesised to be related to the twenty underlying first-order latent personality dimensions is depicted as a matrix equation (equation 1). Whether it is justified to make inferences about the twenty personality dimensions in the manner dictated by the scoring key depends on the fit of the measurement model and the strength of the loading of the items on the underlying latent variables. The substantive hypothesis that the PAPI-N provides a construct valid measure of personality within the workplace as defined by the instrument, on a South African sample, will be tested by testing the statistical hypotheses described in paragraph 3.6.

Equation 1 portrays the first-order measurement model implied by the scoring key of the PAPI-N when the items comprising each of the twenty-one subscales are used individually to represent each of the 20 latent personality dimensions and the SD latent variable.

$$\mathbf{X} = \mathbf{\Lambda}^{\mathbf{X}}\xi + \boldsymbol{\delta} \qquad (1)$$

Where:

- $\mathbf{X}$ is 126x1 column vector of observable indicator (item) scores;
- $\mathbf{\Lambda}^{\mathbf{X}}$ is a 126x21 matrix of factor loadings;

- ξ is a 21x1 column vector of first-order latent personality dimensions, and
- $\boldsymbol{\delta}$ is a 126x1 column vector of unique or measurement error components consisting of the combined effect on **X** of systematic non-relevant influences and random measurement error (Jöreskog & Sörbom, 1993).

Given that a hypothesis-testing, restricted, confirmatory factor analytic approach was utilised in the psychometric evaluation of the PAPI-N, specific structural assumptions were made relating to the number of latent variables that underlie personality, the relations among the latent variables, and the specific pattern of loadings of indicator variables (Theron & Spangenberg, 2005). According to Skrondal and Rabe-Hesketh (2004), the confirmatory factor analysis (CFA) technique is a hypothesis-testing procedure, specifically designed to test hypotheses about the relationships between items and factors whose number and interpretation are pre-determined. Particular model parameters are therefore set to prescribed values in the confirmatory model. These assumptions are primarily, however not exclusively, reflected in the order of the factor loading matrix $\boldsymbol{\Lambda}^{\mathbf{X}}$ (specifically the number of columns in lambda-X) and the pattern of freed and fixed factor loadings within the matrix. The measurement model above implies two additional matrices. The first is a symmetrical 21x21 $\boldsymbol{\Phi}$ (phi) matrix. This matrix contains the co-variance or correlations between the various latent personality dimension. The PAPI-N measurement model mathematically expressed in Equation 1 assumes that the primary personality factors are correlated. The second matrix $\boldsymbol{\theta_\delta}$ (theta-delta) is a 21x21 variance-co-variance matrix, which shows the variance in ($\theta_{\delta ii}$ & $\theta_{\delta jj}$) and co-variance ($\theta_{\delta ij}$) between the measurement error terms δ_i and δ_j. The measurement error terms δ_i and δ_j are normally assumed to be uncorrelated across the indicator variables and thus the co-variance terms are usually fixed to zero in $\boldsymbol{\Theta_\delta}$. Theta-delta is thus normally a diagonal matrix with all off diagonal elements set to zero (Spangenberg & Theron, 2004). By freeing off-diagonals in this matrix it would imply that the measurement error terms δ_i and δ_j may be correlated which means the possibility of additional common factors that are not reflected in the model. However, due to the confirmatory nature of this study, freeing off-diagonals in the matrix would be impossible to justify in terms of the design intentions of the developers of the PAPI-N.

In specifying the model, the scales of the measurement of the latent variables were not specified by setting the factor loadings on the first observed variable to unity. In the case of a single-group analysis Jöreskog and Sörbom (1993; 1998) contend that instead of defining the origin and unit of the latent variable scales in terms of observable reference variables, the latent variables should rather be standardised. In terms of this option the unit of measurement becomes the standard deviation $\sigma_i(\xi)$ (Spangenberg & Theron, 2004). All the factor loadings of each of the items on the latent personality dimensions of the PAPI-N were set free to be estimated. This was however only done with regards to the items designated to reflect each of the twenty personality factors. All the remaining elements of $\mathbf{\Lambda^X}$ were fixed at zero loadings to reflect the assumption that each item only reflects a single specific latent personality dimension and thereby the assumed factor simplicity of the PAPI-N items (Tabachnick & Fidell, 1989). All the elements of the $\mathbf{\Phi}$ matrix and the main diagonal of the θ_δ matrix were treated as free by default.

3.7.2 Evaluation of Model Identification

Diamantopoulos and Siguaw (2000) mention that when evaluating the identification of model, the researcher has to determine whether sufficient information is available in the observed inter-item variance-co-variance matrix order to obtain a unique solution for the parameters set free to be estimated in the measurement model. According to MacCallum (1995), a unique solution of the parameters in the model would be possible if for each free parameter there would have been at least one algebraic function that expresses that parameter as a function of sample variance or co-variance terms. Diamantopoulos and Siguaw (2000) and MacCallum (1995) further make two recommendations with regards to model identification. The first recommendation is that a definite scale should be established for each latent variable. The second recommendation is that model parameters to be estimated should not exceed the number of unique variance or co-variance terms in the sample observed co-variance matrix (Diamantopoulos and Siguaw, 2000; MacCallum, 1995). The following formula can be used to determine whether a specified model meets the minimum requirement for identification

$$t \leq s/2$$

where:

t = the number of parameters to be estimated
s = the number of variances and co-variances amongst the manifest (observable) variables, calculated as (p)(p +1)
p = the number of observed variables (i.e., items in this case).

If t > s/2 the model is unidentified. If a model is unidentified "it is the failure of the combined model and data constraints to identify (locate or determine) unique estimates that results in the identification problem" (Diamantopoulos and Siguaw, 2000 p. 48). If t = s/2 the model is just-identified. This means that a single unique solution can be obtained for the parameter estimates. A just-identified model, however, has zero degrees of freedom and therefore no variance-co-variance information remains to test the derived model solution (Diamantopoulos and Siguaw, 2000). If t < s/2 the model is over-identified. In this regard, it means that more than one estimate of each parameter can be obtained. In a model that is over-identified, the equations available outnumber the number of parameters to be estimated (Diamantopoulos and Siguaw, 2000). A just-identified model has positive degrees of freedom and therefore variance-co-variance information remains to test the derived model solution (Diamantopoulos and Siguaw., 2000). For the model, each latent variable will be treated as a (0; 1) standardised variable (MacCallum, 1995), thereby satisfying the first recommendation. The number of model parameters that are set free to be estimated (t=462) are less than the number of non-redundant elements in the observed sample co-variance matrix ([(p)(p+1)]/2=8001) (Diamantopoulos and Siguaw, 2000).

3.7.3 Item and Dimensionality Analysis

Prior to fitting the measurement model item analysis will be used to examine the assumption that the six items comprising each of the twenty subscales of the PAPI-N do in fact reflect a common underlying latent variable. The design of the PAPI-N reflects the intention to construct essentially one-dimensional sets of items to reflect variance in each of the twenty latent personality dimensions collectively comprising the personality domain. The items are meant to function as relatively homogenous stimulus sets to which test takers respond with behaviour, that is primarily a relatively uncontaminated expression of a specific underlying personality latent variable.

Item analysis is a technique used to identify and eliminate poor items that do not convey a clear representation of the subscale in question. Anastasi and Urbina (1997) indicate that item analysis can be used to create high validity and reliability in tests, thereby suggesting that tests can be improved through the selection, substitution and revision of items. The objective of item analysis is to create one-dimensional sets of items that explain variance in each of the latent personality variables comprising the personality construct. In the case of this study, however, poor items will not be deleted. The objective of the study is to evaluate the psychometric integrity of the South African PAPI-N. The researcher does not have the mandate to modify the instrument. When poor items are detected they will be tagged as such and reported. All poor items will remain in the measurement model that will be fitted.

The SPSS 21 reliability procedure will be used to analyse the items comprising each of the PAPI-N subscales.

High internal consistency reliability for each subscale (i.e., high Cronbach alpha's), high item-subscale total correlations, high squared multiple correlations when regressing items on linear composites of the remaining items comprising the subscale and other favourable item statistics will, however, not provide sufficient evidence that the common underlying latent variable is in fact a uni-dimensional latent variable. In the conceptualisation of the personality construct and in the design of the PAPI-N the fundamental assumption was that each of the twenty first-order personality factors is in fact a uni-dimensional latent variable. The manner in which the PAPI-N interprets personality does not make provision for any further refinement of the 20 first-order personality dimensions into narrower, more specific personality factors.

It is thereby, however not implied that each of the twenty personality dimensions is a narrow, very specific construct. Rather each personality dimension should be interpreted as a relatively broad facet of personality that expresses itself in a wide array of specific behaviours. Nonetheless each of the items comprising each of the twenty subscales of the PAPI-N is expected to load (albeit rather modestly) on a single factor. These items in the measurement model idealistically should function as homogenous stimuli to which test takers would respond in a manner that is a true expression of that specific single underlying latent variable. The intention of dimensionality analysis is to confirm the uni-dimensionality assumption on each subscale.

Although these items comprising each of the 20 subscales of the PAPI are intended to operate as stimulus sets to which the test takers respond with observable behaviour, which is primarily an expression of a specific uni-dimensional latent personality dimension, the behavioural responses to the items are not only dependant on the latent personality dimensions of interest, but also influenced by numerous other non-relevant latent variables and random error influences that are not relevant to the measurement objective (Guion, 1998). The systematic non-relevant latent variables that influence the response to item *i* are, however, assumed not to affect the response to item *j*. The focal latent personality dimension is therefore assumed to be the only common source of variance across all the items comprising a subscale. If the latent variable of interest therefore would be statistically controlled, the partial correlation between items will approach zero (Hulin, Drasgow, & Parsons, 1983). In addition if all the items in a subscale reflect a single underlying latent variable the partial bivariate inter-item correlations with all remaining items held constant should approach zero.

Dimensionality analysis normally allows the researcher to detect and remove those items with inadequate factor loadings, and/or to split heterogeneous subscales into two or more homogeneous subscales if necessary. Again it needs to be stressed that the researcher does not have the mandate to modify the instrument. Poor items will not be deleted. When poor items are detected they will be tagged as such and reported. All poor items will remain in the measurement model that will be fitted. When factor fission is found the nature of the fission will be examined, but the measurement model will not be modified. The objective of the research is to evaluate the psychometric credentials of the PAPI-N as it is currently being used in South Africa. Further to this, the researcher does not hold any intellectual property rights on the instrument and therefore cannot re-word or remove any items. Neither can the researcher modify the scoring key of the instrument.

The dimensionality analysis will be performed by subjecting each of the twenty personality subscales to an unrestricted principal axis factor analysis with oblique rotation. The exploratory factor analyses performed on the subscales will moreover shed additional light (via the magnitude of the factor loadings) on the success with which each item represents the common core underlying the subscale of items of which it forms part of. Principle axis factor analysis was chosen over principle component analysis because the intention was to determine the number of underlying factors that need to be assumed to account for the observed co-variance between the items comprising each subscale, whereas principle component analysis analyses common variance as well as error- and unique variance (Tabachnick & Fidell, 1996). Oblique rotation was chosen as rotational technique rather than orthogonal rotation because if more than one factor would emerge, oblique rotation would allow the extracted factors to correlate in the rotated solution (Tabachnick & Fidell, 1996).

Orthogonal rotation would have restricted the extracted factors to be uncorrelated in the rotated solution. Oblique rotation does not preclude the latter outcome. Although the expectancy is that the dimensionality analyses will verify the uni-dimensionality assumption and that rotation of the extracted solution will therefore not be required, oblique rotation provides for a more realistic spectrum of possible outcome in case of factor fission.

The uni-dimensionality assumption will be tested on all twenty subscales through exploratory factor analysis using SPSS 21. The eigenvalue-greater-than-unity rule of thumb along with the scree plot was used to determine the number of factors to be extracted. The uni-dimensionality assumption was considered to be supported if the eigenvalue-greater-than-unity rule results in the extraction of a single factor, if the magnitude of the factor loadings are reasonably high ($\lambda_{ij} \geq .50$ were considered large) and a small percentage of the residual correlations are greater than .05.

3.7.4 Estimation of Model Parameters

3.7.4.1 Variable Type

The PAPI-N utilises a seven-point Likert type scale, to which respondents are requested to indicate the degree to which they agree or disagree with the item statement. The individual PAPI-N items strictly speaking have to be treated as ordinal variables due to the nature of this type of scale used to capture the respondents' responses. Muthén and Kaplan (1985) stated that it is, however, acceptable practice to specify the data obtained on a five-point (or longer) scale as continuous data, especially in the case of CFA (maximum likelihood) SEM analyses. An alternative strategy to convert ordinal categorical data to continuous data would be to use item parcels rather than item level raw data.

For this study an important factor that needed to be considered was whether to fit the measurement model representing the twenty latent personality dimensions and the SD dimension with individual items or to create item parcels. Various considerations regarding psychometric properties, factor-solution and model fit needed to be considered before this decision could be made.

The creation of item parcels could serve as a solution for a variety of data problems, i.e. non-normal variables, small sample sizes and unstable parameter estimates. The use of item parcels presents itself as a practical measure to reduce the number of indicators in lengthy scales and therefore the number of parameters that needs to be estimated (Bandalos & Finney, 2001). To obtain valid results when analysing data with maximum likelihood estimation, the SEM process requires normally distributed continuous variables. Compared to individual items, it has been found that item parcels could better approximate normally distributed continuous variables when they serve as indicators of the latent construct. The use of item parcels would then have had the advantage of creating more reliable indicator variables thereby reducing the misrepresentation of model parameter estimates (Dunbar-Isaacson, 2006). It was also found that the use of item parcels could significantly improve model fit in some circumstances, for example model-fit indices like the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and the Chi-Square Test improves as the number of items in a parcel increases, but only with items that have a uni-dimensional structure (Bandalos, 2002; Bandalos & Finney, 2001). In some circumstances, item parcels may also help to assure that the assumption of multivariate normality is satisfied when handling data using maximum likelihood estimation methods (Sass & Smith, 2006). Dunbar-Isaacson (2006) also indicated lower skewness and kurtosis and higher validity for item parcels.

Disadvantages of using item parcels, however, do exist. Item parcels are only effective when uni-dimensionality is met. Another disadvantage is that item parcelling may increase the likelihood of misrepresenting the latent construct (Kim & Hagtvet, 2003). The reason for improved model fit could be due to the fact that the parcel-based models cancel out random and systematic error by adding these errors and therefore improved model fit. As a result it could reduce the probability of identifying miss specified models, thereby increasing the probability of Type II errors and consequently failure of rejecting a wrong model correctly (Little, Cunningham, Shahar & Widaman, 2002).

The advantages of item parcels would have been seriously considered if the fit of a structural model would have been evaluated. In the current study, however, the aim is not to evaluate the use of the PAPI-N to provide item parcel indicator variable measures for personality latent variables in a structural model. The aim is rather to evaluate the PAPI-N psychometrically as a freestanding measure of personality within the workplace. The aim is to evaluate whether the design intentions of the PAPI-N succeeded. The design intention was that each individual item would provide a valid and reliable measure of a specific latent personality dimension. The ideal approach therefore would have been to evaluate the success with which items represent the latent personality dimensions they were meant to reflect by fitting the measurement model with the individual items as indicator variables. Structural equation modelling on the PAPI-N in which individual items serve as indicator variables of the twenty latent personality dimensions will result in a cumbersome and extensive exercise that places high demands on memory capacity, simply due to the number of items involved. More specifically, fitting a measurement model in which individual items serve as indicator variables of the latent personality dimensions will require the estimation of 462 model parameters (126 factor loadings, 126 measurement error variances and 210 co-variance terms). This places formidable demands on the required sample size since the number of observations at least has to exceed the number of parameters to be estimated (Jöreskog & Sörbom, 1996a; 1996b). A reasonably large archival database of 5817 PAPI-N protocols has fortunately been developed at Work Dynamics. This allows the construct validity of the PAPI-N to be evaluated by fitting the measurement model with individual PAPI-N items as indicator variables.

3.7.4.2 Univariate and Multivariate Normality

As items measured on a 7-point scale were used as indicator variables in this study, they may be interpreted as approximating continuous variables (Muthén and Kaplan, 1985). When fitting a measurement model to continuous data, the method of maximum likelihood estimation is used to derive estimates for the measurement model parameters. This method of estimation, however, requires that the data follows a multivariate normal distribution (Kaplan, 2000). This is also true for generalised least squares (GLS) and full information maximum likelihood (FIML) as possible alternative estimation methods for structural equation modelling with continuous data (Holtzkamp, 2013).

In the event of working with non-normal data, additional estimation methods that could be used for structural equation modelling are robust maximum likelihood (RML), weighted least squares (WLS) and diagonally weighted least squares (DWLS) (Mels; 2003). Robust maximum likelihood is recommended in cases where the assumption of a multivariate normal distribution does not hold (Mels; 2003). The inappropriate analysis of continuous non-normal variables in structural equation models can however result in incorrect standard errors and chi-square estimates (Du Toit & Du Toit, 2001; Mels, 2003). The univariate and multivariate normality of the composite indicator variables will consequently be evaluated via PRELIS (Jöreskog & Sörbom, 1996b). If the null hypothesis of multivariate normality is rejected ($p<.05$) the item indicator variables will be normalised and the assumption of multivariate normality tested again. Most likely the multivariate assumption will still not hold although the departure from multivariate normality most likely will be less severe. Univariate normality will most probably have improved.

3.7.5 Testing of Model Fit

The 64 bit version of LISREL 9.1 was used to fit the measurement model defined in Equation 1. The measurement model syntax will be run from DOS. In the case of a large number of indicator variables the calculation of the inverse of the estimated asymptotic co-variance matrix requires extremely large memory capacity. The interactive use of the standard 32 bit version of LISREL 8.8 will not be able to cope with this (Holtzkamp, 2013).

According to Hooper, Coughlan and Mullen (2008), model fit refers to how well the proposed model reflecting the underlying theory is able to account for the observations made on the latent variables comprising the model. The objective of structural equation modelling is to determine how well the model "fits" the data of the underlying theory or more specifically, how well the model can account for the observed co-variance matrix. The model fits the data well when the estimated model parameters can closely reproduce the observed co-variance matrix. The model could then be considered as providing a plausible account of the process that generated the observed co-variance matrix. However, if the model fits well it could never be concluded that the process described in the model is necessarily the one that underpins the phenomenon of interest.

A full spectrum of fit indices provided by LISREL is available to guide the researcher to assess the absolute and comparative fit of the proposed measurement model. A number of cut-off values for these indices and the lack of agreement on which indices should be report on could however lead to conflicting information. Another concern that necessitates the researchers' attention is the fact that over the past few years newly and improved indices has been developed to obtain better model fit. It is therefore essential that researchers should use information with caution, as model fit is one of the utmost important steps in the structural equation modelling process (Diamantopoulos & Siguaw, 2000; Hooper et al., 2008).

For this reason, a conclusive verdict will not be pronounced on the fit of a model based on only a single statistical index. The full spectrum of fit indices available in LISREL will be considered to evaluate how well the model fits the underlying data and theory. This will include the following fit indices.

3.7.5.1 LISREL Fit Indices

ABSOLUTE FIT INDICES

Absolute fit indices are used to describe how well an *a priori* model fits the data and provides the best indication of how well the proposed theory fits the data. In contrast to incremental fit indices, absolute fit indices do not compare its calculations to a baseline model but rather measures how well the model fits on its own. The following indices are included in this category; Model Chi-Square (X^2), Root Mean Square Error of Approximation (RMSEA), Goodness-of-Fit (GFI), Adjusted Goodness-of-Fit (AGFI), the Root Mean Square residual (RMR) and the Standardised Root Mean Square residual (SRMR) (Jöreskog & Sörbom, 1993).

Model Chi-square (X^2)

The normal theory chi-square value is traditionally used to evaluate overall model fit when the multivariate normality assumption is met whilst the Satorra-Bentler chi-square values is used when the multivariate normality assumption does not hold (Mels, 2003). The Satorra-Bentler chi-square statistic results from using the robust maximum likelihood parameter estimation method and is morer suitable for multivariate non-normal data (Mels, 2003). Both chi-square statistics determine the magnitude of discrepancy between the observed and reproduced co-variance matrices. The chi-square statistic is used to test the exact fit null hypothesis (H_{01}: RMSEA=0) that the measurement model fits the data in the population perfectly and that the model can reproduce the observed co-variance matrix to a degree of accuracy that could be explained in terms of sampling error only. An insignificant p-value ($p>.05$) therefore indicates a good model fit. Both the normal theory weighted least squares chi-square statistic and the Satorra-Bentler statistic are sensitive to sample size. Large sample sizes nearly always result in model rejections whereas small sample sizes often results in the chi-square having a lack of power to discriminate between good fitting models and poor fitting models (Hooper et al., 2008).

Root Mean Square Error of Approximation (RMSEA)

The root mean square error of approximation (hereafter referred to as RMSEA) is a popular measure of fit that expresses the difference between the observed and estimated sample co-variance matrices. Moreover, it determines how well the model, with unfamiliar but optimally selected parameter estimates would fit the populations co-variance matrix. Diamantopoulos and Siguaw (2000) (cited in Hooper et al., 2008) mention that the RMSEA statistic has become one of the most informative fit indices mainly because of its sensitivity to the number of model parameters. In this regard, RMSEA prefers parsimony and will usually choose the model with the lesser number of model parameters. The RMSEA also focuses on the inconsistency amongst the observed and reproduced co-variance matrices in the population but expresses the inconsistency function value in terms of the degrees of freedom of the model (Diamantopoulos & Siguaw, 2000). A value below .05 is interpreted as an indication of a good model fit and values smaller than .08 indicates a reasonable fit. Values greater than .08 indicate poor model fit (Browne & Cudeck, 1993). Furthermore, LISREL provides for a test of closeness of model fit by formally computing the probability of the sample RMSEA value being observed in the sample under the null hypothesis H_{02}: RMSEA$\leq$.05 (Du Toit & Du Toit, 2001). Failure to reject the close fit null hypothesis would permit the conclusion that the measurement model fits closely in the parameter.

Goodness-of-Fit statistic (GFI) and the Adjusted Goodness-of-Fit statistic (AGFI)

The Goodness-of Fit (GFI) statistic was created as an alternative to the chi-square test (Jöreskog & Sörbom, 2003). This statistic serves to calculate the proportion of variance that is accounted for by the estimated population co-variance and determines how closely the model comes to replicating the observed co-variance matrix (Hooper et al., 2008). Recommendations for GFI cut-off values are .90 and when factor loadings and samples sizes are low, a cut-off value of .95 is required. Related to the GFI is the adjusted goodness-of-fit statistic (AGFI) which adjusts the GFI based upon degrees of freedom. Like the RMSEA, these indices also favour more parsimonious models but get penalised for model complexity. As with the GFI, indications of good model fit is confirmed by values in the range of 0 and 1 with a generally accepted value of .90. Furthermore, AGFI also tends to increase with sample size (Hooper et al., 2008).

Root Mean Square Residual (RMR) and Standardised Root Mean Square Residual (SRMR)

Hooper et al. (2008) describe these two indices as "the square root of the differences between the residuals of the sample co-variance matrix and the hypothesised co-variance model" (p. 54). The scale of each indicator is used to calculate the range of the RMR. It will however have complications in the interpretation of the RMR when a questionnaire contains items with varying scale lengths. For example some items may be measured on a scale ranging from 1-5 while others may range from 1-7. The standardised RMR (SRMR) counteracts this problem and is therefore much more helpful in interpretation. Values for the SRMR statistic range between 0 and 1.0, with values smaller than .05 indicating a good fit while values as high as .08 are still considered acceptable (Hooper et al., 2008). These authors however warn that when perfect model fit is obtained (i.e. a SRMR value of 0), consideration should be given to a high number of parameters and a large sample size as it often results in a lower SRMR value (Hooper et al., 2008).

INCREMENTAL FIT INDICES

Incremental fit indices, also known as comparative or relative fit indices, are a group of indices that do not use the chi-square in isolation to determine model fit, but rather compares the chi-square value of the fitted model to that of a baseline model. The null hypothesis for the baseline model is that all variables are structurally unrelated (Hooper et al., 2008). This category includes the normed-fit index (NFI) and the comparative fit index (CFI).

Normed-Fit Index (NFI)

This statistic evaluates the model by comparing the X^2 value of the model to the X^2 of the null model. The null or independence model represents the worst possible model as it specifies that all the variables are structurally unrelated. NFI values ranges between 0 and 1, where a value of 0 indicates the worst possible model and a value of 1 indicates the best possible model. Values greater than .90 are considered to reflect good model fit, however NNFI ≥ .95 is suggested as an acceptable cut-off value. NFI statistic is also very sensitive to sample size (Hooper et al., 2008).

Comparative Fit Index (CFI)

The Comparative Fit Index (CFI) is a revised form of the NFI but takes sample size into account. This index has however become one of the most popular fit indices in SEM to report on because of its insensitivity for sample size. As with the NFI, this statistic also assumes a model in which all the latent variables are structurally unrelated. Similar to NFI, CFI values also ranges between 0 and 1.0 with good fitting models obtaining values closer to 1.0. Cut-off values of CFI ≥ .95 are deemed acceptable (Hooper et al., 2008).

PARSIMONY FIT INDICES

Parsimony fit indices has been developed to overcome the problem of model complexity and includes the parsimony goodness-of-fit index (PCFI) and the parsimonious normed fit index (PNFI). The PGFI is obtained from the GFI by adjusting for loss of degrees of freedom. The PNFI also adjusts for degrees of freedom but is derived from the NFI statistic. Compared to other goodness of fit indices, parsimony fit index values are considerable lower because of the way they get penalised for model complexity. Values of .50 indicate a good model fit (Hooper et al., 2008).

A second form of parsimony fit indices, also known as information criteria indices, is available to determine model fit. The Akaike information criterion (AIC) and the consistent version of the AIC (CAIC) are the two indices developed that are used to compare non-nested or non-hierarchical models. Smaller values seem to indicate a good model fit, but because of an absence of a 0 to 1 scale it is difficult to recommend a cut-off value (Hooper et al., 2008). Smaller values are indicative of a better fitting model (Diamantopoulos & Siguaw, 2000).

3.7.6 Interpreting Measure Standardised Residuals and Modification Indices

Residuals represent the difference between elements of the observed and fitted/estimated co-variance matrices, in other words $\Sigma_{observed} - \Sigma_{fitted}$, where Σ represents a co-variance matrix (Diamantopoulos & Siguaw, 2000; Jöreskog & Sörbom, 1993). Residuals, especially standardised residuals, provide important diagnostic information on the sources of lack of fit in models (Kelloway, 1998). Jöreskog and Sörbom (1993) explain that a standardised residual refers to a residual that is divided by its approximate standard error and can be interpreted as a z-score (i.e. in terms of its standard deviation above or below the mean). On a 1% significance level standardised residuals can be considered large if they exceed +2.58 or -2.58 (Diamantopoulos & Siguaw, 2000). Ideally standardised residuals should be distributed approximately symmetrical around zero. Large residuals would be indicative of relationships (or the lack thereof) between indicator variables that the model fails to explain. More specifically, large positive residuals would indicate that the model underestimates the relationship between two variables and thus would imply the need for additional explanatory paths, whereas, large negative residuals would suggests that the model overestimates the relationship between two observed variables, implying the need to prune paths away. LISREL provides a summary of the largest and smallest standardised residuals as well as a stem-and-leaf plot that describes how the residuals are distributed around the median of zero. The Q-plot provides an additional graphical display of residuals, by plotting the standardised residuals (horisontal axis) against the quantiles of the normal distribution (Diamantopoulos & Siguaw, 2000).

Model modification indices indicates the extent to which the model's χ^2 fit statistic would decrease if a currently fixed parameter were set free and the model were re-estimated (Jöreskog & Sörbom, 1993). Model modification indices are aimed at answering the question whether any of the currently fixed parameters, when set free in the model, would significantly ($p<.01$) improve the parsimonious fit of the model. Large modification index values (>6.6349) provide an indication as to which parameters, if set free, would improve the fit of the model at a 1% significant level (Diamantopoulos & Siguaw, 2000). This, however, should only be done if the researcher can substantively justify such modifications as well as to ensure that it would make theoretical sense to do so (Diamantopoulos & Siguaw, 2000; Kelloway, 1998).

3.7.7 Interpreting measurement model parameter estimates

In addition to the standardised residuals and modification indices, parameter estimates were also examined to arrive at a verdict on the fit of the measurement model. As part of the evaluation process the focus would fall upon the hypothesised relationships between the latent variables and their indicators (De Goede, 2007). Moreover, the focus would fall on the validity and reliability of the observed variables used to represent the latent constructs of interest (De Goede, 2007).

According to Diamantopoulos and Siguaw (2000), evidence on the validity of the measures could be obtained through the examination of the magnitude and the significance of the paths between the latent variables and their indicators. The unstandardised factor loading matrix (**Λx**) indicates the regression coefficients of the regression of the individual items X_j on the latent personality dimensions ξ_i, and is used to evaluate the significance of the first-order factor loadings hypothesised by the proposed measurement model as expressed in equation 1. The loadings of the first-order factor loadings are significant ($p<.05$) if the t-values in the matrix $>|1.96|$.

Diamantopoulos and Siguaw (2000) however caution that there is a problem with solely relying on unstandardised loadings and their associated t-values. The problem is that unstandardised loadings retain scaling information of variables and thus can only be interpreted with reference to the scales of the variable. Hence, these scholars suggest that this problem could be avoided by examining the magnitude of the completely standardised solution, in which both latent and manifest variables have been standardised. The magnitude of the completely standardised loadings were therefore also examined.

The reliability of the indicators was further investigated by means of the squared multiple correlations (R^2) of the indicators. The squared multiple correlations (R^2) indicate the proportion of the indicator variance that is explained by its latent variable (Diamantopoulos &

Siguaw, 2000). A high R^2 value would be preferred as it would be indicative that the variance for the concerned indictor, to a large degree, reflects variance in the latent variable to which it has been linked. The rest of the variance, not explained by the latent variable can then be attributed to systematic and random measurement error (Diamantopoulos & Siguaw, 2000).

To explain the total variance in the i^{th} item (X_i) Spangenberg and Theron (2005) indicate that it could be decomposed into variance due to variance in the latent variable the item was designed to reflect (ξ_i), variance due to variance in other systematic latent effects the item was not designed to reflect as well as random measurement error. The latter two sources of variance in the indicator variable are acknowledged in equation 1 through the measurement error term (δ_i). The measurement error terms (δ_i) thus do not differentiate between systematic and random sources of error or non-relevant variance. The square of the completely standardised factor loadings λ_{ij} will therefore be interpreted as the proportion systematic-relevant item variance given that each item load on one latent variable only.

3.7.8 Discriminant Validity

According to Bagozzi, Yi, and Phillips (1991) discriminant validity can be examined by determining whether the correlations contained in the phi matrix are significantly different from one. This can be achieved by calculating a 95% confidence interval for each phi estimate. The following formula will be used: parameter estimate (phi value) ±1.96 * standard error. Discriminant validity is achieved when the correlations between the factors are significantly less than 1.0 ($p<.05$). Discriminant validity will therefore exist if the 95% confidence intervals calculated for the 190 phi estimates do not contain 1.

The average variance-extracted proportions will in addition be calculated for all 20 latent personality dimensions and compared to the square of the phi estimates between the latent personality dimensions (Diamantopoulos & Siguaw, 2000; Farrell, 2010; Hair, Black, Babin & Anderson., 2006). The average variance extracted is defined as the proportion of variance in the indicator variables that is due to the latent variable being measured rather than to measurement error. The average variance extracted is calculated as (Diamantopoulos & Siguaw, 2000, p. 91):

$$p_v = (\Sigma\lambda^2)/[\Sigma\lambda^2 + \Sigma(\theta_\delta)]$$

The variance-extracted estimates should be greater than the squared correlation estimate, indicating that a latent personality dimension explains its item measures better than it explains another construct (i.e. latent personality dimension), and thereby providing support for discriminant validity. The average variance extracted should exceed at least .50 so that the latent variable being measured by the indicators account for more of the variance in the indicators than measurement error.

If discriminant validity has not been shown, confidence in the claims that two latent variables are qualitatively distinct constructs would then be seriously compromised. Subscales that fail the test of discriminant validity should therefore be further evaluated to determine whether there are any items that may contribute to poor discriminant validity and possible poor fit. Items with high lambda-X modification indices may be possible culprits that contribute towards the discriminant validity problem and as complex items should be considered for deletion. By deleting indiscriminant items the fit of the model should improve (Hooper et al., 2008).

3.7.9 Model Re-specification

Given the complexity associated with structural equation modelling, it is not unusual to initially find poor model fit (Hooper et al., 2008). Some alterations to the model can however, significantly improve these results. This however should only be done if the researcher can substantively justify such modifications as well as ensure that it would make theoretical sense to do so (Diamantopoulos & Siguaw, 2000; Kelloway, 1998).

Hooper et al. (2008) recommend that a good starting point is to assess the fit of each construct and each item to determine whether there are any items that contribute to poor model fit. Items with a low multiple r^2 (<.02) generally presents very high levels of error and should therefore be removed. A second way in which model fit can be improved is by making provision for correlated error terms by freeing off-diagonals elements in the θ_δ (theta-delta) matrix. By allowing for correlated error terms would imply the existence of additional common factors that is not reflected in the model. Moreover, the correlation of error terms would mean that shared sources of systematic but non-relevant variance are affecting item scores and thereby causing co-variance between the measurement error terms amongst items.

Again, this should only be considered if the researcher can substantively justify such a decision (Hooper et al., 2008). In this study, however, the purpose was not to improve the fit of the measurement model but rather to evaluate the fit of the *a priori* model indicated by the test developers.

3.8 SAMPLE DESIGN

This section delineates the nature, details, size and limitations of the sample being used for the study. It also aims to provide information in support of the decision to undertake a confirmatory factor analysis into the first-order factor structure of the PAPI-N within the target population. The target population at which the PAPI-N was aimed at, and for which it has been standardised, consists of South African employees in managerial and non-managerial positions.

In this study the non-probability sampling method of convenience sampling was used. The use of non-probability sampling procedures means that the findings of this study can only be generalised to the target population with great caution. More specifically, the extent to which observations can or may be generalised to the target population depends on the representativeness of the sample.

The data used in this study was obtained from the data archives of Cubiks (Pty) Ltd, an International HR Consultancy Company, with written permission of the test distributor to utilise the sample data for the purpose of this research. The initial South African PAPI-N database comprised all respondents who were assessed by Work Dynamics for various assessment purposes as requested by their client organisations across different industries and occupations or jobs. The assessments were completed between 2007 and 2012 in different settings but under the same standardised conditions to aid in the assessment process. The data contained no missing values as the computerised version of the PAPI-N does not allow respondents the option to leave questions unanswered.

The database contained the scale scores and individual raw item scores for each of the items comprising the PAPI-N, and self-reported information on each respondent's gender, nationality, ethnicity, and mother tongue. However, records of the sample have been provided in an anonymous format which can at no point be traced back to an individual's score or profile as anonymity should always be protected. The majority of the respondent's information did not include information such as age, educational level, occupation, and management responsibility. This lack of information is rather unfortunate as it prevents the proper characterisation of the sample's age, educational level, occupation and managerial/non-managerial responsibility. A more accurate description of the research sample would have been desirable because these characteristics probably affect how respondents respond to the items comprising the PAPI-N. This is certainly an unfortunate shortcoming in this study, which will need to be taken cognisance of in future research.

The total sample consists of 5817 respondents of which 3438 (59%) were male and 2379 (41%) were female. From the total sample, the ethnic make-up consists of 2961 (51%) Blacks, 1580 (27.1%) Whites, 982 (16.8%) Coloureds, 264 (4.5%) Indians and 30 (0.5%) respondents who did not provide clear details of their ethnicity. All respondents have completed the PAPI-N in South African English. Because this study excludes other demographic data, it would be difficult to gain further insight into the impact that educational background, age, occupation or managerial/non-managerial positions have on the data. As this research aims to determine whether the PAPI-N provides a valid and reliable measure of the personality construct in the South African work context, the sample could be considered suitable for the purpose of this study.

The nature of the findings of the study, in conjunction with the limitations thereof, will as a result, limit any definitive conclusion on the applicability of the PAPI-N to the South African work context.

3.9 MEASURING INSTRUMENT

The study used the South African adapted PAPI-N English version, a self-report personality questionnaire which was specifically designed to elicit behaviours and preferences that are relevant to the workplace. It has been standardised for the South African population and is structured around "needs" (i.e. drivers or motivators) and "roles" (or behaviours) as experienced or displayed in the workplace.

The PAPI-N contains a hundred and twenty six statement items to which respondents are asked to rate themselves to the extent to which they agree or disagree with each statement on a 7-point scale (1 = absolutely disagree with the statement, to 7 = absolutely agree with the statement). Respondents reply on a Likert type scale, producing a total score for each of the ten need scales, ten role scales, and a social desirability scale, which make up their PAPI profile. Scoring is done by computer as data is captured and no subjective interpretation is involved. Scoring could, however, also be done manually with detailed instructions (available in the technical manual, p 116-120). Each of the PAPI-N scales include a reverse-keyed item which is opposite in meaning to the other items of a specific scale. The social responsibility scale has no reversed items (Anderson & Lewis, 1998).

In undertaking the assessment, the majority of respondents completed the PAPI-N on-line. In cases where respondents completed the paper and pencil version of PAPI-N, item responses were immediately captured into the database once they have completed the questionnaire. None of these contained missing values. The test was administered by qualified administrators (psychometrists and psychologists) who followed standardised procedures and testing conditions in all venues. Before the commencement of testing, every respondent completed consent and biographical information forms. The assurance was given that when results are used for research purposes, individual information will be kept confidential. The PAPI-N has no time limit, however respondents were informed of how long it normally takes to complete the questionnaire, for example “the average person takes about 25 minutes”. Respondents were also asked not to over think the deeper, underlying meaning of the items, but to give their initial responses.

3.10 SUMMARY

The purpose of this chapter was to describe the research methodology and the hypotheses that will be tested. It also discussed the statistical procedures that will be used to evaluate the data followed by the expected results. Chapter 4 will present and discuss the outcome results from the analyses.

CHAPTER 4

RESEARCH RESULTS

4.1 INTRODUCTION

The measurement model derived from the architecture of the PAPI-N and the design intentions of its developers hypothesises specific systematic relationships between specific indicator variables and specific latent personality variables. In accordance with the proposed relationships amongst the specific indicator variables and the underlying latent variables they were meant to reflect, specific statistical hypothesis were formulated. Two overarching statistical hypotheses were formulated on overall measurement model fit and 464 specific statistical hypotheses on the significance of the freed factor loadings in the factor loading matrix, the significance of the measurement error variances and the correlations between the twenty latent variables measured by the PAPI-N. The purpose of this chapter is to present and discuss the results of the various statistical analyses performed. Chapter 4 moreover aims to provide evidence on the validity of the operational hypotheses presented at the beginning of chapter 3.

The results in this chapter are presented in the following order: (i) item analyses, (ii) dimensionality analyses, (iii) test for multivariate normality on individual items, (iv) confirmatory factor analysis of the PAPI-N measurement model with individual items and (v) an evaluation of the discriminant validity of the 20 need and role latent variables. Prior to performing all the analysis reverse scoring was utilized for the following items due to the fact that they are stated in a negative sense. Item p1[12]; l5; c1; h1; d1; w6; r1; z5; n1; x1; b1; s1; o3; i3; t1; k6; e1; a1; f6; and g6 were all recoded in the positive direction.

[12] Items has been recorded according to the name of the subscale as to protect the scoring key of the PAPI-N. Items are recorded as such: p1, p2, p3, p4, p5, p6 for the P-subscale; l1, l2, l3, l4, l5, l6 for the L-subscale etc. Only the Social Desirability scales has been recorded differently: q1, q2, q3, q4, q5, q6

4.2 ITEM ANALYSIS

As described in the previous chapter, item analysis was conducted on each of the PAPI-N subscales by means of the reliability procedure of SPSS 21. This was done in order to identify items not contributing to an internal consistent description of the subscale in question. Again it is important to note that the decision on the psychometric credentials of any item should be based on a basket of evidence and not only on a single item statistic. The reliability procedure involved the calculation and evaluation of the following classical measurement theory item statistics for each of the PAPI-N subscales:

- Item means, variances and standard deviation,
- Inter-item correlations, and
- Item-total correlation statistics, which includes the change in subscale mean, variance, item-total- and squared multiple correlation and the coefficient of internal consistency if each item would be removed.

The classical measurement theory item statistics were calculated for each of the PAPI-N subscales. Analysis of these item statistics would typically result in the deletion of one or more items. It should be stressed, however, that item analysis in the case of this study only served the purpose of screening the suitability of the items that would represent the latent variables and not to propose permanent modifications to any of the scales in question. Decisions as to whether the items of the PAPI-N should be culled, modified or replaced should come from the developers of the instrument based on research feedback.

4.2.1 ITEM ANALYSIS FINDINGS: PAPI-N SUBSCALES

A summary of the item analysis results for each of the PAPI-N subscales as well as for the SD scale are presented in Table 4.1. The detailed output of the item analyses is electronically available on the enclosed CD, folder: Item Analysis, in Appendix A. The only subscales to achieve a Cronbach alpha value above .80 were H and T, while only eight of the remaining PAPI-N subscales (C, D, W, Z, N, X, B, I and SD) obtained values above the reliability standard of .70. However, it needs to be stressed that measures of personality generally tends to show somewhat lower coefficients of internal consistency (Smit, 1996). The results reported in Table 4.1 generally tend to correspond to the reliability results reported in Table

2.5. The Cronbach alpha values obtained for each of the personality subscales in the earlier study conducted on the smaller South African sample (see Table 2.5) correlated .791 with the reliability coefficient values obtained in the current study. The reliability coefficients obtained in the current study generally tended to be slightly higher than those reported in the earlier South African study. In only six cases was this trend reversed. Only in the case of subscale R did the Cronbach alpha show a substantial increase in value (+.309) from the value reported in the earlier study. It, however, should be remembered that the current sample includes the previous smaller sample.

Table 4.1

Reliability results of the PAPI-N Subscales

Subscale	Sample size	Mean	Variance	Standard deviation	Alpha
P	5817	28.76	30.051	5.482	.641
L	5817	33.33	21.042	4.587	.693
C	5817	33.77	28.560	5.344	.774
H	5817	34.43	21.111	4.595	.814
D	5817	35.48	17.728	4.210	.758
W	5817	34.90	21.248	4.610	.728
R	5817	31.71	19.159	4.377	.608
Z	5817	33.30	26.197	5.118	.770
N	5817	35.08	19.950	4.467	.748
X	5817	22.42	41.565	6.447	.759
B	5817	34.68	25.483	5.048	.750
S	5817	30.86	17.904	4.231	.594
O	5817	30.88	24.060	4.905	.648
I	5817	26.68	35.830	5.986	.771
T	5817	30.77	35.177	5.931	.831
K	5817	27.64	23.908	4.890	.424
E	5817	28.61	31.888	5.647	.657
A	5817	32.17	18.857	4.342	.508
F	5817	31.60	28.271	5.317	.675
G	5817	33.85	17.478	4.181	.638
SD	5817	18.65	49.434	7.031	.789

The descriptive statistics for the reliability coefficients obtained in the two studies are shown in Table 4.2. Generally the distributions showed agreement to a reasonable degree in terms of location and dispersion. The original distribution of reliability coefficients does, however, tend to be slightly more negatively skewed and slightly more leptokurtic than the distribution in the current study.

Table 4.2

Descriptive statistics for the PAPI-N Subscale reliability coefficients across two reliability studies

		ALPHA_5817	ALPHA_555
N	Valid	20	20
	Missing	0	0
Mean		.68945	.65635
Median		.71050	.71000
Mode		.424[a]	.299[a]
Std. Deviation		.102873	.128445
Variance		.011	.016
Skewness		-1.014	-1.649
Std. Error of Skewness		.512	.512
Kurtosis		1.005	2.397
Std. Error of Kurtosis		.992	.992
Range		.407	.468
Minimum		.424	.299
Maximum		.831	.767
Percentiles	25	.63875	.60000
	50	.71050	.71000
	75	.76725	.74825

a. Multiple modes exist. The smallest value is shown

Internal consistency reliability coefficients were calculated for subgroups based on ethnicity and gender. Cronbach alpha for each of these groups are reported in Table 4.3. Visual inspection of the results in Table 4.3 revealed alpha coefficients across the different ethnic groups, which ranged between .395 and .907. Remarkably lower reliability coefficients were found on the K- and A-subscale for Black respondents compared to the other ethnic groups as well as the general population. The reliability coefficients for the Black respondents were generally somewhat lower than for the White, Coloured and Indian respondents.

Internal consistency reliability coefficients were also calculated for male and female respondents. Table 4.3 reveals remarkable stability across both groups on most of the scales. The most variability were found on the G, D and B-subscales with alpha coefficients ranging between .523 and .566 for the G-subscale, .732 and .774 for the D-subscale, and .726 and .767 for the B-subscale.

Table 4.3

Internal consistency reliability of the PAPI-N Subscales by ethnic groups and gender

Subscale	Black (N = 2961)	Coloured (N = 982)	Indian (N = 264)	Other (N = 30)	White (N = 1580)	Female (N = 2370)	Male (N = 3438)
P	.582	.681	.668	.568	.700	.639	.626
L	.612	.746	.683	.797	.785	.701	.685
C	.748	.777	.790	.877	.804	.782	.766
H	.794	.829	.820	.857	.829	.805	.820
D	.742	.725	.756	.869	.791	.732	.774
W	.690	.671	.721	.789	.783	.707	.738
R	.552	.606	.701	.651	.669	.605	.609
Z	.672	.783	.747	.603	.832	.764	.773
N	.718	.773	.793	.732	.795	.727	.759
X	.718	.790	.779	.840	.810	.758	.759
B	.636	.765	.784	.799	.825	.726	.767
S	.595	.637	.622	.666	.564	.594	.593
O	.616	.659	.676	.622	.687	.665	.634
I	.728	.782	.776	.865	.826	.769	.772
T	.808	.831	.816	.907	.862	.835	.826
K	.345	.423	.549	.469	.517	.423	.427
E	.567	.722	.734	.576	.758	.653	.658
A	.395	.551	.463	.654	.615	.505	.510
F	.643	.689	.738	.763	.733	.663	.682
G	.539	.613	.521	.573	.557	.523	.566
SD	.777	.786	.818	.684	.797	.795	.785

4.2.1.1 Need to control others (P) scale

Table 4.4 presents the more detailed results of the item analysis for the P-subscale. The *Need to control others* scale obtained a questionable (see rules of thumb for describing reliability coefficients in Table 2.6) alpha coefficient of .641 (see Table 4.1 above). Visual inspection of the item means and item standard deviations revealed the absence of extreme means and small standard deviations and therefore suggested the absence of poor items. The item statistics showed the item means to range from 4.41 to 5.15 on a 7-point scale and the standard deviations from 1.347 to 1.700. The inter-item correlation matrix, however, revealed that all the items correlated below .50, with items p1 and p2 showing the lowest correlations. Furthermore, these two items obtained the lowest corrected item total correlation and squared multiple correlation values (< .30). The item-total statistics, however, indicated that the alpha coefficient would increase to .660 if item p2 would be deleted, and decrease to .637 if item p1 were deleted. Item p2's questionable status is further supported in Table 4.4 in that the scale variance decreases substantially less when p2 is deleted from the scale.

Item p2 was consequently flagged as a problematic item that lowered the homogeneity of the P-subscale. The low inter-item correlations, the relative magnitude of the corrected item-total correlation (.210), the squared multiple correlation (.051) and the increase in alpha affected by the removal of this item (from .641 to .660) would have justified the deletion of this item. Usually when an instrument's items is used to test a structural model one has created, one would delete such item (p2) to create psychometrically satisfactory measures to credibly test the merits of the model. The results shown in Table 4.4 explain the unsatisfactory alpha coefficient (.641) of this latent variable reported in Table 4.1, which subsequently also suggests that item p2 do not reflect the same underlying factor as the rest in the subscale.

The analysis was re-run to investigate whether the deletion of item p2 brought additional problematic items to the fore or not. The deletion of item p2, now resulted in item p1 coming to the fore as a problematic item. The relative magnitude of the corrected item-total correlation (.297), the squared multiple correlation (.105) and the increase in alpha affected by the removal of this item (.699 from .660) would have justified the deletion of this item from the P-subscale. It was however indicated that the P-subscale's alpha would, in the case of the deletion of any of the other items, not increase.

Table 4.4

Item analysis results for the Need to control others scale

	Mean	**Std. Deviation**	**N**
p1	4.96	1.700	5817
p2	4.41	1.623	5817
p3	4.54	1.558	5817
p4	4.92	1.435	5817
p5	4.76	1.474	5817
p6	5.15	1.347	5817

	p1	p2	p3	p4	p5	p6
p1	1.000					
p2	.062	1.000				
p3	.196	.176	1.000			
p4	.298	.156	.425	1.000		
p5	.194	.164	.325	.298	1.000	
p6	.152	.138	.341	.272	.354	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
p1	23.80	22.652	.278	.105	.637
p2	24.35	24.071	.210	.051	**.660**
p3	24.22	20.871	.474	.263	.557
p4	23.84	21.598	.480	.258	.559
p5	24.00	21.965	.428	.204	.577
p6	23.61	23.044	.402	.194	.589

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.794	4.414	5.154	.740	1.168	.077	6
Item Variances	2.333	1.814	2.892	1.078	1.594	.156	6
Inter-Item Correlations	.237	.062	.425	.363	6.862	.010	6

4.2.1.2 Leadership role (L) scale

The results for the item analysis for this subscale are depicted in Table 4.5. The *Leadership role* scale obtained a questionable alpha coefficient of .693 (see Table 4.1 above). When looking at Table 4.5 the item means fell in a range from 4.80 to 6.17 on a 7-point scale and the item standard deviations from .873 to 1.485. Although responses to item l3 reflected a slightly smaller standard deviation (.873), the mean of 6.17 could not really be considered extreme even though it approached the upper end of the 7-point scale. In the inter-item correlation matrix, a number of items failed to correlate at least moderately (r_{ij}>.30) with the other items in the subscale. Items l2 and l6 returning the lowest inter-item correlation values. While all the corrected item total correlations were larger than .30, three items (l2, l5 and l6) obtained lower squared multiple correlation values. In addition, item l6 was flagged as a problematic item due to the low inter-item correlations, the low squared multiple correlation (.114) and the increase in alpha affected by the removal of this item from .693 to .699.

The analysis was re-run to investigate whether the deletion of item l6 brought additional problematic items to the fore or not. After item l6 was deleted a Cronbach's alpha of .699 was obtained. Moreover, the item-total statistics revealed that the deletion of item l2, would further increase the Cronbach alpha from .699 to .718. After a second re-run of the analysis, the deletion of item l2 now resulted in item l5 coming to the fore as a problematic item. The squared multiple correlation (.192) and the increase in alpha affected by the removal of this item (.756 from .718) would have justified the culling of this item from L-subscale. It was further concluded that the alpha coefficient would, in the case of the deletion of any of the other indicated items, not increase. These findings raised the concern that the items flagged as problematic items are in fact items loading on a second factor that explains less variance in the item data. The results for the exploratory factor analysis reported in paragraph 4.4.3.2, however, indicated that the 6 subscale items all measure a single underlying latent variable but that the three problematic items have rather low factor loadings on the single factor.

Table 4.5:

Item analysis results for the Leadership role scale

	Mean	Std. Deviation	N
l1	5.85	.916	5817
l2	5.23	1.377	5817
l3	6.17	.873	5817
l4	5.64	1.002	5817
l5	5.63	1.477	5817
l6	4.80	1.485	5817

	l1	l2	l3	l4	l5	l6
l1	1.000					
l2	.298	1.000				
l3	.548	.257	1.000			
l4	.511	.258	.475	1.000		
l5	.367	.195	.377	.331	1.000	
l6	.247	.259	.224	.237	.153	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
l1	27.47	15.973	.578	.412	.620
l2	28.10	15.209	.366	.143	.676
l3	27.15	16.400	.549	.380	.631
l4	27.69	15.872	.522	.337	.629
l5	27.69	14.451	.393	.195	.670
l6	28.53	15.150	.319	.114	**.699**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.554	4.798	6.174	1.376	1.287	.233	6
Item Variances	1.481	.762	2.204	1.441	2.891	.469	6
Inter-Item Correlations	.316	.153	.548	.395	3.581	.013	6

4.2.1.3 Organised type (C) scale

Table 4.6 presents the results of the item analysis for the *Organised type* subscale. This scale obtained an acceptable alpha coefficient of .774 (see Table 4.1 above). The item means ranged from 5.45 to 5.98 (on a 7-point scale) and the item standard deviations from .952 to 1.758. No extreme means or small standard deviations therefore exist. In the inter-item correlation matrix all the items, except for items c2 and c4, correlated below .50. All the remaining items, except for item c1, correlated moderately (r_{ij}>.30) with each other. Item c2 obtained the lowest inter-item correlation ranging from .218 to 315. All the corrected item total correlations were larger than .30. Furthermore in the item-total statistics, items c1 and c3 obtained the lowest squared multiple correlation and corrected item total correlation values. Item c1 was, however, the only item that would increase the alpha coefficient quite substantially from .774 to .811 and therefore presented itself as problematic.

The analysis was run again to investigate whether the deletion of item c1 brought additional problematic items to the fore. The results indicated that none of the remaining items, if deleted, would further increase the alpha coefficient of .811 and item c1 should thus be the only item that should be flagged as problematic.

Table 4.6

Item analysis results for the Organised type scale

	Mean	Std. Deviation	N
c1	5.49	1.758	5817
c2	5.79	1.120	5817
c3	5.98	.952	5817
c4	5.59	1.272	5817
c5	5.46	1.240	5817
c6	5.45	1.318	5817

	c1	c2	c3	c4	c5	c6
c1	1.000					
c2	.270	1.000				
c3	.226	.426	1.000			
c4	.315	.576	.448	1.000		
c5	.218	.411	.329	.526	1.000	
c6	.257	.486	.409	.601	.417	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
c1	28.27	20.100	.340	.120	**.811**
c2	27.98	21.117	.601	.398	.724
c3	27.79	23.102	.498	.268	.750
c4	28.18	19.137	.701	.536	.693
c5	28.30	21.131	.517	.309	.741
c6	28.31	19.797	.599	.413	.719

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.628	5.455	5.982	.527	1.097	.046	6
Item Variances	1.691	.905	3.092	2.186	3.415	.560	6
Inter-Item Correlations	.394	.218	.601	.383	2.758	.015	6

4.2.1.4 Integrative planner (H) scale

The results for the item analysis for this subscale are depicted in Table 4.7. The *Integrative planner* scale returned one of the highest alpha coefficients (.814) of all the scales (see Table 4.1). When looking at the item statistics the item means fell in a range from 5.33 to 5.98 (on a 7-point scale) and the item standard deviations from .875 to 1.522. Except for item h1, all the other items obtained higher inter-item correlations than .30. In addition, the low squared multiple correlation (.152) and a substantial increase in alpha affected by the removal of item h1 (from .814 to .860) suggested that the item was not successfully reflecting the same underlying latent variable than the majority of the items in the H-subscale and should therefore be deleted.

The analysis was re-run to investigate whether the deletion of item h1 brought additional problematic items to the fore. The results indicated that none of the remaining items, if deleted, would further increase the alpha coefficient of .860 and it concluded, therefore, that h1 should be the only item to be flagged as problematic.

Table 4.7

Item analysis results for the Integrative planner scale

	Mean	Std. Deviation	N
h1	5.33	1.522	5817
h2	5.87	.955	5817
h3	5.77	.896	5817
h4	5.98	.875	5817
h5	5.79	.939	5817
h6	5.69	1.052	5817

	h1	h2	h3	h4	h5	h6
h1	1.000					
h2	.307	1.000				
h3	**.265**	.466	1.000			
h4	.308	.614	.545	1.000		
h5	.309	.505	.466	.521	1.000	
h6	.357	.588	.508	.616	.691	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
h1	29.10	14.336	.387	**.152**	**.860**
h2	28.56	15.397	.641	.464	.773
h3	28.66	16.189	.571	.367	.788
h4	28.45	15.651	.678	.521	.769
h5	28.64	15.436	.650	.508	.772
h6	28.75	14.222	.729	.603	.751

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.738	5.332	5.977	.645	1.121	.049	6
Item Variances	1.131	.766	2.316	1.550	3.022	.351	6
Inter-Item Correlations	.471	.265	.691	.426	2.607	.017	6

4.2.1.5 Attention to detail (D) scale

Table 4.8 presents the results of the item analysis for the D-subscale. The *Attention to detail* scale obtained an acceptable alpha coefficient of .758 (Table 4.1). Further inspection showed that the item means ranged from 5.76 to 6.02 on a 7-point scale and the item standard deviations from .836 to 1.486. While the responses to items d4 and d6 reflected smaller standard deviations (.836 and .863), none of the means were on either one of the extreme ends on the 7-point scale. Table 4.8 further reveals a relatively more coherent set of items which tends to respond more in unison to systematic differences in the underlying latent variable. This can be concluded from the moderate correlations in the inter-item correlation matrix and the higher item-total correlations in Table 4.8.

One of the items however was flagged as problematic. Item d1 was flagged as it correlates low with the majority of the other items in the subscale (r_{ij}<.30). Furthermore, the corrected item-total correlation (.283) and the squared multiple correlation (.083) also point to the problematic status of item d1. The results further indicated that this item, if deleted, would increase the alpha coefficient substantially from .758 to .820.

The analysis was re-run to investigate whether any additional problematic items came to the fore. The results indicated that none of the remaining items, if deleted, would further increase the Cronbach's alpha of .820. Item d1 therefore presents itself as the only item to be problematic.

Table 4.8

Item analysis results for the Attention to detail scale

	Mean	Std. Deviation	N
d1	5.93	1.486	5817
d2	5.89	.979	5817
d3	5.76	1.029	5817
d4	5.93	.836	5817
d5	5.96	.925	5817
d6	6.02	.863	5817

	d1	d2	d3	d4	d5	d6
d1	1.000					
d2	.236	1.000				
d3	.238	.498	1.000			
d4	.182	.445	.444	1.000		
d5	.201	.502	.458	.481	1.000	
d6	.217	.499	.465	.467	.545	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
d1	29.55	12.547	.283	.083	**.820**
d2	29.59	12.598	.600	.397	.698
d3	29.72	12.477	.577	.362	.703
d4	29.55	13.671	.543	.341	.717
d5	29.52	12.926	.594	.417	.702
d6	29.46	13.214	.601	.413	.703

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.913	5.757	6.017	.260	1.045	.008	6
Item Variances	1.087	.698	2.207	1.509	3.161	.319	6
Inter-Item Correlations	.392	.182	.545	.364	3.002	.017	6

4.2.1.6 Need for rules and supervision (W) scale

Table 4.9 presents the detailed results of the item analysis for the W-subscale. The *Need for rules and supervision* subscale items returned an acceptable alpha coefficient of .728 (see

Table 4.1). Visual inspection of the item means and item standard deviations revealed the absence of extreme means and small standard deviations and therefore suggests the absence of poor items. The mean ranged from 5.60 to 6.34 on a 7-point scale and the standard deviation from .994 to 1.454. Furthermore, the inter-item correlation matrix revealed that all the items correlated below .50. All the corrected item total correlations were larger than .30 with the lowest being .336 for item w6. Squared multiple correlations below .30 were nevertheless obtained for items w1, w2 and w6.

Item w6 was, however, flagged as the only item that showed itself as problematic in as far as the low inter-item correlations along with the relative magnitude of the squared multiple correlation (.154) and a substantial increase in alpha affected by the removal of this item (from .728 to .740).

The analysis was run again to investigate whether the deletion of item w6 brought additional problematic items to the fore. The item-total statistics revealed that none of the remaining items, if deleted, would further increase the alpha coefficient of .740 and item w6 should thus be the only item that should be flagged as problematic.

Table 4.9

Item analysis results for the Need for rules and supervision scale

	Mean	**Std. Deviation**	**N**
w1	6.34	.994	5817
w2	5.85	1.059	5817
w3	5.81	1.183	5817
w4	5.60	1.173	5817
w5	5.64	1.168	5817
w6	5.67	1.454	5817

	w1	w2	w3	w4	w5	w6
w1	1.000					
w2	.210	1.000				
w3	.338	.271	1.000			
w4	.263	.498	.437	1.000		
w5	.322	.297	.498	.453	1.000	
w6	.148	.348	.156	.319	.204	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
w1	28.56	17.237	.366	.157	.715
w2	29.05	15.964	.492	.296	.683
w3	29.09	15.254	.497	.329	.680
w4	29.30	14.484	.604	.400	.647
w5	29.26	15.096	.528	.337	.671
w6	29.23	15.316	.336	.154	**.740**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.817	5.597	6.339	.743	1.133	.076	6
Item Variances	1.393	.987	2.114	1.126	2.141	.152	6
Inter-Item Correlations	.317	.148	.498	.350	3.354	.013	6

4.2.1.7 Conceptual thinker (R) scale

The results for the item analysis for the *Conceptual thinker* subscale are depicted in Table 4.10. The reliability statistics indicated a questionable alpha coefficient of .608 (Table 4.1). The item statistics showed the item means range from 4.71 to 5.76 (on a 7-point scale) for the 6 items included in this subscale. Standard deviations ranged from .945 and 1.500. Table 4.10 however displays a somewhat incoherent set of items which, although they were all meant to reflect factor R, nonetheless do not seem to respond in unison to systematic differences in a single underlying latent variable. This is apparent from the low inter-item correlations (.001 to .486) and the low squared multiple correlations (.038 to .287).

Item r1 was flagged as an item that lowered the homogeneity of the R-subscale. The deletion of the item could be justified by the low inter-item correlations (.001 to .164), the magnitude of the corrected item-total correlation (.153), the squared multiple correlation (.038) and the increase in alpha affected by the removal of this item (from .608 to .655). The results shown in Table 4.10 explain the unsatisfactory alpha coefficient (.608) of this latent variable reported in Table 4.1, which subsequently also suggests that item r1 do not reflect the same underlying factor as the rest of the items in the subscale.

The analysis was re-run to investigate whether the deletion of item r1 brought additional problematic items to the fore or not. The deletion of item r1, now resulted in item r6 coming to the fore as a problematic item. The relative magnitude of the corrected item-total correlation (.292), the squared multiple correlation (.093) and the increase in alpha affected by the removal of this item (.664 from .655) would have justified the deletion of this item as well from the scale. The results, however, further suggested that the alpha coefficient would, in the case of the deletion of any of the other indicated items, not increase.

Table 4.10

Item analysis results for the Conceptual thinker scale

	Mean	Std. Deviation	N
r1	5.46	1.500	5817
r2	4.98	1.375	5817
r3	5.08	1.266	5817
r4	5.72	.991	5817
r5	5.76	.945	5817
r6	4.71	1.353	5817

	r1	r2	r3	r4	r5	r6
r1	1.000					
r2	.133	1.000				
r3	.164	.445	1.000			
r4	.112	.285	.278	1.000		
r5	.099	.265	.295	.486	1.000	
r6	.001	.198	.182	.212	.258	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
r1	26.25	15.119	.153	.038	**.655**
r2	26.74	12.966	.435	.241	.521
r3	26.64	13.355	.454	.250	.516
r4	25.99	14.889	.430	.275	.539
r5	25.95	15.019	.443	.287	.538
r6	27.00	14.737	.249	.096	.604

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.285	4.712	5.764	1.052	1.223	.184	6
Item Variances	1.575	.893	2.251	1.358	2.520	.288	6
Inter-Item Correlations	.228	.001	.486	.485	728.563	.015	6

4.2.1.8 Need for change (Z) scale

Table 4.11 presents the results of the item analysis for the Z-subscale. The *Need for change* scale obtained an acceptable alpha coefficient of .770 (see Table 4.1). Visual inspection of the item means and item standard deviations revealed the absence of extreme means and small standard deviations and therefore suggested the absence of poor items. The mean ranged from 4.55 to 6.08 on a 7-point scale and the standard deviation ranged from .991 to 1.649. In the inter-item correlation matrix some items (z5 and z6) failed to correlate moderately (r_{ij}>.30) with the remainder of the items in the subscale. All the corrected item total and squared multiple correlations were larger than .30. Item z5 was, however, the only item to present itself as problematic. The low inter-item correlations, the relative magnitude of the squared multiple correlation (.120) and the increase in alpha affected by the removal of this item (from .770 to .779) would have justified the elimination of this item.

The analysis was subsequently re-run to investigate whether any additional problematic items came to the fore after the deletion of item z5. The item-total statistics revealed that none of the remaining items if deleted would further increase the alpha coefficient of .779. Item z5 was therefore the only problematic item that should be considered to be deleted from the Z-subscale.

Table 4.11

Item analysis results for the Need for change scale

	Mean	**Std. Deviation**	**N**
z1	5.85	1.246	5817
z2	6.03	.991	5817
z3	6.08	.949	5817
z4	4.96	1.488	5817
z5	4.55	1.649	5817
z6	5.84	1.014	5817

	z1	z2	z3	z4	z5	z6
z1	1.000					
z2	.415	1.000				
z3	.417	.488	1.000			
z4	.458	.338	.358	1.000		
z5	.333	.244	.281	.349	1.000	
z6	.450	.466	.531	.440	.292	1.000

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach Alpha if Item Deleted**
z1	27.45	18.427	.581	.347	.718
z2	27.27	20.535	.521	.330	.738
z3	27.23	20.480	.561	.381	.731
z4	28.34	17.232	.546	.313	.729
z5	28.76	17.744	.413	.177	**.779**
z6	27.47	19.769	.599	.407	.721

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.550	4.546	6.076	1.530	1.337	.407	6
Item Variances	1.566	.900	2.718	1.818	3.020	.565	6
Inter-Item Correlations	.391	.244	.531	.287	2.174	.007	6

4.2.1.9 Need to finish a task (N) scale

The results for the item analysis for the *Need to finish a task* subscale are depicted in Table 4.12. This subscale returned an acceptable alpha coefficient of .748 (see Table 4.1). Further investigation showed that the item means ranged from 4.84 to 6.21 on a 7-point scale and the item standard deviations ranged from .833 to 1.773. In the inter-item correlation matrix, item n1 stood out dramatically as all its obtained correlations were below .30. In addition, item n1 obtained the lowest corrected item total correlation (.301) and squared multiple correlation (.100) values. The results further indicated that this item, if deleted, would increase the alpha substantially from .748 to .828.

The analysis was run again to investigate whether the deletion of item n1 brought additional problematic items to the fore. The item-total statistics revealed that none of the remaining items if deleted would further increase the alpha coefficient of .828 and item n1 should thus be the only item that should be considered to be deleted from the N-subscale.

Table 4.12

Item analysis results for the Need to finish a task scale

	Mean	Std. Deviation	N
n1	4.84	1.773	5817
n2	6.11	.981	5817
n3	6.08	.878	5817
n4	5.97	1.011	5817
n5	5.87	.963	5817
n6	6.21	.833	5817

	n1	n2	n3	n4	n5	n6
n1	1.000					
n2	.223	1.000				
n3	.196	.481	1.000			
n4	.289	.497	.459	1.000		
n5	.199	.467	.473	.517	1.000	
n6	.253	.495	.491	.555	.499	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
n1	30.24	12.959	.301	.100	**.828**
n2	28.97	14.708	.569	.379	.693
n3	29.00	15.404	.547	.363	.702
n4	29.11	14.149	.628	.439	.676
n5	29.21	14.852	.562	.386	.695
n6	28.87	15.248	.616	.433	.690

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.847	4.841	6.213	1.372	1.283	.257	6
Item Variances	1.253	.694	3.143	2.449	4.529	.872	6
Inter-Item Correlations	.406	.196	.555	.359	2.834	.017	6

4.2.1.10 Need to be noticed (X) scale

Table 4.13 presents the results of the item analysis for the X-subscale. The *Need to be noticed* subscale items obtained an acceptable alpha coefficient of .759 (see Table 4.1). The results showed the item means ranged from 2.77 to 5.76 on a 7-point scale and the standard deviation ranged from 1.109 to 1.849. In the inter-item correlation matrix, items x1 and x6, failed to correlate moderately (r_{ij}>.30) with the remainder of the items in the subscale. All the corrected item total and squared multiple correlations were larger than .30, except for x1 and x6. The item-total statistics further indicated that the alpha coefficient would increase from .759 to .793 if item x6 were to be deleted. Although item x1 does have somewhat lower inter-item and squared multiple correlations the alpha coefficient would nonetheless not be positively affected by the removal of this item. Item x6 was, therefore, flagged as the only item that did not reflect the same underlying latent variable as the rest of the items.

The analysis was re-run to investigate whether the deletion of item x6 brought additional problematic items to the fore or not. The deletion of item x6, resulted in item x1 coming to the fore as a problematic item. The squared multiple correlation (.195) and the increase in alpha affected by the removal of this item (.812 from .793) would have justified the deletion of this item from the X-subscale. It was further indicated that the X-subscale's alpha would, in the case of the deletion of any other items, not increase. After deleting x1, none of the other items revealed itself as problematic.

Table 4.13

Item analysis results for the Need to be noticed scale

	Mean	Std. Deviation	N
x1	3.72	1.849	5817
x2	3.06	1.667	5817
x3	2.77	1.507	5817
x4	3.79	1.640	5817
x5	3.31	1.699	5817
x6	5.76	1.109	5817

	x1	x2	x3	x4	x5	x6
x1	1.000					
x2	.422	1.000				
x3	.307	.548	1.000			
x4	.274	.423	.485	1.000		
x5	.300	.489	.540	.636	1.000	
x6	.155	.102	.040	.112	.088	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
x1	18.69	29.648	.422	.206	.750
x2	19.36	27.973	.613	.411	.691
x3	19.65	29.530	.596	.418	.699
x4	18.63	28.557	.589	.441	.699
x5	19.11	27.375	.636	.496	.684
x6	16.66	38.452	.137	.033	**.793**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.736	2.772	5.762	2.990	2.079	1.134	6
Item Variances	2.546	1.231	3.420	2.190	2.780	.552	6
Inter-Item Correlations	.328	.040	.636	.596	16.089	.037	6

4.2.1.11 Need to belong to groups (B) scale

The results of the item analysis for the *Need to belong to groups* subscale are depicted in Table 4.14. The reliability statistics indicated an acceptable alpha coefficient of .750 (see Table 4.1). Visual inspection of the item means and item standard deviations revealed the absence of extreme means and small standard deviations and therefore suggested the absence of poor items. The item statistics showed the means to range from 5.33 to 6.25 (on a 7-point scale) and the standard deviations from 1.017 to 1.735. Furthermore, the inter-item correlation matrix revealed that all the items, but for item b1, correlated moderately ((r_{ij}>.30) with each other. Besides item b1, all the corrected item total correlations were larger than .30.

Furthermore, items b1 and b4 obtained the lowest squared multiple correlation values (.087 and .295). The item-total statistics, however, indicated that the alpha coefficient would increase to .803 if only item b1 would be deleted.

Item b1 was as a result flagged as a problematic item due to the low inter-item correlations (.158 to .250), the relative magnitude of the corrected item-total correlation (.282), the squared multiple correlation (.087) and the substantial increase in alpha affected by the removal of this item (from .750 to .803).

The analysis was run again to investigate whether the deletion of item b1 brought additional problematic items to the fore. The results indicated that none of the remaining items if deleted would further increase the alpha coefficient of .803 and item b1 should thus be the only item that should be considered to be deleted from the B-subscale.

Table 4.14

Item analysis results for the Need to belong to groups scale

	Mean	**Std. Deviation**	**N**
b1	5.33	1.735	5817
b2	6.25	1.037	5817
b3	5.35	1.473	5817
b4	5.86	1.048	5817
b5	5.93	1.077	5817
b6	5.95	1.017	5817

	b1	b2	b3	b4	b5	b6
b1	1.000					
b2	.242	1.000				
b3	.188	.398	1.000			
b4	.158	.348	.328	1.000		
b5	.238	.473	.482	.499	1.000	
b6	.250	.497	.486	.487	.692	1.000

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach Alpha if Item Deleted**
b1	29.35	18.290	.282	.087	**.803**
b2	28.43	19.463	.541	.313	.706
b3	29.32	17.139	.506	.300	.712
b4	28.81	19.876	.482	.295	.719
b5	28.75	18.204	.666	.545	.674
b6	28.72	18.509	.679	.552	.675

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.779	5.327	6.250	.924	1.173	.134	6
Item Variances	1.591	1.034	3.012	1.978	2.913	.671	6
Inter-Item Correlations	.384	.158	.692	.534	4.379	.022	6

4.2.1.12 Social harmoniser (S) scale

Table 4.15 presents the results of the item analysis for the *Social harmonizer*. This scale obtained a poor alpha coefficient of .594. The results showed the item means to range from 2.83 to 5.86 on a 7-point scale and the item standard deviation to range from .969 to 1.518. Table 4.15 displays a worrisome lack of coherence in the set of items designed to measure factor S. The low (and quite often negative) inter-item correlations and low (and at times negative) item-total correlations as well as the low squared multiple correlations indicated that the items comprising this subscale do not respond in unison to systematic differences in a single underlying latent variable.

Furthermore, item s1 presented itself as a problematic item that lowered the homogeneity of the S-subscale. The negatively low inter-item correlations (-.075 to .006), the relative magnitude of the corrected item-total correlation (-.057), the squared multiple correlation (.011) and the substantial increase in alpha affected by the removal of this item (from .594 to .733) would have warranted the deletion of this item.

The analysis was re-run to investigate whether any additional problematic items came to the fore. The item-total statistics revealed that none of the items if deleted would further increase the alpha coefficient of .733. Item s1 was therefore the only item that should be considered to be deleted from the S-subscale.

Table 4.15

Item analysis results for the Social harmoniser scale

	Mean	Std. Deviation	N
s1	2.83	1.518	5817
s2	5.82	.969	5817
s3	5.16	1.359	5817
s4	5.86	1.043	5817
s5	5.46	1.275	5817
s6	5.73	1.113	5817

	s1	s2	s3	s4	s5	s6
s1	1.000					
s2	-.075	1.000				
s3	-.039	.204	1.000			
s4	-.060	.262	.382	1.000		
s5	.006	.303	.383	.481	1.000	
s6	-.045	.494	.277	.342	.477	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
s1	28.03	16.296	-.057	.011	**.733**
s2	25.04	14.305	.363	.259	.542
s3	25.70	12.431	.378	.205	.526
s4	25.00	13.322	.459	.294	.503
s5	25.40	11.541	.546	.372	.447
s6	25.13	12.763	.491	.371	.486

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.143	2.830	5.855	3.025	2.069	1.352	6
Item Variances	1.507	.938	2.304	1.366	2.455	.267	6
Inter-Item Correlations	.226	-.075	.494	.569	-6.589	.044	6

4.2.1.13 Need to relate closely to individuals (O) scale

The results for the item analysis for the O-subscale are depicted in Table 4.16. The *Need to relate closely to individuals* scale returned a questionable alpha coefficient of .648 (see Table 4.1). The item means ranged from 3.61 to 5.81 (on a 7-point scale) and the item standard deviations from 1.092 to 1.665. Table 4.16 displays a somewhat incoherent set of items which, although they were all meant to reflect factor O, nonetheless did not seem to respond in unison to systematic differences in a single underlying latent variable. This was apparent from the low inter-item correlations and some of the low squared multiple correlations in Table 4.16. Items o3, o5 and o6 obtained the lowest inter-item correlation and the lowest squared multiple correlation values.

Moreover, item o3 presented itself as a problematic item that lowered the homogeneity of the O-subscale. The low inter-item correlations (.051 to .114), the relative magnitude of the corrected item-total correlation (.114), the squared multiple correlation (.016) and the substantial increase in alpha affected by the removal of this item (from .648 to .721) would, under the normal circumstances, justified the deletion of this item.

The analysis was run again to investigate whether the deletion of item o3 brought additional problematic items to the fore. The deletion of item o3, now resulted in item o6 coming to the fore as a problematic item. The relative magnitude of the corrected item-total correlation (.309), the squared multiple correlation (.098) and the increase in alpha affected by the removal of this item (.759 from .721) would have justified the deletion of this item from the scale. When item o6 was deleted and a re-run of the analysis was completed, it could be concluded that none of the further items, if deleted, would increase the alpha coefficient of .759.

Table 4.16

Item analysis results for the Need to relate closely to individuals scale

	Mean	Std. Deviation	N
o1	5.81	1.322	5817
o2	5.75	1.237	5817
o3	3.61	1.665	5817
o4	5.51	1.179	5817
o5	5.71	1.092	5817
o6	4.50	1.561	5817

	o1	o2	o3	o4	o5	o6
o1	1.000					
o2	.606	1.000				
o3	.081	.114	1.000			
o4	.433	.478	.098	1.000		
o5	.319	.390	.051	.407	1.000	
o6	.222	.238	.053	.258	.229	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
o1	25.07	16.786	.510	.398	.556
o2	25.13	16.710	.575	.448	.536
o3	27.27	19.611	.114	.016	**.721**
o4	25.37	17.543	.519	.321	.560
o5	25.17	18.877	.421	.227	.596
o6	26.38	17.782	.292	.099	.643

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.146	3.606	5.809	2.202	1.611	.809	6
Item Variances	1.845	1.192	2.774	1.582	2.327	.391	6
Inter-Item Correlations	.265	.051	.606	.555	11.839	.028	6

4.2.1.14 Ease in decision making (I) scale

Table 4.17 presents the results of the item analysis for the I-subscale. The *Ease in decision making* scale obtained an acceptable alpha coefficient of .771. Visual inspection of the item means and item standard deviations revealed the absence of extreme means and small standard deviations and therefore suggested the absence of poor items. The item statistics showed the means to range from 3.32 to 5.11 on a 7-point scale and the standard deviations from 1.312 to 1.681.

The results further showed a relatively more coherent set of items which tended to respond in more unison to systematic differences in a single underlying latent variable. This was evident from the moderate (although at times still low) inter-item correlations and the somewhat higher item-total correlations and squared multiple correlations. However, two items were flagged as problematic. Item i2 and item i3 were identified as items that lowered the homogeneity of the I-subscale. The low inter-item correlations, the relative magnitude of the corrected item-total correlations (.367 and .233), the squared multiple correlations (.152 and .058) and the increase in alpha affected by the removal of item i3 (from .771 to .811) would have justified the removal of these items from the subscale.

The analysis was re-run to investigate whether the deletion of item i2 and item i3 brought additional problematic items to the fore[13]. After the two items were deleted the Cronbach's alpha increased from .771 to .853. Inspection of the item-total statistics for the remaining items revealed that none of the items if deleted would further increase the alpha coefficient of .853 and thus i2 and i3 should be the only two items that should be considered for deletion from the I-subscale.

Table 4.17

Item analysis results for the Ease in decision making scale

	Mean	Std. Deviation	N
i1	5.09	1.385	5817
i2	3.42	1.681	5817
i3	3.32	1.577	5817
i4	4.80	1.385	5817
i5	5.11	1.312	5817
i6	4.93	1.391	5817

[13] It is acknowledged that the practice of deleting more than one item at a time is not methodological best practice since the response of the second problematic item to the deletion of the first item cannot be determined.

	i1	i2	i3	i4	i5	i6
i1	1.000					
i2	.317	1.000				
i3	.187	.123	1.000			
i4	.606	.365	.222	1.000		
i5	.490	.252	.162	.535	1.000	
i6	.567	.305	.200	.656	.702	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
i1	21.589	25.331	.616	.433	.713
i2	23.257	26.637	.367	.152	**.782**
i3	23.357	29.357	.233	.058	**.811**
i4	21.880	24.444	.691	.533	.693
i5	21.576	26.039	.602	.509	.718
i6	21.749	24.319	.698	.618	.691

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.447	3.325	5.106	1.781	1.536	.703	6
Item Variances	2.134	1.723	2.826	1.104	1.641	.181	6
Inter-Item Correlations	.379	.123	.702	.579	5.701	.037	6

4.2.1.15 Work pace (T) scale

The results for the item analysis for the *Work pace* subscale are depicted in Table 4.18. This scale obtained the highest alpha coefficient (.831) of all the scales (see Table 4.1). Further inspection showed that the item means ranged from 4.89 to 5.49 on a 7-point scale and the item standard deviations from 1.088 to 1.672. In the inter-item correlation matrix all the items, except for item t1, correlated moderately (r_{ij}>.30) to high (r_{ij}>.50) with each other. All the corrected item total correlations were larger than .30. As is evident from Table 4.18, item t1 was the only item where the squared multiple correlation was smaller than .30. Furthermore it was indicated that the deletion of item t1 would increase the alpha from .831 to .855 whilst none of the other items, if deleted, would result in an increase in the alpha value. Item t1 was therefore flagged as a problematic item that lowered the homogeneity of the scale.

The analysis was run again to investigate whether the deletion of item t1 brought additional problematic items to the fore. The item-total statistics revealed that none of the remaining items, if deleted, would further increase the alpha coefficient of .855. Item t1 was therefore the only item that should be considered to be deleted from the T-subscale.

Table 4.18

Item analysis results for the Work pace scale

	Mean	Std. Deviation	N
t1	5.18	1.672	5817
t2	4.90	1.377	5817
t3	5.49	1.088	5817
t4	5.19	1.246	5817
t5	5.13	1.277	5817
t6	4.89	1.321	5817

	t1	t2	t3	t4	t5	t6
t1	1.000					
t2	.312	1.000				
t3	.381	.498	1.000			
t4	.435	.532	.656	1.000		
t5	.287	.462	.517	.600	1.000	
t6	.265	.414	.483	.548	.764	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
t1	25.59	25.347	.418	.210	**.855**
t2	25.88	25.393	.568	.346	.812
t3	25.28	26.510	.668	.486	.796
t4	25.58	24.502	.740	.574	.778
t5	25.64	24.771	.690	.640	.787
t6	25.88	24.991	.639	.600	.797

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.129	4.887	5.490	.603	1.123	.050	6
Item Variances	1.801	1.185	2.796	1.611	2.360	.295	6
Inter-Item Correlations	.477	.265	.764	.500	2.888	.018	6

4.2.1.16 Need to be forceful (K) scale

Table 4.19 presents the results of the item analysis for the K-subscale. The *Need to be forceful* scale returned an unacceptable alpha coefficient of .424. This was the lowest alpha coefficient of all the subscales. The results showed the item means to range from 3.43 to 5.85 and the item standard deviation to range from 1.415 to 1.921. Table 4.19 further showed a worrisome lack of coherence in the set of items which were all meant to reflect factor K. The low (and at times negative) inter-item correlations, the low item-total correlations and squared multiple correlations indicated that the items comprising this subscale do not respond in unison to systematic differences in a single underlying latent variable.

Furthermore, two items (k3 and k6) were flagged as problematic that contributed towards lowering the homogeneity of the scale. The low inter-item correlations, the relative magnitude of the corrected item-total correlations (.092 and .080), the squared multiple correlations (.021 and .032) and the increase in alpha affected by the removal of item k3 (from .424 to .463) would warrant the deletion of these items from the subscale.

The analysis was run again to investigate whether the deletion of these two items brought additional problematic items to the fore. After item k3 and item k6 were deleted, the alpha coefficient increased from .463 to .501. The item-total statistics, however' revealed that none of the remaining items, if deleted, would further increase Cronbach's alpha. It was therefore concluded that only these two items should be considered to be deleted from this scale.

Table 4.19

Item analysis results for the Need to be forceful scale

	Mean	Std. Deviation	N
k1	5.85	1.172	5817
k2	4.44	1.614	5817
k3	4.47	1.921	5817
k4	4.76	1.415	5817
k5	3.43	1.775	5817
k6	4.69	1.626	5817

	k1	k2	k3	k4	k5	k6
k1	1.000					
k2	.319	1.000				
k3	.069	.060	1.000			
k4	.197	.216	.056	1.000		
k5	.143	.220	.116	.157	1.000	
k6	.163	.041	-.041	.059	.055	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
k1	21.79	19.144	.330	.144	.330
k2	23.20	17.262	.301	.151	.319
k3	23.17	18.697	.092	.021	**.463**
k4	22.88	18.903	.244	.077	.360
k5	24.21	17.001	.257	.075	.344
k6	22.95	20.093	.080	.032	**.452**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.607	3.428	5.849	2.421	1.706	.602	6
Item Variances	2.578	1.375	3.690	2.316	2.685	.670	6
Inter-Item Correlations	.122	-.041	.319	.359	-7.787	.008	6

4.2.1.17 Emotional restraint (E) scale

The results for the item analysis for the *Emotional restraint* scale are depicted in Table 4.20. This subscale indicated a questionable alpha coefficient of .657. The item means ranged from 3.70 to 5.45 and the standard deviations from 1.288 to 1.849. In the inter-item correlation matrix, a number of items failed to correlate at least moderately (r_{ij}>.30) with the other items in the subscale. Furthermore, item e1 and item e3 were flagged as problematic as they contributed towards lowering the homogeneity of the scale. The low inter-item correlations, the relative magnitude of the corrected item-total correlations (.258 and .226), the squared multiple correlations (.074 and .064) and the increase in alpha affected by the removal of item e1 (from .657 to .672) would have justified deleting these items from the subscale.

The analysis was re-run to investigate whether the deletion of item e1 and item e3 brought additional problematic items to the fore. After the two items were deleted a Cronbach alpha of .709 was obtained. This, however, resulted in item e2 coming to the fore as a problematic item. The relative magnitude of the corrected item-total correlation (.384), the squared multiple correlation (.149) and the increase in alpha affected by the removal of this item (.717 from .709) would have justified the removal of this item from the E-subscale. After a further re-run of the analysis after the deletion of item e2, the results suggested that none of the remaining items, if deleted, would further increase the alpha coefficient.

Table 4.20

Item analysis results for the Emotional restraint scale

	Mean	Std. Deviation	N
e1	3.70	1.849	5817
e2	4.57	1.550	5817
e3	4.90	1.577	5817
e4	4.90	1.553	5817
e5	5.09	1.432	5817
e6	5.45	1.288	5817

	e1	e2	e3	e4	e5	e6
e1	1.000					
e2	.205	1.000				
e3	.083	.134	1.000			
e4	.200	.337	.218	1.000		
e5	.213	.308	.136	.486	1.000	
e6	.151	.272	.199	.425	.471	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
e1	24.91	23.822	.258	.074	**.672**
e2	24.03	23.585	.392	.165	.613
e3	23.70	25.778	.226	.064	**.671**
e4	23.71	21.776	.531	.328	.560
e5	23.52	22.839	.512	.340	.573
e6	23.16	24.177	.478	.289	.590

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.768	3.698	5.448	1.750	1.473	.356	6
Item Variances	2.404	1.659	3.417	1.758	2.060	.343	6
Inter-Item Correlations	.256	.083	.486	.403	5.873	.015	6

4.2.1.18 Need to achieve (A) scale

Table 4.21 presents the results of the item analysis for the A-subscale. The *Need to achieve* scale returned a poor alpha coefficient of .508. The item means ranged from 3.54 to 6.51 and the item standard deviations from .750 to 1.898. The results (Table 4.21) revealed a somewhat incoherent set of items which do not seem to respond in unison to systematic differences in a single underlying latent variable, although they were all meant to reflect factor A. This can be seen from the low (and at times negative) inter-item correlations, the low item-total correlations and low squared multiple correlations.

Furthermore, item a1 and item a5 were flagged as problematic items. Their low inter-item correlations, the relative magnitude of the corrected item-total correlations (.203 and .169), the squared multiple correlations (.044 and .074) and the increase in alpha affected by the removal of item a5 (from .508 to .532) would have warranted deleting these items from the subscale.

The analysis was run again to investigate whether the deletion of these two items brought additional problematic items to the fore. After item a1 and item a5 were deleted the alpha coefficient increased from .508 to .632. The item-total statistics, however revealed that none of the remaining items if deleted would further increase the alpha coefficient and should thus item a1 and item a5 should be the only two items that should be considered to be deleted from this scale.

Table 4.21

Item analysis results for the Need to achieve scale

	Mean	**Std. Deviation**	**N**
a1	4.54	1.898	5817
a2	6.28	.844	5817
a3	6.51	.750	5817
a4	6.00	1.065	5817
a5	3.54	1.738	5817
a6	5.31	1.357	5817

	a1	a2	a3	a4	a5	a6
a1	1.000					
a2	.073	1.000				
a3	.101	.346	1.000			
a4	.131	.348	.410	1.000		
a5	.114	-.020	.035	.058	1.000	
a6	.180	.204	.259	.387	.250	1.000

	Scale Mean if Item Deleted	**Scale Variance if Item Deleted**	**Corrected Item-Total Correlation**	**Squared Multiple Correlation**	**Cronbach Alpha if Item Deleted**
a1	27.63	12.525	.203	.044	**.525**
a2	25.89	16.438	.250	.177	.478
a3	25.66	16.297	.330	.224	.459
a4	26.17	14.581	.386	.289	.415
a5	28.63	13.661	.169	.074	**.532**
a6	26.87	12.797	.435	.227	.368

	Mean	**Minimum**	**Maximum**	**Range**	**Maximum / Minimum**	**Variance**	**N of Items**
Item Means	5.362	3.539	6.509	2.970	1.839	1.311	6
Item Variances	1.812	.562	3.601	3.039	6.405	1.578	6
Inter-Item Correlations	.192	-.020	.410	.430	-20.797	.018	6

4.2.1.19 Need to be supportive (F) scale

The results for the item analysis for the F-subscale are depicted in Table 4.22. The *Need to be supportive* scale obtained a questionable alpha coefficient of .675. Further inspection of the results showed that the item means ranged from 4.29 to 6.09 and the item standard deviations from 1.115 to 1.683. The results revealed a somewhat incoherent set of items which, although they were all meant to reflect factor F, nonetheless did not seem to respond in unison to systematic differences in the single underlying latent variable. This is apparent from the low inter-item correlations and the low squared multiple correlations in Table 4.22.

Item f6 was flagged as a problematic item that lowered the homogeneity of the subscale. The low inter-item correlations, the relative magnitude of the corrected item-total correlation (.107), the squared multiple correlation (.043) and the substantial increase in alpha that would be affected by the removal of this item (from .675 to .722) would have warranted the deletion of this item. Moreover, the deletion of item f6 may reveal additional items as problematic that are currently not shown as clear cut problematic items in Table 4.22.

The analysis was re-run to investigate whether the deletion of item f6 brought additional problematic items to the fore or not. After item f6 was deleted a Cronbach alpha of .722 was obtained. Moreover, the item-total statistics revealed that the removal of item f4, would further increase the alpha coefficient from .722 to .745. After a second run of the analysis, the deletion of item f4 now resulted in item f3 being flagged as a problematic item. The squared multiple correlation (.177) and the increase in alpha affected by the removal of this item (.755 from .745) would have justified the deletion of this item from the F-subscale. An inspection of the results further revealed that the alpha coefficient would, in the case of the deletion of any of the other items, not increase. The sequential eroding of items from the subscale illustrated by these results suggested the presence of a second factor that explains somewhat less variance in the item data on which the three problematic items load. The results obtained on the exploratory factor analysis reported in paragraph 4.4.3.19 supported this inference.

Table 4.22

Item analysis results for the Need to be supportive scale

	Mean	Std. Deviation	N
f1	5.08	1.535	5817
f2	4.29	1.683	5817
f3	5.89	1.115	5817
f4	5.62	1.169	5817
f5	4.63	1.666	5817
f6	6.09	1.343	5817

	f1	f2	f3	f4	f5	f6
f1	1.000					
f2	.472	1.000				
f3	.395	.264	1.000			
f4	.186	.157	.244	1.000		
f5	.569	.485	.339	.223	1.000	
f6	.081	.000	.160	.145	.038	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
f1	26.52	18.214	.588	.411	.562
f2	27.31	18.675	.465	.297	.611
f3	25.71	22.309	.448	.214	.626
f4	25.98	23.665	.285	.095	.668
f5	26.97	17.569	.567	.401	.567
f6	25.51	25.032	.107	.043	**.722**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.267	4.293	6.089	1.796	1.418	.514	6
Item Variances	2.063	1.244	2.833	1.589	2.277	.483	6
Inter-Item Correlations	.250	.000	.569	.569	-1282.547	.028	6

4.2.1.20 Role of the hard worker (G) scale

Table 4.23 presents the results of the item analysis for the G-subscale. The *Role of the hard worker* scale obtained a questionable alpha coefficient of .638 as depicted in Table 4.1. Further investigation showed that the item means ranged from 3.63 to 6.35 and the item standard deviations from .812 to 1.748. The inter-item correlation matrix further revealed that all the items, except for item g3, correlated below .50. Item g6 returned the lowest correlations ranging from .048 to .083. Besides item g6, all the corrected item total correlations were larger than .30. When regressing each item on a weighted linear composite of the remaining variables, interesting observations were made. While the squared multiple correlations further suggested that item g6 was a poor item, items g1 and g4 also returned values lower than .30. Item g6 was, however, flagged as the only item that showed itself as problematic due to the low inter-item correlations along with the relative magnitude of the corrected item-total correlation (.090), the squared multiple correlation (.010) and a substantial increase in alpha affected by the removal of this item (from .638 to .767). This suggested that item g6 was not successfully reflecting the same underlying latent variable than the majority of the items in the G-subscale were reflecting.

Usually when an instrument's items is used to test a structural model one has created, one would delete such item (g6) to create psychometrically satisfactory measures to credibly test the merits of the model. However, due to the nature of this study, the above mentioned item was retained for subsequent analysis.

The analysis was run again to investigate whether the deletion of item g6 brought any additional problematic items to the fore. The results revealed that none of the remaining items, if deleted, would further increase the alpha coefficient of .767 and g6 should thus be the only item to be considered to be deleted from the G-subscale.

Table 4.23

Item analysis results for the Role of the hard worker scale

	Mean	**Std. Deviation**	**N**
g1	6.29	.831	5817
g2	6.02	1.015	5817
g3	5.70	1.244	5817
g4	6.35	.811	5817
g5	5.86	1.097	5817
g6	3.63	1.748	5817

	g1	g2	g3	g4	g5	g6
g1	1.000					
g2	.417	1.000				
g3	.304	.501	1.000			
g4	.299	.389	.371	1.000		
g5	.323	.487	.493	.412	1.000	
g6	.083	.071	.048	.058	.074	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
g1	27.56	14.221	.409	.211	.591
g2	27.83	12.503	.550	.390	.535
g3	28.15	11.734	.493	.348	.543
g4	27.49	14.130	.441	.243	.584
g5	27.98	12.237	.526	.359	.537
g6	30.22	13.272	.090	.010	**.767**

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	5.641	3.627	6.355	2.728	1.752	1.036	6
Item Variances	1.364	.659	3.057	2.398	4.642	.797	6
Inter-Item Correlations	.289	.048	.501	.453	10.477	.029	6

4.2.1.21 Social Desirability (SD) scale

The results for the item analysis for *Social Desirability* scale are depicted in Table 4.24. The SD scale returned an acceptable alpha coefficient of .789. The results further showed a more coherent set of items which seems to respond in unison to systematic differences in a single underlying latent variable. This was evident from the higher (although at times modest) inter-item correlations, the higher item-total correlations and squared multiple correlations in Table 4.24. The subscale alpha would decrease for all subscale items if any one of them were to be deleted. The small squared multiple correlations, however, point to the fact that the items are not without problems in the sense that they share relatively small proportions of variance despite the fact that they were developed to all reflect a single underlying latent personality dimension.

Table 4.24

Item analysis results for the Social desirability scale

	Mean	Std. Deviation	N
q1	2.80	1.552	5817
q2	3.35	1.655	5817
q3	3.83	1.877	5817
q4	2.61	1.454	5817
q5	2.63	1.687	5817
q6	3.42	1.822	5817

	q1	q2	q3	q4	q5	q6
q1	1.000					
q2	.376	1.000				
q3	.390	.396	1.000			
q4	.446	.360	.376	1.000		
q5	.384	.403	.402	.426	1.000	
q6	.309	.344	.378	.335	.481	1.000

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach Alpha if Item Deleted
q1	15.85	37.055	.528	.299	.760
q2	15.30	36.253	.524	.277	.760
q3	14.81	33.996	.544	.297	.756
q4	16.04	37.658	.542	.310	.758
q5	16.02	34.744	.596	.367	.743
q6	15.22	34.975	.517	.293	.763

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3.108	2.612	3.833	1.222	1.468	.251	6
Item Variances	2.824	2.113	3.523	1.410	1.667	.284	6
Inter-Item Correlations	.387	.309	.481	.172	1.556	.002	6

4.3 SUMMARY OF THE ITEM ANALYSIS RESULTS

Item analysis was conducted on each of the PAPI-N subscales to determine the success with which they represent the various personality dimensions. A variety of item statistics was calculated for the items of each subscale. Two of the subscales returned good internal reliability coefficients, nine subscales returned acceptable coefficients, seven subscales returned questionable coefficients, two scales returned poor coefficients and only a single subscale returned an unacceptable reliability coefficient in terms of the interpretative taxonomy shown in Table 2.6. When considering the basket of evidence provided by these item statistics it had to be concluded that for about 50% of the PAPI-N subscales the items of the subscales did not respond in unison to systematic differences in a single underlying latent variable. Internal consistency in these subscales was either questionable, poor or unacceptable. In addition, visual inspection of the inter-item correlation statistics and item-total statistics, particularly the corrected item-total correlations, squared multiple correlations and increases in alpha if the item were deleted, further revealed a number of items that contributed towards lowering the homogeneity of the scales. This would normally have resulted in the deletion of these items. However, due to the nature of this study, all items were retained for the subsequent analyses.

Inspection of the problematic items[14] indicated no logical explanation as to why the particular items were flagged as problematic. The items in question do not appear to be phrased in a manner that makes them difficult to comprehend. Neither do they appear to be worded in such a way that they clearly reflect a different facet of the personality dimension than the rest of the items of the subscale. This latter possibility was subsequently investigated more rigorously by performing exploratory factor analysis on each of the subscales of the PAPI-N.

4.4 DIMENSIONALITY ANALYSIS

As discussed in the previous chapter, unrestricted principal axis factor analysis with oblique rotation was performed on each of the PAPI-N subscales. This was done to confirm the uni-dimensionality of each subscale and to evaluate the success with which each item, along with the rest in that particular item set, measures the specific latent personality dimension it purports to measure. The purpose of the analyses would also be to recommend the removal or rewriting of items with insufficient factor loadings and where necessary, to split heterogeneous subscales into two or more homogeneous subsets of items. If the latter happens, this would require the concomitant adjustments to the underlying PAPI-N measurement model and that the item and dimensionality analysis be repeated on the newly created subscales. However, due to the nature of the study, as explained earlier, this option was not possible. Furthermore, the analyses aimed at getting a better understanding of the item functioning per scale in the questionnaire and not to re-word or to delete any of the items in question.

Spangenberg and Theron (2004) refer to Hulin, Grasgrow and Parsons (1983) who caution that the factor analysis performed on a matrix of product moment correlations might not be the most appropriate procedure for establishing the uni-dimensionality of a scale due to the danger of extracting artefact factors reflecting differences in item difficulty or variance only. To counter this shortcoming, Spangenberg and Theron (2004, p. 7) cite Schepers (1992) who

[14] The wording of the problematic items are not listed here so as to prevent the items of the PAPI-N to leak into the public domain.

argues for the need to calculate the descriptive statistics for the items of each subscale to ascertain the possibility of multiple factors appearing as an artefact of differential item characteristics like skewness. The dimensionality procedure therefore involved the calculation of the following statistics for each of the PAPI-N subscales:

- Calculation of descriptive statistics for the items of each subscale,
- Calculation of the inter-item correlation matrix, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's measure of sphericity for the items of each subscale to evaluate the factor analysability of each matrix,
- Calculation of scree plots and the eigenvalues associated with all the possible factors that could be extracted to explain the observed correlations between the items of each subscale,
- Calculation of the reproduced and residual correlation matrices for the items of each subscale given the extracted factor structure to evaluate the adequacy of the structure,
- Calculation of the factor loadings in the factor structure for the items of each subscale to evaluate the adequacy of the loadings, and
- Calculation of the rotated factor structure if more than one factor had to be extracted.

The eigenvalue-greater-than-unity rule of thumb was used to determine the number of factors to extract. SPSS 21 for Windows (2013) was used to perform a series of 20 exploratory factor analyses on the items comprising the subscales of the PAPI-N. Table 4.25 provides a summary of the results of the factor analyses. The detailed output of the dimensionality analyses is electronically available on the enclosed CD, folder: Dimensionality Analysis, in Appendix B.

The results of the dimensionality analysis are presented in the following order: (i) assessing the factor analysability of the inter-item correlation matrices, (ii) overview of the principal axis factor analyses results, and (iii) a discussion of the individual PAPI-N scale uni-dimensionality results.

4.4.1 ASSESSING THE FACTOR ANALYSABILITY OF THE INTER-ITEM CORRELATION MATRIX

Rather than presenting the results of the analysis of the factor analysability of the 20 inter-item correlation matrices separately for each of the twenty subscales, the findings of all twenty analyses are interpreted and summarised centrally. Reporting on the results of item analysis and dimensionality analysis performed on a sizable number of subscales tends to make for rather tedious and repetitive reading. The approach used in this study will hopefully to some degree counteract this tendency.

Several criteria were used to evaluate the factor analysability of the 20 inter-item correlation matrices. Firstly, the correlation matrix should contain numerous moderately large (r_{ij}>.30) and statistically significant (p<.05) correlations to be factor analysable. Visual inspection of the inter-item correlation matrix revealed several sizable (r_{ij}>.30) and significant (p<.05) correlations for all the 20 PAPI-N subscales. Some of the correlation matrices did, however, contain inter-item correlations lower than .30. The discussion of the item analysis results presented in paragraph 4.3 pointed out the specific subscales where these problems occurred.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy should be greater than .60 for the observed correlation matrix to be factor analysable. This measure approaches unity for matrices with high degree of off-diagonal correlations, with values greater than .80 indicating that the data set is likely to factor analyse well. The KMO values were all above .60 (see Table 4.25 below), varying between .639 and .865, which suggests that all the correlation matrices were factor analysable. Furthermore, Bartlett's test of sphericity was used to test the null hypothesis (at p<.001) that the correlation matrix is an identity matrix in the parameter. In the case of all 20 scales, the identity matrix null hypothesis could be rejected which means that the off-diagonal elements in the correlation matrix in the parameter are not zero and that the matrix is factor analysable.

4.4.2 OVERVIEW OF THE PRINCIPAL AXIS FACTOR ANALYSES RESULTS

The results of the principal axis factor analyses are depicted in Table 4.25. Seven of the twenty subscales failed the uni-dimensionality test. The affected scales were: (i) W – *Need for Rules and Supervision*, (ii) R – *Conceptual Thinker*, (iii) X – *Need to be noticed*, (iv) S –

Social Harmoniser, (v) K – *Need to be forceful*, (vi) A – *Need to achieve* and (vii) F – *Need to be supportive*. For these scales, the possibility of meaningful factor fission was investigated. Attempts were subsequently made to determine the identities of the two extracted factors based on an inspection of the common theme shared by the items loadings on each factor. In addition, descriptive statistics were reviewed to determine if the two factor structure may have emerged as an artefact of differential skewness of the items or as an artefact of systematic differences in some other descriptive statistic across items. Irrespective of whether meaningful factor fission occurred, the ability of a single factor to account for the observed inter-item correlation matrix was nevertheless still investigated. This was done to examine the magnitude of the factor loadings when a single factor is forced and to examine the credibility of such a solution based on the magnitude of the residual correlations.

A summary of the item factor loadings for each subscale are also provided in Table 4.25. The following criteria were used to interpret the magnitude of the factor loadings: In terms of guidelines provided by Hair et al. (2006) .30 to .40 are considered acceptable, .50 or higher are considered significant and loadings above .70 are considered indicative of well-defined structure. In the case of personality research, the general accepted cut-off value of .30 was used as a benchmark for these analyses (Laher, 2010).

Item factor loadings from the different subscales were generally satisfactory. Eighty-nine of the 126 items (70.63%) obtained loadings higher than .50, 25 of the items (19.84%) obtained loadings between .30 and .49, and the remaining 12 of the items (9.52%) obtained loadings lower than the benchmark cut-off value of .30. The problematic items were: p2 (.264), r1 (.204), r6 (.296), x6 (.140), s1 (-.065), o3 (.133), i3 (.254), k3 (.231), k6 (.267), e3 (.274), a1 (.246) and g6 (.106)[15]. It is interesting that 5 out of the 10 scales affected by these items with low factor loadings passed the uni-dimensionality test. All twelve of these items were also flagged as problematic during the item analysis. An additional 9 items were flagged as problematic in the item analyses[16]. A more detailed discussion of the results of the dimensionality analyses will subsequently be presented.

---

[15] Items has been recorded according to the name of the subscale as to protect the scoring key of the PAPI-N. Items are recorded as such: p1, p2, p3, p4, p5, p6 for the P-subscale; l1, l2, l3, l4, l5, l6 for the L-subscale etc. Only the Social Desirability scales has been recorded differently: q1, q2, q3, q4, q5, q6

[16] In total 21 items were flagged as problematic in the item analysis.

Table 4.25

Summary of the results of the principal axis factor analyses

Sub-Scale	Determinant	KMO	Variance Explained	Min factor loading	Max factor loading	% Residual r > .05
P	.486	.755	25.941	.264	.639	13
L	.304	.811	34.268	.355	.753	6
C	.183	.852	41.863	.377	.836	0
H	.092	.861	49.288	.416	.836	13
D	.191	.861	41.718	.310	.720	0
W	.294	.780	**Factor 1: 34.923**	.434	.777	
			Factor 2: 7.256	.490	.737	0
			Single Factor: 33.499	.391	.741	53
R	.455	.714	**Factor 1: 27.732**	.282	.827	
			Factor 2: 6.642	.215	.692	6
			Single Factor: 25.760	.197	.620	53
Z	.214	.843	39.942	.451	.721	26
N	.173	.865	43.014	.332	.739	0
X	.201	.792	**Factor 1: 39.931**	.151	.819	
			Factor 2: 5.595	-.392	-.841	13
			Single Factor: 38.461	.138	.771	40
B	.170	.841	42.003	.311	.829	6
S	.340	.747	**Factor 1: 32.564**	.495	.745	
			Factor 2: 6.311	-.344	.312	0
			Single Factor: 30.904	-.065	.715	53
O	.324	.769	32.363	.133	.778	13
I	.130	.819	43.559	.254	.850	6
T	.073	.828	**Factor 1: 51.203**	.522	.772	
			Factor 2: 7.408	-.850	-.864	0
			Single Factor: 49.511	.453	.821	53
K	.747	.639	**Factor 1: 16.327**	.241	.436	
			Factor 2: 3.406	-.560	.076	0
			Single Factor: 15.536	.140	.571	26
E	.410	.772	29.179	.274	.699	13
A	.513	.696	**Factor 1: 24.525**	.573	.624	
			Factor 2: 7.264	.235	.564	0
			Single Factor: 23.513	.150	.708	40
F	.346	.757	**Factor 1: 31.449**	.668	.755	
			Factor 2: 5.758	.341	.415	0
			Single Factor: 30.921	.117	.761	26
G	.306	.812	34.232	.103	.732	0
SD	.240	.848	38.820	.591	.687	20

4.4.3 DISCUSSION OF THE INDIVIDUAL PAPI-N DIMENSIONALITY RESULTS

4.4.3.1 Need to control others (P) scale uni-dimensionality results

The design intention was that the 6 items written for the *Need to control others* subscale should all reflect a single underlying personality dimension. Although item p2 and item p1 were found to be poor items in the item analysis, due to the nature of this study, all items were included in the dimensionality analysis of *the Need to control others* subscale. The SPSS exploratory factor analysis results suggested that a single factor is required to satisfactorily explain the observed correlations between the items of the subscale. Only one factor obtained an eigenvalue greater than one. The scree plot also suggested the extraction of a single factor.

The factor matrix indicated that all the items, besides item p2 which loaded .264, loaded reasonably acceptable (.358 $< \lambda <$ .639) onto one factor. Item p2 was also flagged as problematic during the item reliability analysis. Table 4.26 presents the extracted factor structure. Furthermore, there were only 2 (13%) non-redundant residuals with absolute values greater than .05 suggesting that the rotated solution provided a credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the P-subscale.

Table 4.26

Factor matrix for the P-subscale

	Factor 1
Item p1	.358
Item p2	.264
Item p3	.639
Item p4	.615
Item p5	.548
Item p6	.522

4.4.3.2 Leadership role (L) scale uni-dimensionality results

The uni-dimensionality assumption that the 6 items comprising the *Leadership role* subscale all reflect a single underlying personality factor was tested. The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.27. The scree plot, in-line with the above, also suggested the extraction of a single factor. Items l1, l3 and l4 have satisfactory loadings on the single factor ($\lambda >$ .50). The factor matrix further indicated that reasonably acceptable loadings for items l2 and l5 were obtained, while item l6 had a borderline loading approaching .40. Based on the combined evidence generated by the item and factor analyses it was decided to flag item l6, as a questionable item for future analysis on the PAPI-N but to retain this item in the subscale for the current analysis. Furthermore, in the residual correlation matrix only one (6%) of the residual correlations was greater than .05 suggesting that the factor solution provides a credible explanation for the observed inter-item correlation matrix. The uni-dimensionality assumption was therefore corroborated for this subscale.

Table 4.27

Factor matrix for the L-subscale

	Factor 1
Item l1	.753
Item l2	.406
Item l3	.710
Item l4	.668
Item l5	.497
Item l6	.355

4.4.3.3 Organised type (C) scale uni-dimensionality results

The design intention was that the 6 items written for the *Organised type* subscale should all reflect a single underlying personality dimension. The results suggested that a single factor was required to satisfactorily explain the observed correlations between the items of the subscale. Only one factor obtained an eigenvalue greater than unity. The scree plot also suggested the extraction of a single factor. The factor matrix indicated that all the items, besides item c1 which loaded .377, loaded satisfactory ($.569 < \lambda < .836$) onto one factor. Item c1 was also flagged as problematic during the item reliability analysis. Table 4.28 presents the extracted factor structure. Furthermore, none (0%) of the residual correlations were greater than .05 suggesting that the rotated solution provided a very credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the C-subscale.

Table 4.28

Factor matrix for the C-subscale

	Factor 1
Item c1	.377
Item c2	.700
Item c3	.569
Item c4	.836
Item c5	.600
Item c6	.706

4.4.3.4 Integrative planner (H) scale uni-dimensionality results

The uni-dimensionality assumption that the 6 items comprising the *Integrative planner* subscale all reflect a single underlying personality factor was tested. The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue larger than one. The scree plot also suggested the extraction of a single factor. The factor matrix revealed that all the items, besides item h1 (λ = .416), loaded satisfactorily on the single extracted factor as all the remaining factor loadings were greater than .50. Item h1 was also flagged as problematic during the item reliability analysis. The extracted factor structure is shown in Table 4.29. Furthermore, there were only 2 (13%) non-redundant residuals with absolute values greater than .05 suggesting that the rotated solution provided a credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated.

Table 4.29

Factor matrix for the H-subscale

	Factor 1
Item h1	.416
Item h2	.727
Item h3	.645
Item h4	.774
Item h5	.736
Item h6	.836

4.4.3.5 Attention to detail (D) scale uni-dimensionality results

The design intention was that the 6 items written for the *Attention to detail* subscale should all reflect a single underlying personality dimension. The results suggested that a single factor was required to satisfactorily explain the observed correlations between the items of the D-subscale. Only one factor had an eigenvalue greater than one. The scree plot also suggested the extraction of a single factor. The factor matrix indicated that all the items, except item d1 which loaded .310, loaded satisfactory ($.649 < \lambda < .720$) onto one factor. Item d1 was also flagged as problematic during the item reliability analysis. Table 4.30 presents the extracted factor structure. Furthermore, none (0%) of the residual correlations were greater than .05 suggesting that the rotated solution provided a very credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the D-subscale.

Table 4.30

Factor matrix for the D-subscale

	Factor 1
Item d1	.310
Item d2	.707
Item d3	.672
Item d4	.649
Item d5	.720
Item d6	.719

4.4.3.6 Need for rules and supervision (W) scale uni-dimensionality results

The design intention was that the 6 items written for the *need for rules and supervision* subscale should all reflect a single underlying personality dimension. The results suggested that two factors were required to satisfactorily explain the observed correlations between the items of the W-subscale. Two factors obtained eigenvalues greater than one. The scree plot also suggested the extraction of two factors.

The two factors that that were extracted for the W – (Need for rules and supervision) subscale could be meaningfully interpreted. Factor 1 seemed to focus on *an individual's preference for receiving clear guidelines and structure before or when doing a job*, whereas Factor 2 tended to focus more on *an employee's adherence to follow rules and procedures at work*. Both these dimensions seemed to represent logical facets of an overarching *need for rules and supervision*. Table 4.31a shows the items that load on the respective factors. The descriptive statistics calculated for the items of the W-subscale (Table 4.31b) suggested that it was highly unlikely that the two factors emerged as an artefact of differential item characteristics of the items. The item descriptive statistics revealed that all the items were consistently significantly negatively skew ($p<.05$) and consistently significantly ($p<.05$) leptokurtic. Differential item statistics therefore do not provide a plausible explanation for the extracted factor matrix. Table 4.31a moreover suggested that neither could the emergence of a second factor be explained in terms of the problematic items that emerged in the item analysis.

Table 4.31a

Factor loadings for the W-subscale (rotated pattern matrix)

	Factor	
	1	2
Item w1	**.423**	.165
Item w2	.207	**.699**
Item w3	**.717**	.169
Item w4	.464	**.571**
Item w5	**.647**	.250
Item w6	.128	**.461**

a. Rotation converged in 3 iterations.

Table 4.31b

Descriptive statistics for the W-subscale

	Item	w1	w2	w3	w4	w5	w6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	6.34	5.85	5.81	5.60	5.64	5.67
	Std. Deviation	.994	1.059	1.183	1.173	1.168	1.454
	Variance	.987	1.121	1.399	1.375	1.364	2.114
	Skewness	-2.436*	-1.347*	-1.437*	-1.135*	-1.362*	-1.302*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	7.878*	2.474*	2.397*	1.511*	2.356*	1.093*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

To examine how well the 6 items represent a single underlying factor the item are meant to represent, a single factor was forced on the data. When forcing a single factor reasonably acceptable factor loadings ($.391 < \lambda < .741$) for all items in this subscale were obtained. The results are shown in Table 4.31c below. Item w6 that came to the fore as a problematic item in the item analysis again expressed itself as a marginal item in Table 4.31a and Table 4.31c.

Table 4.31c

Factor matrix when forcing the extraction of a single factor (W-subscale)

	Factor
	1
Item w1	.433
Item w2	.565
Item w3	.618
Item w4	.741
Item w5	.647
Item w6	.391

1 factors extracted. 7 iterations required.

The residuals correlations (the discrepancy between the observed and reproduced inter-item correlations) were examined for both the two-factor and the one-factor solutions. For the two-factor solution, all non-redundant residuals had absolute values less than .05 thus suggesting that the rotated factor solution provides an extremely convincing explanation for the observed inter-item correlation matrix. This solution explained 60.82% of the total subscale variance in the initial solution but only 42.179% of the observed variance in the extracted solution. For the one-factor solution a large percentage (53%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the forced factor solution does not provide a convincing explanation for the observed inter-item correlation matrix. The single extracted factor accounted for 33.499% of the total subscale variance. This would imply that the initial factor fissure solution provides a more plausible explanation of the observed correlation matrix, and that there is little support for the design assumption that all items comprising the W-subscale reflect one inseparable underlying theme. The items comprising the W-subscale seem to reflect two latent dimensions of a second-order *need for rules and supervision* personality factor. The fitting of a second-order measurement model in which two first-order *need for rules and supervision* factors load on a single second-order factor could shed more light on the validity of this hypothesis.

4.4.3.7 Conceptual thinker (R) scale uni-dimensionality results

The design intention was that the 6 items written for the *conceptual thinker* subscale should all reflect a single underlying personality dimension. The results suggested that two factors were required to satisfactorily explain the observed correlations between the items of the R-subscale. Two factors obtained eigenvalues greater than one. The scree plot also suggested the extraction of two factors.

The two factors that were extracted in the dimensionality analysis (Table 4.32a) for the Conceptual thinker scale were two conceptually meaningful factors. Factor 1 seemed to represent *an individual's tendency to come up with new and creative ideas/solutions*. Factor 2 could be interpreted to represent *an individual's tendency to reflect on theoretical issues and concepts*. The two extracted factors seem to be logical sub-dimensions of the *Conceptual thinker* factor. Upon review of the descriptive statistics (Table 4.32b), all the items showed statistically significant negative skewness ($p<.05$) and significantly positive kurtosis ($p<.05$) thus suggesting that the extracted factor structure was not a mere artefact of differential item statistics, but rather reflected the fact that the items of the subscale assessed two distinct substantive factors. The results therefore pointed to meaningful factor fission.

Table 4.32a

Factor loadings for the R-subscale (rotated pattern matrix)

	Factor	
	1	2
Item r1	.075	**.204**
Item r2	.225	**.609**
Item r3	.226	**.652**
Item r4	**.590**	.240
Item r5	**.748**	.180
Item r6	**.296**	.179

a. Rotation converged in 3 iterations.

Table 4.32b

Descriptive statistics for the R-subscale

	Item	r1	r2	r3	r4	r5	r6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	5.46	4.98	5.08	5.72	5.76	4.71
	Std. Deviation	1.500	1.375	1.266	.991	.945	1.353
	Variance	2.251	1.890	1.603	.983	.893	1.830
	Skewness	-1.104*	-.775*	-.855*	-1.037*	-1.148*	-.578*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	.463*	.124	.416*	1.750*	2.625*	-.151*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

When forcing the extraction of a single factor, factor loadings ranged between .197 and .620. Of concern were items r1 and r6 that returned low factor loadings in the extracted single-factor (.197; .353) solution. Item r1 and item r6 were both flagged as problematic in the item reliability analysis (sub-section 4.2.1.7). The results of the one-factor solution are shown in Table 4.32c below.

Table 4.32c

Factor matrix when forcing the extraction of a single factor (R-subscale)

	Factor
	1
Item r1	.197
Item r2	.555
Item r3	.569
Item r4	.606
Item r5	.620
Item r6	.353

1 factors extracted. 6 iterations required.

The residual correlations were examined for both the two-factor and the one-factor solutions. For the two-factor solution, a small percentage (6%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the rotated solution provided a reasonably credible explanation for the observed correlation matrix. The two-factor solution explained 54.30% of the total subscale variance in the initial solution but only 34.374% of the observed variance in the extracted solution. For the one-factor solution, a large percentage (53%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the forced factor solution did not provide a credible explanation for the observed correlation matrix. The one-factor solution accounted for 45.526% of the total subscale variance. The foregoing basket of evidence suggested that the initial factor solution provided a more plausible explanation of the observed inter-item correlation matrix. There was little evidence to support the design assumption that all items comprising the R-subscale reflect a single inseparable underlying theme. The position that the items comprising the R-subscale reflect two latent dimensions of a second-order *conceptual thinker* personality factor does, however seem plausible. Formally fitting a second-order measurement model in a confirmatory factor analysis could provide further credence to this position.

4.4.3.8 Need for change (Z) scale uni-dimensionality results

The uni-dimensionality assumption that the 6 items comprising the *need for change* subscale all reflect a single underlying personality factor was tested. The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.33. The scree plot, in-line with the above, also suggested the extraction of a single factor. The factor matrix indicated that all the items, besides item z5 (λ = .451), loaded satisfactorily on the single extracted factor, as all factor loadings were larger than .50. Item z5 was also flagged as problematic during the item reliability analysis. Furthermore, 4 (26%) of the reproduced correlations were larger than .05, implying that the factor solution provided a reasonably credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated.

Table 4.33

Factor matrix for the Z-subscale

	Factor 1
Item z1	.665
Item z2	.630
Item z3	.679
Item z4	.611
Item z5	.451
Item z6	.721

4.4.3.8 Need to finish a task (N) scale uni-dimensionality results

The dimensionality analysis (Table 4.34) for the *Need to finish a task* scale resulted in the extraction of a single factor since only a single factor obtained an eigenvalue greater than one. The scree plot also suggested the extraction of one factor. The factor matrix indicated that all the items, except item n1 which loaded .332, loaded satisfactory ($.664 < \lambda < .739$) onto the single factor. Item n1 was also flagged as problematic during the item reliability analysis. Table 4.34 presents the extracted factor structure. Furthermore, none (0%) of the residual correlations were greater than .05 suggesting that the rotated solution provided a very credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the N-subscale.

Table 4.34

Factor Matrix for the N-subscale

	Factor 1
Item n1	.332
Item n2	.685
Item n3	.664
Item n4	.739
Item n5	.687
Item n6	.736

4.4.3.10 Need to be noticed (X) scale uni-dimensionality results

For the X - *Need to be noticed* –subscale two factors emerged. The results indicated that two factors should be extracted since two factors obtained an eigenvalue greater than one. The inflection point in the scree plot corroborated this result. From the rotated factor matrix (Table

4.35a), a pattern of loadings emerged that only allowed a somewhat tenuous stance on the identity of the two extracted factors. Factor 1 had three items (x3, x4 and x5) with factor loadings greater than .50. Three items (x1, x2 and x3) loaded on Factor 2. One item (Q125) did not load on any of the two extracted factors. Item x3 revealed itself as a complex item simultaneously loading on two factors. Factor 1 could somewhat tentatively be described as *an individual's tendency to actively take steps to draw attention to them* while Factor 2 could somewhat cautiously be interpreted as *an individual's preference to experience attention*. Descriptive statistics were reviewed to determine if the two-factor structure may be an artefact of differential skewness or kurtosis of the item distributions. The item descriptive statistics for the X-subscale is reported in Table 4.35b. Four items (x1, x2, x3 and x5) showed significant positive skewness, whilst only one (x6) of the two negatively skewed items was statistically significant (p<.05). Only one item showed significant positive kurtosis (p<.05). Differential skewness or kurtosis therefore did not offer a credible explanation for the extracted solution.

Table 4.35a

Factor loadings for the X-subscale (rotated pattern matrix)

	Factor	
	1	2
Item x1	.190	**.483**
Item x2	.324	**.759**
Item x3	**.502**	**.478**
Item x4	**.707**	.271
Item x5	**.778**	.313
Item x6	.059	.140

a. Rotation converged in 3 iterations.

Table 4.35b

Descriptive statistics for the X-subscale

	Item	x1	x2	x3	x4	x5	x6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	3.72	3.06	2.77	3.79	3.31	5.76
	Std. Deviation	1.849	1.667	1.507	1.640	1.699	1.109
	Variance	3.420	2.777	2.272	2.689	2.888	1.231
	Skewness	.183*	.490*	.671*	-.029	.264*	-1.359*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	-1.162*	-.800*	-.446*	-1.069*	-1.117*	2.541*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

When forcing the extraction of a single factor reasonably acceptable factor loadings ($.456 < \lambda < .771$) for all items were obtained except for item x6 ($\lambda = .138$). Item x6 was also flagged as problematic during the item reliability analysis (sub-section 4.2.1.10). The results for the single forced extracted factor are shown in Table 4.35c below.

Table 4.35c

Factor matrix when forcing the extraction of a single factor (X-subscale)

	Factor 1
Item x1	.456
Item x2	.698
Item x3	.714
Item x4	.699
Item x5	.771
Item x6	.138

1 factors extracted. 6 iterations required.

The residual correlations were examined for both the two-factor and the one-factor solution. For the two-factor solution a small percentage (13%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the rotated two-factor solution provided a reasonably credible explanation for the observed correlation matrix. The two-factor solution explained 63.99% of the total subscale variance in the initial solution, but only 45.526% of the observed variance in the extracted solution. For the one-factor solution, a large percentage (40%) of non-redundant residuals had absolute values greater than .05, thus suggesting that the forced factor solution did not provide a credible explanation for the observed correlation matrix. The one-factor solution accounted for 38.461% of the total subscale variance. The foregoing basket of evidence would suggest that the initial factor solution provided a more credible explanation of the observed inter-item correlation matrix. There was little evidence to support the design assumption that all items comprising the X-subscale reflect a single inseparable underlying theme.

4.4.3.11 Need to belong to groups (B) scale uni-dimensionality results

The dimensionality analysis (Table 4.36) for the *Need to belong to groups* scale resulted in the extraction of one factor. The scree plot also suggested the extraction of a single factor. The factor matrix indicated that all the items, besides item b1 ($\lambda = .311$), loaded satisfactorily on the single extracted factor, as all factor loadings were larger than .50. Item b1 was also flagged as problematic during the item reliability analysis. The extracted factor structure is shown in Table 4.36. Furthermore, in the residual correlation matrix only one (6%) of the residual correlations were greater than .05 suggesting that the factor solution provided a credible explanation for the observed inter-item correlation matrix. The uni-dimensionality assumption was therefore corroborated for the B-subscale.

Table 4.36

Factor Matrix for the B-subscale

	Factor 1
Item b1	.311
Item b2	.612
Item b3	.596
Item b4	.581
Item b5	.817
Item b6	.829

4.4.3.12 Social harmoniser (S) scale uni-dimensionality results

The uni-dimensionality assumption that the 6 items comprising the *social harmoniser* subscale all reflect a single underlying personality factor, was tested. The results indicated that two factors should be extracted since two factors obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.37a. The scree plot, in-line with the above, also suggested the extraction of two factors. The Social harmoniser scale split into two conceptually meaningful factors. Factor 1 seemed to focus on *a person's social ability to build and maintain mutually beneficial interpersonal relationships with others*, whereas Factor 2 tended to focus more on *an individual's desirability to get to know his/her fellow employees at work*. The rotated factor matrix (Table 4.37a) contains the items that loaded on the two factors. Factor 1 has three items (s3, s4 and s5) with loadings greater than .50. Two items (s2 and s6) load on Factor 2. Item s1 was the only item that did not load on any of the extracted factors. Furthermore, with s2 and s6 as the only items loading on Factor 2, the interpretation of this factor becomes somewhat tentative. The descriptive statistics (Table 4.37b) further revealed that all the distributions of all the items were significantly negatively skewed ($p<.05$) and statistically significantly leptokurtic ($p<.05$) thereby suggesting that the emergence of the two factors cannot be attributed to differential skewness or kurtosis in the items.

Table 4.37a

Factor loadings for the S-subscale (rotated pattern matrix)

	Factor	
	1	2
Item s1	-.027	-.065
Item s2	.194	**.565**
Item s3	**.525**	.170
Item s4	**.656**	.215
Item s5	**.614**	.363
Item s6	.287	**.774**

a. Rotation converged in 3 iterations.

Table 4.37b

Descriptive statistics for the S-subscale

	Item	s1	s2	s3	s4	s5	s6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	2.83	5.82	5.16	5.86	5.46	5.73
	Std. Deviation	1.518	.969	1.359	1.043	1.275	1.113
	Variance	2.304	.938	1.848	1.087	1.627	1.239
	Skewness	1.086*	-1.352*	-.819*	-1.125*	-1.004*	-1.255*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	.488*	2.990*	.081	1.732*	.831*	1.889*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

When forcing the extraction of a single factor the factor fusion resulted in a one-factor solution ($-.065 < \lambda < .715$) in which all items, but s1 showed acceptable factor loadings. Item s1 returned a negligible low loading of -.065. This item was also flagged as problematic during the item reliability analysis (sub-section 4.2.1.12). The results for the single forced factor are shown in Table 4.37c below.

Table 4.37c

Factor matrix when forcing the extraction of a single factor (S-subscale)

	Factor 1
Item s1	-.065
Item s2	.513
Item s3	.499
Item s4	.609
Item s5	.715
Item s6	.675

1 factors extracted. 7 iterations required.

The residuals correlations were examined for both the two-factor and the one-factor solutions. For the two-factor solution, all non-redundant residuals had absolute values less than .05, thus suggesting that the rotated factor solution provided an extremely convincing explanation for the observed inter-item correlation matrix. This solution explained 57.86% of the total subscale variance in the initial solution but only 38.875% of the observed variance in the extracted solution. For the one-factor solution, a large percentage (53%) of non-redundant residuals obtained absolute values greater than .05, thus suggesting that the forced factor solution did not provide a convincing explanation for the observed inter-item correlation matrix. The one extracted factor accounted for 30.904% of the total subscale variance. This would suggest that the initial factor fissure solution would provide a more plausible explanation of the observed correlation matrix, and that there is little support for the design assumption that all items comprising the S-subscale reflect one undividable underlying theme. The position that the items comprising the S-subscale reflect two latent dimensions of a second-order *social harmoniser* personality factor does, however seem plausible. Formally fitting a second-order measurement model in a confirmatory factor analysis could provide further credence to this position.

4.4.3.13 Need to relate closely to individuals (O) scale uni-dimensionality results

The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.38. The scree plot, in-line with the above, also suggested the extraction of a single factor. The results of the single factor matrix indicated satisfactory loadings for items o1, o2, o4 and o5 ($\lambda > .50$) but the loadings of items o3 and o6 fell below the acceptable loading criterion. Based on the combined evidence generated by the item and factor analyses it was decided to flag item o3 as a questionable item for future analysis on the PAPI-N, but to retain this item in the subscale for the current analysis. Furthermore, there were 2 (13%) non-redundant residuals with absolute values greater than .05, suggesting that the rotated solution provided a credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the O-subscale.

Table 4.38

Factor matrix for the O-subscale

	Factor 1
Item o1	.695
Item o2	.778
Item o3	.133
Item o4	.657
Item o5	.530
Item o6	.350

4.4.3.14 Ease in decision making (I) scale uni-dimensionality results

The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.39. The scree plot also suggested the extraction of a single factor. The results indicated satisfactory loadings for items i1, i4, i5 and i6 ($\lambda > .70$) and a reasonably acceptable loading for item i2, while item i3 fell below the acceptable loading criterion. Based on the combined evidence generated by the item and factor analyses it was decided to flag item i2 and item i3 as questionable items for future analysis on the PAPI-N, but to retain these items in the subscale for the current analysis. Furthermore, in the residual correlation matrix only one (6%) of the residual correlations were greater than .05 suggesting that the factor solution provides a credible explanation for the observed inter-item correlation matrix. The uni-dimensionality assumption was thus corroborated for the I-subscale.

Table 4.39

Factor matrix for the I-subscale

	Factor 1
Item i1	.708
Item i2	.404
Item i3	.254
Item i4	.798
Item i5	.725
Item i6	.850

4.4.3.15 Work pace (T) scale uni-dimensionality results

The dimensionality analysis (Table 4.40a) for the *Work pace* scale resulted in the extraction of one factor. The scree plot also suggested the extraction of a single factor. The factor matrix indicated that all the items, besides item t1 (λ = .453), loaded satisfactorily on the single extracted factor as all factor loadings were larger than .50. Item t1 was also flagged as problematic during the item reliability analysis. The extracted factor structure is shown in Table 4.39. There were, however, 8 (53%) non-redundant residuals with absolute values greater than .05. The credibility of the single factor solution as explanation for the observed inter-item correlation matrix was therefore somewhat tenuous. The uni-dimensionality assumption for the T-subscale was therefore not convincingly illustrated by the results depicted in Table 4.40a.

Table 4.40a

Single factor matrix for the T-subscale

	Factor 1
Item t1	.453
Item t2	.625
Item t3	.734
Item t4	.821
Item t5	.789
Item t6	.735

The high percentage large residual correlations again suggest the presence of a second factor that failed to be extracted because the eigenvalue of the second factor marginally fell below the cut-off value of 1. The extraction of a second factor was consequently forced. The obliquely rotated two-factor solution is shown in Table 4.40b.

Table 4.40b

Factor loadings for the the T-subscale (rotated pattern matrix)

	Factor	
	1	**2**
Item t1	**.566**	.084
Item t2	**.552**	-.111
Item t3	**.731**	-.056
Item t4	**.772**	-.110
Item t5	.049	**-.864**
Item t6	.002	**-.850**

Table 4.40b indicates that items t5 and t6 load on factor 2 while the remaining four items load on factor 1. The factor fission seems to be rather subtle. Items t1, t2, t3 and t4 all share the theme of the tempo at which work activities is performed whereas items t5 and t6 both refer to the time it takes to finish things. Factor 1 therefore seems to refer to *a behavioural pace or tempo factor* whereas factor 2 seems *to represent a completion time factor*. The results shown in Table 4.40 can therefore be considered a meaningful factor fission as both these two themes represent logical facets of the work pace theme that the subscale was designed to measure. The results depicted in Table 4.40b moreover provided a credible explanation of the observed inter-item correlation matrix.

4.4.3.16 Need to be forceful (K) scale uni-dimensionality results

The assumption that the 6 items comprising the *need to be forceful* subscale all reflect a single underlying personality factor was tested. Dimensionality analysis for the K – *Need to be forceful* –subscale revealed a two factor structure. The first two factors had eigenvalues greater than 1. The scree plot also suggested the extraction of a single factor. From the rotated factor matrix (Table 4.41a), no clear interpretable pattern of loadings emerged that would suggest a meaningful fission of the *need to be forceful* factor (i.e. no conceptually meaningful themes underlying the two factors could be identified). Factor 1 had three items (k2, k4 and k5) with loadings greater than .30. Two items (k1 and k2) loaded on factor 2. Item k3 and item k6 did not load on any of the two extracted factors. Furthermore, item k2 presented itself as a complex item simultaneously loading on two factors. The descriptive statistics were reviewed (Table 4.40b) and revealed that five of the six items distributions (k1, k2, k3, k4 and k6) indicated statistically significant negative skewness, whilst the distribution of only one item (k5) demonstrated significant positive skewness ($p<.05$). The distribution of only one item (k1) was significantly leptokurtic ($p<.05$). The distributions of the remaining items were all significantly platikurtic ($p<.05$). Neither the differential skewness or kurtosis corresponded to the factor loading pattern in Table 4.41a. The extracted factor structure can therefore not be ascribed to differential skewness in the items.

Table 4.41a

Factor loadings for the K-subscale (rotated pattern matrix)

	Factor	
	1	2
Item k1	.256	**.596**
Item k2	**.442**	**.326**
Item k3	.231	-.033
Item k4	**.314**	.212
Item k5	**.430**	.088
Item k6	-.021	.267

a. Rotation converged in 3 iterations.

Table 4.41b

Descriptive statistics for the K-subscale

	Item	k1	k2	k3	k4	k5	k6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	5.85	4.44	4.47	4.76	3.43	4.69
	Std. Deviation	1.172	1.614	1.921	1.415	1.775	1.626
	Variance	1.375	2.606	3.690	2.003	3.150	2.644
	Skewness	-1.461*	-.384*	-.460*	-.551*	.156*	-.282*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	2.470*	-.766*	-1.129*	-.346*	-1.237*	-1.028*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

When forcing the extraction of a single factor, the loadings of the 6 items were generally low. k1 and k2 (complex item) were the only items to return factor loadings greater than .50. Two items (k4 and k5) returned loadings higher than .30. Of concern were items k3 and k6 as they obtained loadings of less than .30. Both items were flagged as problematic in the item reliability analysis (sub-section 4.2.1.16). The results for the single forced extraction factor are shown in Table 4.41c below.

Table 4.41c

Factor matrix when forcing the extraction of a single factor (K-subscale)

	Factor
	1
Item k1	.533
Item k2	.571
Item k3	.140
Item k4	.385
Item k5	.358
Item k6	.162

1 factors extracted. 8 iterations required.

The residuals correlations were examined for both the two-factor and the one-factor solution. For the two-factor solution all non-redundant residuals had absolute values less than .05 thus suggesting that the rotated factor solution provides an extremely convincing explanation for the observed inter-item correlation matrix. This solution explained 46.13% of the total subscale variance in the initial solution but only 19.733% of the observed variance in the extracted solution. For the one-factor solution a moderate percentage (26%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the forced single factor solution did not provide a convincing explanation for the observed inter-item correlation matrix. The one extracted factor accounted for 15.536% of the total subscale variance. This would suggest that the initial factor fissure solution would statistically provide a more plausible explanation of the observed correlation matrix, and that statistically there is questionable support for the design assumption that all items comprising the K-subscale reflect one indivisible underlying theme. The fact that the extracted factors could not be meaningfully interpreted combined with the fact that only one item had a strong loading on the second factor erodes theoretical confidence in the two-factor solution.

4.4.3.17 Emotional restraint (E) scale uni-dimensionality results

The results indicated that only one factor should be extracted since only a single factor obtained an eigenvalue greater than one. The extracted factor structure is shown in Table 4.42. The scree plot also suggested the extraction of a single factor. The factor matrix results indicated satisfactory loadings for items e4, e5 and e6 ($\lambda > .50$) and reasonably acceptable loadings for items e1 and e2, while item e3 fell below the acceptable loading criterion. Based on the combined evidence generated by the item and factor analyses it was decided to flag these three items, as questionable items for future analysis on the PAPI-N, but to retain these items in the subscale for the current analysis. Furthermore, there were 2 (13%) non-redundant residuals with absolute values greater than .05 suggesting that the rotated solution provided a credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the E-subscale.

Table 4.42

Factor matrix for the E-subscale

	Factor 1
Item e1	.302
Item e2	.470
Item e3	.274
Item e4	.699
Item e5	.694
Item e6	.628

4.4.3.18 Need to achieve (A) scale uni-dimensionality results

The Need to achieve scale also presented a two-factor oblique factor structure as illustrated in Table 4.43a below. The two factors seem to distinguish between *a person's need for personal achievement* (Factor 1) and *a person's level of competitiveness (competing with the achievements of others)* (Factor 2). Factor 1 had three items (a2, a3 and a4) with loadings higher than .50. Factor 2 obtained only two items (a5 and a6) with loadings greater than .30. Although loading primarily on Factor 2, item a1 did nonetheless not load strongly on this factor. The descriptive statistics calculated for the items of the A-subscale (Table 4.43b) indicated that the two factors may have emerged as a result of the differential kurtosis of the item distributions. The item distributions of the three items that loaded on factor 2 were all statistically significantly ($p<.05$) platikurtic whereas the distributions of the three items loading on items were all statistically significantly ($p<.05$) leptokurtic. Five of the six items (a1, a2, a3, a4 and a6) displayed significant negative skewness, while one item (a5) displayed significant positive skewness ($p<.05$). The pattern of differential item skewness did, however, not correspond to the factor loading pattern in the rotated two-factor solution. The themes underlying the items loading on the two factors were, however, sufficiently convincing not to regard the two factors as merely representing artefacts of differential item statistics.

Table 4.43a

Factor loadings for the A-subscale (rotated pattern matrix)

	Factor	
	1	2
Item a1	.114	**.246**
Item a2	**.555**	.022
Item a3	**.600**	.113
Item a4	**.637**	.252
Item a5	-.055	**.446**
Item a6	.344	**.602**

a. Rotation converged in 3 iterations.

Table 4.43b

Descriptive statistics for the A-subscale

	Item	a1	a2	a3	a4	a5	a6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	4.54	6.28	6.51	6.00	3.54	5.31
	Std. Deviation	1.898	.844	.750	1.065	1.738	1.357
	Variance	3.601	.712	.562	1.134	3.020	1.841
	Skewness	-.342*	-1.928*	-2.329*	-1.471*	.159*	-1.045*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	-1.138*	6.831*	9.357*	2.799*	-1.114*	-.842*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

Upon forcing the extraction of a single-factor solution, the majority of items loaded satisfactorily on the single factor (.150 < λ < .708), however items a1 and a5 returned low loadings (.219; .150 respectively). Both items were flagged as problematic during the item reliability analysis (sub-section 4.2.1.18). The results for the single forced extraction factor are shown in Table 4.43c below.

Table 4.43c

Factor matrix when forcing the extraction of a single factor (A-subscale)

	Factor
	1
Item a1	.219
Item a2	.480
Item a3	.574
Item a4	.708
Item a5	.150
Item a6	.529

1 factors extracted. 10 iterations required.

The residuals correlations were examined for both the two-factor and the one-factor solution. For the two-factor solution all non-redundant residuals had absolute values less than .05 thus suggesting that the rotated factor solution provided an extremely convincing explanation for the observed inter-item correlation matrix. The two-factor solution explained 53.91% of the total subscale variance in the initial solution but only 31.789% of the observed variance in the extracted solution. For the one-factor solution a large percentage (40%) of non-redundant residuals had absolute values greater than .05 thus suggesting that the forced factor solution did not provide a credible explanation for the observed correlation matrix. The one extracted factor accounted for 23.513% of the total subscale variance. This would suggest that the initial factor fissure solution would provide a more credible explanation of the observed correlation matrix, and that there was little support for the design assumption that all items comprising the A-subscale reflect one inseparable underlying theme. The position that the items comprising the A-subscale reflect two latent dimensions of a second-order *need to achieve* personality factor does, however seem plausible. Formally fitting a second-order measurement model in a confirmatory factor analysis could provide further credence to this position.

4.4.3.19 Need to be supportive (F) scale uni-dimensionality results

The results of the dimensionality analysis for the *Need to be supportive* scale revealed a two-factor oblique factor structure, differentiating *a person's desire to meet his/her manager's expectations* (Factor 1) from *a person's loyalty and respect for authority* (Factor 2). Factor 1 had to do with the extent to which someone wants to help and be seen to help their manager and those in higher positions of authority. It often indicates a person's awareness of the political subtleties of the organisation (Anderson & Lewis, 1998). Factor 2 had more to do with a person's respect for authority and also his/her willingness to question his/her boss's decision if they do not agree with it. As the rotated factor matrix for the F-subscale depicted below in Table 4.44a illustrates, items f1, f2, and f5 loaded on Factor 1 whereas items f3, f4 and f6 loaded on Factor 2. Item f3 however presented itself as a complex item simultaneously loading on two factors. The item descriptive statistics (Table 4.44b) were reviewed and revealed that all item distributions showed to be significantly negatively skew ($p<.05$), thus confirming that the emergence of the two factors cannot be attributed to differential skewness.

Table 4.44a

Factor loadings for the F-subscale (rotated pattern matrix)

	Factor	
	1	2
Item f1	**.720**	.220
Item f2	**.643**	.050
Item f3	.385	**.449**
Item f4	.196	**.366**
Item f5	**.741**	.152
Item f6	-.024	**.394**

a. Rotation converged in 3 iterations.

Table 4.44b

Descriptive statistics for the F-subscale

	Item	f1	f2	f3	f4	f5	f6
N	Valid	5817	5817	5817	5817	5817	5817
	Mean	5.08	4.29	5.89	5.62	4.63	6.09
	Std. Deviation	1.535	1.683	1.115	1.169	1.666	1.343
	Variance	2.356	2.833	1.244	1.365	2.777	1.803
	Skewness	-.912*	-.379*	-1.532*	-1.079*	-.626*	-2.102*
	Std. Error of Skewness	.032	.032	.032	.032	.032	.032
	Kurtosis	.265*	-.682*	3.178*	1.351*	-.577*	4.252*
	Std. Error of Kurtosis	.064	.064	.064	.064	.064	.064

When forcing the extraction of a single factor, the loadings of the 6 items were generally moderate. Four items (f1, f2, f3 and f5) obtained loadings greater than .50 and one item (f4) returned a loading just above .30. Item f6 was the only item that returned a loading lower than .30, and was also flagged as problematic during the item reliability analysis (sub-section 4.2.1.19). The results for the single forced extraction factor are shown in Table 4.44c below.

Table 4.44c

Factor matrix when forcing the extraction of a single factor (F-subscale)

	Factor
	1
Item f1	.761
Item f2	.604
Item f3	.502
Item f4	.307
Item f5	.743
Item f6	.117

1 factors extracted. 7 iterations required.

The residuals correlations were examined for both the two-factor and the one-factor solution. For the two-factor solution all non-redundant residuals had absolute values less than .05 thus suggesting that the rotated factor solution provided an extremely convincing explanation for the observed inter-item correlation matrix. This solution explained 58.74% of the total subscale variance in the initial solution but only 37.207% of the observed variance in the extracted solution. For the one-factor solution a moderate percentage (26%) of non-redundant residuals had absolute values greater than .05, thus suggesting that the forced factor solution did not provide a convincing explanation for the observed inter-item correlation matrix. The one extracted factor accounted for 30.921% of the total subscale variance. The foregoing basket of evidence would suggest that the initial factor solution provided a more credible explanation of the observed inter-item correlation matrix. There was little evidence to support the design assumption that all items comprising the F-subscale reflect a single inseparable underlying theme. The position that the items comprising the F-subscale reflect two latent dimensions of a second-order *need to be supportive* personality factor does, however seem plausible. Formally fitting a second-order measurement model in a confirmatory factor analysis could provide further credence to this position.

4.4.3.20 Role of the hard worker (G) scale uni-dimensionality results

The dimensionality analysis (Table 4.45) for the *Role of the hard worker* scale resulted in the extraction of a single factor. The scree plot also suggested the extraction of one factor. The factor matrix indicated that all the items, except item g6 which loaded .103, loaded satisfactory ($.507 < \lambda < .732$) onto one factor. Item g6 was also flagged as problematic during the item reliability analysis. Table 4.45 presents the extracted factor structure. Furthermore, none (0%) of the residual correlations were greater than .05 suggesting that the rotated solution provided a very credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the G-subscale.

Table 4.45

Factor matrix for the G-subscale

	Factor 1
Item g1	.507
Item g2	.732
Item g3	.672
Item g4	.563
Item g5	.694
Item g6	.103

4.4.3.21 Social Desirability (SD) scale uni-dimensionality results

The dimensionality analysis (Table 4.46) for the *Social desirability* scale resulted in the extraction of a single factor. The scree plot also suggested the extraction of one factor. All the obtained factor loadings were greater than .50 and only 3 (20%) of the reproduced correlations were greater than .05. This suggested that the rotated factor solution provided a reasonably credible explanation for the observed correlation matrix. The uni-dimensionality assumption was thus corroborated for the SD scale.

Table 4.46

Factor Matrix for the SD-subscale

	Factor 1
Item q1	.609
Item q2	.599
Item q3	.623
Item q4	.625
Item q5	.687
Item q6	.591

4.5 SUMMARY OF THE DIMENSIONALITY RESULTS

The architecture of the PAPI-N reflects a specific design intention. The design intention is also reflected in the scoring key of the PAPI-N. The design of the PAPI-N questionnaire reflects the intention to construct twenty essentially one-dimensional sets of six items each to reflect variance in each of the twenty latent personality domains collectively comprising the personality construct. The PAPI-N items were designed to function as relatively homogenous stimulus sets to which applicants respond with behaviour that is a relatively uncontaminated expression primarily of a specific underlying latent personality dimension. Specific items were therefore written for each subscale because they are believed to reflect a specific first-order personality dimension. Specific items were written for each subscale because they are believed to express behavioural manifestations of a specific first-order personality dimension. These twenty latent personality dimensions are conceptualised as inseparable dimensions of the personality construct that cannot be divided further. The PAPI-N does not formally make provision for the further subdivision of the personality construct into finer facets of personality.

The results of the dimensionality analyses indicated that the assumption that the twenty latent personality dimensions are inseparable dimensions of the personality construct, was not been supported for eight of the twenty subscales. In all eight of the subscales a obliquely rotated two-factor solution had to be assumed to be an adequate explanation of the observed correlation matrix. In seven of these subscales, the rotated factor solution could be meaningfully theoretically interpreted. In one case (subscale K) the identity of the two underlying factors proved to be somewhat elusive. In seven of the cases the two-factor solution was suggested by the eigenvalue-greater-than-unity rule. In one case (subscale T) the extraction of a second factor was suggested by the large percentage of large residual correlations.

The results of the dimensionality analyses seem to suggest that the majority (90%) of the items generally do systematically reflect their designated latent variables with reasonable success. In the case of the eight subscales that failed the uni-dimensionality assumption, the researcher forced SPSS to extract a single factor. In many of these cases the loadings were generally satisfactory. Differential skewness and kurtosis was examined to determine whether the extracted factor structure could be explained in terms of systematic differences in item statistics. For all the subscales, except for the A-subscale, neither differential skewness nor differential kurtosis was found to offer a plausible account for the extracted factors.

The residuals correlations were examined for both the two-factor and the one-factor solution. Of major importance is that the residuals calculated from the inter-item correlation matrix and the reproduced matrix indicated that the initial solutions, prior to forcing a single factor, provided a substantially more convincing explanation for the observed inter-item correlation matrix than the single-factor solution. This is suggestive that the personality dimensions measured by at least seven of the eight subscales should rather be interpreted as second-order personality factors on which two first-order factors load.

The purpose of the foregoing analyses was to provide insight into the functioning of the PAPI-N scales. Further to this, the analyses assisted in gaining an understanding of the psychometric integrity of the indicator variables that were tasked to represent each of the latent personality variables.

Although no conclusive evidence in this regard can be derived from the current data set, the evidence that emerged from the dimensionality analysis is generally consistent with the position that the subscales do reflect the intended latent variables. Confronting the respective subscale measurement models with the current data set via a series of confirmatory factor analyses utilising LISREL would, however, have shed further light on the credibility of this assumption. To really substantiate such a claim would however require firstly fitting the comprehensive PAPI-N measurement model and (assuming acceptable model fit) subsequently expanding the measurement model implied by the scoring key into a structural model implied by the constitutive definition of personality as seen by the PAPI-N.

The following section will discuss the analyses performed on the data to test the statistical null hypotheses formulated in Chapter 3.

4.6 EVALUATION OF THE PRIMARY (FIRST-ORDER) PAPI-N MEASUREMENT MODEL

4.6.1 UNIVARIATE AND MULTIVARIATE NORMALITY

As items measured on a 7-point scale were used as indicator variables in this study, they may be interpreted as approximating continuous variables (Muthén & Kaplan, 1985). When fitting a measurement model to continuous data, the method of maximum likelihood estimation is used to derive estimates for the measurement model parameters. This method of estimation, however, requires that the data follows a multivariate normal distribution (Kaplan, 2000). This is also true for generalised least squares (GLS) and full information maximum likelihood (FIML) as possible alternative estimation methods for structural equation modelling with continuous data (Mels, 2003). The detailed output of the univariate and multivariate normality is electronically available on the enclosed CD, folder: Univariate and Multivariate Normality, in Appendix C.

In the event of working with non-normal data, alternative estimation methods that could be used for structural equation modelling are robust maximum likelihood (RML), weighted least squares (WLS) and diagonally weighted least squares (DWLS) (Mels; 2003). Robust maximum likelihood is recommended in cases where a measurement model is fitted to continuous data and the assumption of a multivariate normal distribution does not hold (Mels; 2003). The inappropriate analysis of continuous non-normal variables in structural equation models can result in incorrect standard errors and chi-square estimates (Du Toit & Du Toit, 2001; Mels, 2003). The univariate and multivariate normality of the composite indicator variables were consequently evaluated via PRELIS (Jöreskog & Sörbom, 1996b) to determine the appropriate estimation method with which the freed measurement model parameter estimates should be derived.

Table 4.47

Test of multivariate normality for continuous variables before normalisation

Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
961.465	492.933	0.000	19759.974	156.756	0.000	267554.951	0.000

The null hypothesis of univariate normality had to be rejected for all indicator items ($p<.05$). The results presented in Table 4.47 indicated that the null hypothesis of multivariate normality also had to be rejected ($p<.05$). A possible solution was to normalise the PAPI-N indicator variables via PRELIS. The results of the test for multivariate normality for the normalised indicator variable distribution are summarised in Table 4.48.

Table 4.48

Test of multivariate normality for continuous variables after normalisation

Skewness			Kurtosis			Skewness and Kurtosis	
Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Square	P-Value
675.548	300.660	0.000	19038.682	146.817	0.000	111951.951	0.000

Table 4.48 indicates that although the skewness and kurtosis of the item indicator variable distributions significantly improved when normalised, the null hypothesis of multivariate normality still had to be rejected ($p<.05$). As the normalised indicator items still failed the test of multivariate normality, it was consequently decided to use robust maximum likelihood as the estimation method for this study. Subsequently, co-variance matrices and asymptotic co-variance matrices were calculated from the transformed/normalised data set to serve as input for the LISREL analyses.

4.6.2 ASSESSING THE OVERALL GOODNESS-OF-FIT OF THE FIRST-ORDER MEASUREMENT MODEL

LISREL 9.1 was used to perform a confirmatory factor analysis on the first-order PAPI-N measurement model to determine the fit of the model. More specifically the null hypothesis of exact model fit was tested:

H_{01}: RMSEA=0

H_{a1}: RMSEA>0

The exact fit null hypothesis represents the somewhat unrealistic but not altogether impossible position that the first-order measurement model is able to reproduce the observed co-variance matrix to a degree of accuracy that can be explained in terms of sampling error only. Browne and Cudeck (1993, p. 137) consequently argue:

> In applications of the analysis of co-variance structures in the social sciences it is implausible that any model that we use is anything more than an approximation to reality. Since a null hypothesis that a model fits exactly in some population is known a priori to be false, it seems pointless even to try to test whether it is true.

Assuming that the measurement model underlying the PAPI-N only approximates the processes that operated in reality to create the observed co-variance matrix, the following close fit null hypothesis (H_{02}) was also tested (Browne & Cudeck, 1993):

H_{02}: RMSEA≤.05

H_{a2}: RMSEA>.05

Robust maximum likelihood estimation was used to estimate the parameters due to the failure of the data to satisfy the multivariate normality assumption. Due to the number of latent variables and indicator variables in the model, and the number of freed parameters that had to be estimated, the SIMPLIS syntax file was ran in batch mode via DOS. The syntax file ran for approximately 24 hours before the solution successfully converged in 21 iterations. The full spectrum of indices provided by LISREL to assess the absolute and comparative fit of the proposed model was used to reach an informed decision regarding the model's overall fit. The detailed output of the fit of the measurement model is electronically available on the enclosed CD, folder: Measurement Model Fit, in Appendix D.

The fit of the PAPI-N measurement model is discussed next. The magnitude and distribution of the standardised residuals, as well as the percentage of large modification indices calculated for $\mathbf{\Lambda^X}$ and $\mathbf{\Theta_\delta}$, were considered in conjunction with the full array of fit statistics produced by LISREL to assess the fit of the measurement model. If adequate model fit was obtained as judged by this portfolio of evidence, the interpretation of the measurement model parameter estimates were regarded as permissible. Specifically, the magnitude and significance of the lambda factor loading estimates were interpreted, the magnitude and significance of the theta-delta error variance estimates, as well as the proportion of variance explained in the indicator variables by the latent variables they represent.

4.6.2.1 Overall Fit Assessment

The full spectrum of the indices provided by LISREL to assess the absolute and comparative fit of the data is presented in Table 4.49.

Table 4.49

Goodness of fit statistics of the PAPI-N measurement model

Degrees of Freedom	7539	
Minimum Fit Function Chi-Square	68372.199	(P = .0)
Normal Theory Weighted Least Squares Chi-Square	96864.831	(P = .0)
Satorra-Bentler Scaled Chi-Square	82082.195	(P = .0)
Estimated Non-centrality Parameter (NCP)	74543.195	
90 Percent Confidence Interval for NCP	(73622.736 ; 75467.311)	
Minimum Fit Function Value	11.756	
Population Discrepancy Function Value (F0)	12.817	
90 Percent Confidence Interval for F0	(12.659 ; 12.976)	
Root Mean Square Error of Approximation (RMSEA)	.0412	
90 Percent Confidence Interval for RMSEA	(.0410 ; .0415)	
P-Value for Test of Close Fit (RMSEA < .05)	1.000	
Expected Cross-Validation Index (ECVI)	14.272	
90 Percent Confidence Interval for ECVI	(14.114 ; 14.431)	
ECVI for Saturated Model	2.751	
ECVI for Independence Model	278.869	
Chi-Square for Independence Model with 7875 Degrees of Freedom	1621652.947	
Independence AIC	1621904.947	
Model AIC	67004.195	
Saturated AIC	16002.000	
Independence CAIC	1622871.183	
Model BIC	16731.369	
Model CAIC	9192.369	
Saturated CAIC	77357.988	
Normed Fit Index (NFI)	.949	
Non-Normed Fit Index (NNFI)	.952	
Parsimony Normed Fit Index (PNFI)	.909	
Comparative Fit Index (CFI)	.954	
Incremental Fit Index (IFI)	.954	
Relative Fit Index (RFI)	.947	
Critical N (CN)	555.631	
Root Mean Square Residual (RMR)	.116	
Standardised RMR	.0646	
Goodness of Fit Index (GFI)	.791	
Adjusted Goodness of Fit Index (AGFI)	.778	
Parsimony Goodness of Fit Index (PGFI)	.745	

The minimum fit function chi-square (computed as (N -1) F_{min}, where N is the sample size and $Fmin$ is the value of the fitting function at convergence) returned a value of 68372.199 with 7539 degrees of freedom (calculated as ½k(k+1)-t, where k equals the number of observed variables and t equals the number of parameters to be estimated) reflecting a significant result ($p < .01$), implying that the model is not adequate (Kaplan, 2000). The Satorra-Bentler chi-square calculated in terms of the robust maximum likelihood estimation procedure likewise returned a statistically significant value (82082.195; $p<.05$). The hypothesis of exact model fit (H_{01}:RMSEA=0) was therefore rejected. This implies that the PAPI-N measurement model was not able to reproduce the observed co-variance matrix to a degree of accuracy that could be explained in terms of sampling error only. However, due to the chi-square statistic's sensitivity to deviations from multivariate normality and to sample size, as well as the somewhat unrealistic stance that the model fits the population perfectly, Kaplan (2000, p. 84) suggests that "instead of regarding χ^2 as a test statistic, one should regard it as a goodness (or badness) of fit measure in the sense that large χ^2-values correspond to bad fit and small values to good fit".

Diamantopoulos and Siguaw (2000) also recommend assessing the degree of lack of fit of the model, which in this case was done through the estimated non-centrality parameter. Treating the chi-square as a descriptive badness-of-fit measure by expressing the minimum fit function chi-square estimate in terms of the degrees of freedom (χ^2/df = 9.069), suggested that the model does not fit the observed data even at the liberal limit of 5.0 (Theron & Spangenberg, 2004). Kelloway (1998), however, advises that these guidelines have little justification and recommends against a strong reliance on the normed chi-square.

In most circumstances the *a priori* measurement model is only an approximation to reality which means the χ^2 test statistic would follow a non-central χ^2 distribution with non-centrality parameter, λ. The estimated λ value (74543.195) assesses the degree of model fit and reflects the estimated discrepancy between the observed (Σ) and estimated population co-variance ($\Sigma(\theta)$) matrices. Based on the 90 percent confidence interval (73622.736 ; 75467.311) the λ value fell somewhat to the lower limit of the interval. The large value obtained indicated a higher level of discrepancy between (Σ) and ($\Sigma(\theta)$) at a 10% significant level.

The measurement model was fitted by minimising the fit function that compares the observed sample co-variance matrix (S) to the reproduced sample co-variance matrix (S^) derived from the model parameter estimates (Jöreskog & Sörbom, 1993). An indication of the model fit achieved in this case, was depicted by the extent to which the minimum fit function value (11.756) approaches zero. The estimated population function value (F_0) reflects the extent to which the observed population co-variance matrix (Σ) is estimated to differ from the reproduced population co-variance (Σ^) resulting from the parameters minimising the selected discrepancy function fitting the model on Σ (Brown & Cudeck, 1993). An estimate value of 12.817 was obtained for F_0 with confidence interval limits of 12.659 and 12.976. According to Spangenberg and Theron (2004), a perfect fit would have been attained if F_0 was equal to 0 because the observed population co-variance matrix would then have been equal to the estimated population co-variance matrix (Σ^).

The root mean square of approximation (RMSEA) indicates how well the model, with unknown but optimally chosen parameter estimates, would fit the population co-variance matrix (Byrne, 1998). The RMSEA expresses the differences between the observed and estimated sample co-variance matrices in terms of the degrees of freedom of the model (Steiger, 1990), with values below .05 indicating good fit and RMSEA smaller than .08 indicating reasonable fit. Values greater than .08 indicate poor fit. According to Diamantopoulos and Siguaw (2000) the RMSEA statistic is one of the most informative fit indices and is calculated as follows: $(F_0/df)^{1/2}$, where F_0 is the population discrepancy function value and *df* represents the degrees of freedom. As such, a value of zero would indicate the absence of any discrepancy, and would therefore entail a perfect model fit to the data (Mulaik & Millsap, 2000). The RMSEA value of .0412 indicated a good model fit. The 90 percent confidence interval for RMSEA shown in Table 4.49 (.0410 - .0415) indicated that the fit of the model could be regarded as good as the upper bound of the interval fell below the critical cut-off value of .05. In addition, a test of significance of the obtained value was also performed by LISREL by testing H_{02}: RMSEA≤ .05 against H_{a2}: RMSEA> .05. The probability of observing a sample RMSEA value of .0412 under H_{02} was significantly larger (1.00) than the critical p-value of .05. The close fit null hypothesis was therefore not rejected at a 5% significant level (p>.05) and thus it is concluded that the measurement model showed close fit in the parameter (and very good fit in the sample).

The expected cross-validation index (ECVI) focuses on the discrepancy between the reproduced sample co-variance matrix (Σ^) derived from fitting the model on the sample at hand, and the expected co-variance matrix that would be obtained in an independent sample of the same size, from the same population (Byrne, 1998; Diamantopoulos & Siguaw, 2000). It focuses on overall error and is therefore a useful indicator of a model's overall fit (Diamantopoulos & Siguaw, 2000). Since the model ECVI (14.272) was far smaller than the value obtained for the independence model (278.869) but larger than the ECVI value associated with the saturated model (2.751), a model more closely resembling the saturated model seemed to have a better chance of being replicated in a cross-validation sample than the fitted model (Diamantopoulos & Siguaw, 2000).

Evaluating the values of the Akaike information criterion (AIC = 67004.195) presented in Table 4.49 suggested that the fitted measurement model provided a more parsimonious fit than both the independent/null model (1621904.947) and the saturated model (16002.00), as smaller values on these indices indicate a more parsimonious model (Kelloway, 1998; Spangenberg & Theron, 2004). The values for consistent Akaike information criterion (CAIC = 9192.369), suggested that the fitted measurement model provided a more parsimonious fit than the independent/null model (1622871.183) but not the saturated model (77357.988). This, in conjunction with the ECVI results, indicated that the measurement model does lack some influential paths.

A number of incremental fit indices reported by LISREL are given in Table 4.49. These indices include the normed fit index (NFI=.949), the non-normed fit index (NNFI=.952), the comparative fit index (CFI=.954), the incremental fit index (IFI=.954) and the relative fit index (RFI=.947) which can all assume values between 0 and 1, with .90 generally considered as indicative of a well-fitting model. All of the aforementioned indices exceeded the critical value of .90, thus indicating good comparative fit relative to the independence model.

The critical sample size statistic (CN) refers to the size of the sample that would have made the obtained minimum fit function χ^2 statistic just significant at the 5% significant level. The estimated CN value (555.631) fell well above the recommended threshold value of 200 which implies that the model provided an adequate representation of the data (Diamantopoulos & Siguaw, 2000), although this proposed threshold should be used with circumspection.

The root mean square residual (RMR) reflects the square root of the mean squared difference between the sample co-variance matrix and the hypothesised co-variance model. The range of the RMR is calculated based upon the scales of each indicator variable, which makes this index sensitive to the unit of measurement of the model variables and as a result becomes difficult to determine what a low score is (Diamantopoulos & Siguaw, 2000). The standardised RMR (SRMR) resolves this problem and is thus more meaningful to interpret. Values for the SRMR range from 0 to 1 with well-fitting models obtaining values less than .05 (Byrne, 1998; Diamantopoulos & Siguaw, 2000), however values as high as .08 are deemed acceptable (Hu & Bentler, 1999). The RMR (.116) and the SRMR (.065) indicated reasonable fit as values less than .05 on the latter index suggest the model fits the data well (Kelloway, 1998).

The Goodness-of Fit (GFI) statistic was created as an alternative to the chi-square test (Jöreskog & Sorböm, 2003). This statistic serves to calculate the proportion of variance that is accounted for by the estimated population co-variance and determines how closely the model comes to replicating the observed co-variance matrix (Hooper et al., 2008). This statistic ranges from 0 to 1 with larger samples increasing its value. Recommendations for GFI cut-off values are .90 and when factor loadings and samples sizes are low, a cut-off value of .95 is required. Related to the GFI is the adjusted goodness-of-fit statistic (AGFI) which adjusts the GFI based upon degrees of freedom. Like the RMSEA, these indices also favour more parsimonious models but get penalised for model complexity. As with the GFI, indications of good model fit is confirmed by values in the range of 0 and 1 with a generally accepted value of .90. Furthermore, AGFI also tends to increase with sample size (Hooper et al., 2008). Evaluating the fit of the model in terms of these two indices, both GFI (.791) and AGFI (.778) portrayed only moderately satisfactory model fit. However, given the often detrimental effect of sample size on these two fit indices they are not really relied upon as stand-alone indices (Kelloway, 1998).

The parsimonious goodness-of-fit index (PGFI) and the parsimonious normed fit index (PNFI) acknowledges that model fit can be improved by adding paths to the model and estimating more parameters until perfect fit is achieved in the form of a saturated or just-identified model with no degrees of freedom (Kelloway, 1998). The ideal is, therefore, to find the most parsimonious model that achieves satisfactory fit with the least parameters as possible (Jöreskog & Sorböm, 1993). The parsimonious goodness-of-fit index (PGFI=.745) and the parsimonious normed fit index (PNFI=.909) approach model fit from this perspective. Both of these indices have a range from 0 to 1, with higher values indicating a more parsimonious fit, however, neither is likely to reach the .90 cut-off score as used for other indices and there is no recommendation for how high either index should be to indicate parsimonious fit (Kelloway, 1998).

After interpreting the above mentioned fit indices, the conclusion had to be drawn that the PAPI-N measurement model fitted the data reasonably well, but not perfectly. To ensure a thorough assessment of model fit, it was necessary to investigate the standardised residuals and modification indices to further determine the success with which the model explained the observed co-variances amongst the indicator/manifest variables (Jöreskog & Sörbom, 1993).

4.6.2.2 Examination of Residuals

LISREL provides a summary of the largest and smallest standardised co-variance residuals (in contrast to observational residuals) as well as a stem-and-leaf plot (Figure 4.1) that describes how the residuals are distributed around the median residual. Standardised co-variance residuals can be interpreted as z-scores (i.e. in terms of the number of standard deviation the co-variance residual fall above or below the mean). On a 1% significance level standardised residuals can be considered large if they exceed +2.58 or -2.58 (Diamantopoulos & Siguaw, 2000). Standardised residuals should also be distributed approximately symmetrical around the median residual and the median residual should ideally be centred on zero. Large residuals would be indicative of co-variance relationships (or the lack thereof) between indicator variables that the model fails to explain. More specifically, large positive residuals would indicate that the model underestimates the co-variance between two variables and thus would imply the need for additional explanatory paths, whereas, large negative residuals would suggests that the model overestimates the co-variance between two observed variables, implying the need to prune paths away. Figure 4.1 provides the stem-and-leaf plot of standardised co-variance residuals for the PAPI-N measurement model.

```
- 5|2
- 4|
- 3|
- 2|20
- 1|987777666666666555555544444444444433333333333333333222222222222222222222+07
- 0|999999999999999999999999999999999999999999999999999999999999999999999999+98
  0|111111111111111111111111111111111111111111111111111111111111111111111111+98
  1|000000000000000000000000000000000000000000000000000000000000000000000000+91
  2|000000011111112222223333333333444444556666779
  3|
  4|
  5|5
```

Figure 4.1 Stem-and-leaf plot of PAPI-N measurement model standardised residuals

The distribution of standardised residuals appears to be distributed approximately symmetrical around a median of zero. The distribution is, however, slightly negatively skewed with numerous large negative standardised residuals and a smaller number of large positive standardised residuals. Figure 4.1 shows that both the smallest (-52.418) and the largest (54.768) standardised residual fell well outside the 1% significance limits of ± 2.58. The fitted measurement model resulted in 2401 large negative residuals and 2169 large positive residuals. This means that 4570 out of 8001 observed co-variance terms in the observed sample co-variance matrix (57.12%) were poorly estimated by the derived model parameter estimates. The slight preponderance of large negative residuals, moreover, suggested that some of the paths have to be pruned away to improve model fit. These results go against the relatively positive conclusion of model fit that was suggested by the goodness-of-fit statistics in Table 4.49. It moreover goes against the conclusion derived from the ECVI results and the consistent Akaike information criterion results that indicate that the measurement model lacks some influential paths. It should, however, be taken into account that the multivariate normality assumption was not met. Strictly speaking the interpretation of the residual co-variances in terms of the normal distribution is therefore not permissible (Mels, personal communication, 14 August, 2014). Stronger emphasis was therefore placed on the inferences derived from the array of fit statistics reported in Table 4.49.

The Q-plot depicted in Figure 4.2, provides an additional graphical display of residuals by plotting the standardised residuals (horisontal axis) against the quantiles of the normal distribution (Diamantopoulos & Siguaw, 2000). The data points should fall on the 45-degree reference line to indicate good model fit (Jöreskog & Sörbom, 1993). To the extent that the data points swivel away from the 45-degree reference line, model fit becomes less satisfactory. The Q-plot in Figure 4.2 indicates less than perfect model fit as the standardised residuals for most pairs of observed variables show progressively large angular deviations from the 45-degree reference line in both the lower- and upper region of the X-axis. The inference derived from the Q-plot also goes against the relatively positive conclusion of model fit that was suggested by the goodness-of-fit statistics in Table 4.49.

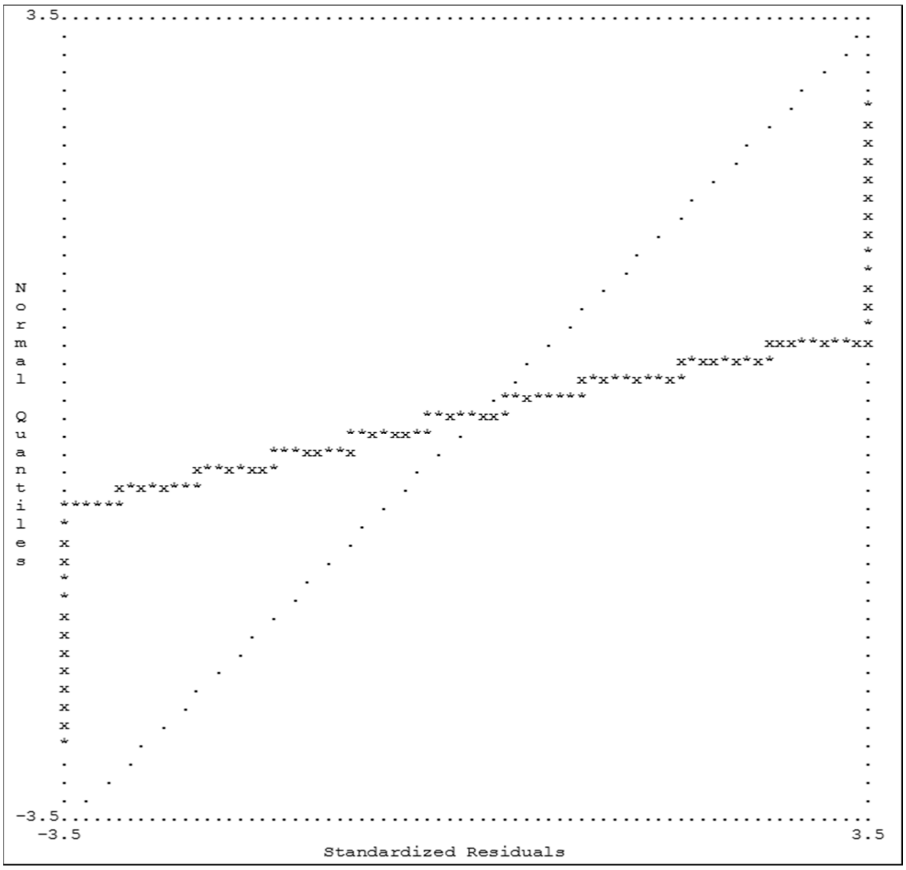

Figure 4.2 Q-plot of PAPI-N measurement model standardised residuals

This finding is in-line with the results reported in Figure 4.1. These results are, however, to some degree in conflict with the model fit conclusion derived from the array of fit statistics portrayed in Table 4.49. The foregoing analysis does not suggest that the fit of the model might be improved through the addition of one or more paths. Rather the preponderance of large negative residuals would suggest a need to remove paths from the model. Subsequently, this possibility was further reviewed and is discussed in the next section.

4.6.2.3 Model modification indices

Model modification indices indicate the extent to which the model's χ^2 fit statistic would decrease if a currently fixed parameter was set free and the model was re-estimated (Jöreskog & Sörbom, 1993). Model modification indices are aimed at answering the question whether any of the currently fixed parameters, when set free in the model, would significantly ($p<.01$) improve the parsimonious fit of the model. Large modification index values (>6.6349) provide an indication as to which parameters, if set free, would improve the fit of the model at a 1% significant level (Diamantopoulos & Siguaw, 2000). This, however, should only be done if the researcher can substantively justify such modifications by presenting a convincing theoretical argument as to why it would make theoretical sense to do so (Diamantopoulos & Siguaw, 2000; Kelloway, 1998). In this study, however, the purpose was not to improve the fit of the measurement model but rather to evaluate the fit of the *a priori* model indicated by the test developers.

Examination of the modification indices calculated for the Λ^X matrix, revealed a number of paths that would significantly improve the fit of the PAPI-N measurement model. Approximately 65.32% (1646 out of 2520 factor loadings) of the currently fixed elements in the Λ^X matrix were identified as statistically significant ($p<.01$). This finding essentially corroborated the conclusion derived from the expected cross-validation index (ECVI) and the consistent Akaike information criterion (CAIC), but contradicted the inference derived from the inspection of the stem-and-leaf plot of the standardised residuals. The large percentage of statistically significant modification indices calculated for Λ^X along with the conclusion derived from the expected cross-validation index (ECVI) and the consistent Akaike information criterion (CAIC) suggested that the current model's claim that each sub-scale of items only reflects a single underlying personality dimension should be questioned. The magnitude of the predicted factor loadings that would be found if currently fixed elements in Λ^X would be freed was also investigated. The completely standardised expected change values calculated by LISREL, however, revealed that none of the loadings exceeded the stringent cut-off level of .50.

Although the large percentage of statistically significant modification indices suggested that substantial improvement in fit might be obtained from modifying the factor loadings from fixed to free, substantive justification could not be found for making any *post hoc* modifications to the measurement model. Specific PAPI-N items were written to function as stimulus sets to which respondents should respond with behaviour which should be a relatively uncontaminated behavioural expression of a specific latent personality dimension. The foregoing results, nonetheless, suggested that some of the items also provide information on latent variables they were not designed to reflect.

Upon inspection of the modification indices calculated for the measurement error variance-co-variance matrix, approximately 15.68% (2470 out of 15750) of the co-variance terms currently fixed to zero, would have significantly improved model fit if they were set free. The magnitude of the completely standardised expected change values, however, do not support making this decision due to the absence of strong correlations between the measurement error terms. This finding therefore commented favourably on the fit of the PAPI-N measurement model.

For this study, a conservative approach of upholding the original design intentions was followed and no alterations were made to the model even though it could significantly improve the fit of the measurement model. The objective of the research is to evaluate the construct validity of the PAPI-N in its current format. The researcher had no mandate to alter the design of the instrument.

4.6.3 EVALUATION OF THE PARAMETER ESTIMATES FOR THE FIRST-ORDER MEASUREMENT MODEL

Diamantopoulos and Siguaw (2000) indicate that when a measure is designed to provide a valid reflection of a specific latent variable, then the slope of the regression of X_j on ξ_i in the fitted measurement model has to be substantial and statistically significant. The unstandardised factor loading matrix ($\mathbf{\Lambda^X}$) (see Table 4.50) contains the slope of the unstandardised individual items X_j on the unstandardised latent personality dimensions ξ_i, and was used to evaluate the statistical significance of the first-order factor loadings hypothesised by the proposed measurement model expressed as equation 1. The results depicted in Table 4.50 signify that all the freed first-order factor loadings were statistically significant ($p<.05$). Significant loadings are indicated by t-values greater than $|1.96|$ in the matrix. All 126 null hypotheses H_{0i}: $\lambda_{jk}=0$; i=3, 4, …, 128; j=1, 2, …, 126; k=1, 2, …, 21 can therefore be rejected in favour of H_{ai}: $\lambda_{jk}\neq 0$; i=3, 4, …, 128; j=1, 2, …, 126; k=1, 2, …, 21. Significant factor loadings provide validity evidence in support of the indicators (Diamantopoulos & Siguaw, 2000).

Table 4.50

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
p1	.661 (.026) 25.634	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p2	.459 (.025) 18.143	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p3	.933 (.021) 43.962	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p4	.873 (.02) 43.728	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p5	.865 (.021) 42.031	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p6	.752 (.019) 39.342	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l1	- -	.679 (.011) 62.512	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l2	- -	.67 (.019) 35.795	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l3	- -	.648 (.01) 67.591	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l4	- -	.692 (.012) 55.46	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
l5	- -	.806 (.018) 44.014	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l6	- -	.643 (.02) 31.404	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c1	- -	- -	.789 (.022) 35.802	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c2	- -	- -	.801 (.013) 62.114	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c3	- -	- -	.615 (.012) 51.93	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c4	- -	- -	1.015 (.013) 77.396	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c5	- -	- -	.788 (.016) 50.455	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c6	- -	- -	.936 (.015) 60.92	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h1	- -	- -	- -	.718 (.02) 36.745	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h2	- -	- -	- -	.725 (.011) 68.839	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
h3	- -	- -	- -	.621 (.011) 56.81	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h4	- -	- -	- -	.683 (.009) 73.861	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h5	- -	- -	- -	.736 (.01) 70.645	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h6	- -	- -	- -	.873 (.011) 81.473	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d1	- -	- -	- -	- -	.582 (.019) 31.215	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d2	- -	- -	- -	- -	.681 (.011) 60.188	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d3	- -	- -	- -	- -	.703 (.013) 56.132	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d4	- -	- -	- -	- -	.598 (.010) 61.135	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d5	- -	- -	- -	- -	.697 (.010) 67.414	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d6	- -	- -	- -	- -	.648 (.010) 67.794	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
w1	- -	- -	- -	- -	- -	.500 (.013) 38.466	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w2	- -	- -	- -	- -	- -	.659 (.014) 48.39	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w3	- -	- -	- -	- -	- -	.76 (.014) 52.994	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w4	- -	- -	- -	- -	- -	.863 (.014) 60.929	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w5	- -	- -	- -	- -	- -	.768 (.015) 52.128	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w6	- -	- -	- -	- -	- -	.694 (.019) 35.736	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r1	- -	- -	- -	- -	- -	- -	.379 (.022) 17.576	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r2	- -	- -	- -	- -	- -	- -	.661 (.019) 34.264	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r3	- -	- -	- -	- -	- -	- -	.651 (.017) 37.726	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
r4	- -	- -	- -	- -	- -	- -	.694 (.012) 59.526	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r5	- -	- -	- -	- -	- -	- -	.685 (.011) 60.298	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r6	- -	- -	- -	- -	- -	- -	.485 (.02) 24.217	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z1	- -	- -	- -	- -	- -	- -	- -	.808 (.014) 55.828	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z2	- -	- -	- -	- -	- -	- -	- -	.665 (.011) 58.675	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z3	- -	- -	- -	- -	- -	- -	- -	.715 (.01) 68.494	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z4	- -	- -	- -	- -	- -	- -	- -	.877 (.019) 46.323	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z5	- -	- -	- -	- -	- -	- -	- -	.682 (.023) 29.321	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z6	- -	- -	- -	- -	- -	- -	- -	.769 (.012) 65.891	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
n1	- -	- -	- -	- -	- -	- -	- -	- -	.576 (.024) 24.022	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n2	- -	- -	- -	- -	- -	- -	- -	- -	.678 (.011) 62.568	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n3	- -	- -	- -	- -	- -	- -	- -	- -	.619 (.01) 61.017	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n4	- -	- -	- -	- -	- -	- -	- -	- -	.757 (.011) 69.302	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n5	- -	- -	- -	- -	- -	- -	- -	- -	.721 (.011) 68.043	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n6	- -	- -	- -	- -	- -	- -	- -	- -	.628 (.009) 70.382	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x1	- -	- -	- -	- -	- -	- -	- -	- -	- -	.796 (.028) 28.533	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x2	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.119 (.021) 52.859	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x3	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.089 (.018) 60.382	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x4	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.191 (.02) 59.073	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
x5	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.324 (.019) 67.951	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x6	- -	- -	- -	- -	- -	- -	- -	- -	- -	.131 (.017) 7.852	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.591 (.024) 24.814	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.686 (.012) 58.405	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.896 (.018) 50.13	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.682 (.012) 54.624	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.903 (.011) 80.658	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.875 (.01) 83.654	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
s1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.25 (.025) 9.999	- -	- -	- -	- -	- -	- -	- -	- -	- -
s2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.579 (.013) 46.012	- -	- -	- -	- -	- -	- -	- -	- -	- -
s3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.745 (.018) 40.758	- -	- -	- -	- -	- -	- -	- -	- -	- -
s4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.667 (.012) 53.758	- -	- -	- -	- -	- -	- -	- -	- -	- -
s5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.884 (.015) 59.143	- -	- -	- -	- -	- -	- -	- -	- -	- -
s6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.773 (.014) 57.097	- -	- -	- -	- -	- -	- -	- -	- -	- -
o1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.95 (.015) 61.681	- -	- -	- -	- -	- -	- -	- -	- -
o2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.956 (.014) 69.821	- -	- -	- -	- -	- -	- -	- -	- -
o3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.197 (.026) 7.443	- -	- -	- -	- -	- -	- -	- -	- -
o4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.804 (.015) 52.109	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
o5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.638 (.015) 42.636	- -	- -	- -	- -	- -	- -	- -	- -
o6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.59 (.023) 25.428	- -	- -	- -	- -	- -	- -	- -	- -
i1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.992 (.017) 59.521	- -	- -	- -	- -	- -	- -	- -
i2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.596 (.027) 22.305	- -	- -	- -	- -	- -	- -	- -
i3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.333 (.026) 12.571	- -	- -	- -	- -	- -	- -	- -
i4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.111 (.016) 70.387	- -	- -	- -	- -	- -	- -	- -
i5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.012 (.016) 61.977	- -	- -	- -	- -	- -	- -	- -
i6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.208 (.015) 82.514	- -	- -	- -	- -	- -	- -	- -
t1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.769 (.023) 34.157	- -	- -	- -	- -	- -	- -
t2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.862 (.018) 48.458	- -	- -	- -	- -	- -	- -
t3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.818 (.013) 64.667	- -	- -	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
t4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.997 (.014) 70.213	- -	- -	- -	- -	- -	- -
t5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.055 (.014) 76.149	- -	- -	- -	- -	- -	- -
t6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.034 (.015) 68.572	- -	- -	- -	- -	- -	- -
k1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.693 (.017) 41.218	- -	- -	- -	- -	- -
k2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.743 (.024) 30.784	- -	- -	- -	- -	- -
k3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.293 (.031) 9.349	- -	- -	- -	- -	- -
k4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.586 (.021) 27.444	- -	- -	- -	- -	- -
k5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.634 (.027) 23.141	- -	- -	- -	- -	- -
k6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.375 (.026) 14.258	- -	- -	- -	- -	- -
e1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.497 (.029) 17.084	- -	- -	- -	- -
e2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.703 (.024) 29.108	- -	- -	- -	- -
e3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.454 (.026) 17.401	- -	- -	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
e4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.077 (.02) 52.717	- -	- -	- -	- -
e5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.014 (.019) 52.74	- -	- -	- -	- -
e6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.857 (.017) 49.71	- -	- -	- -	- -
a1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.375 (.027) 13.665	- -	- -	- -
a2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.547 (.01) 55.709	- -	- -	- -
a3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.462 (.009) 52.11	- -	- -	- -
a4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.731 (.012) 58.908	- -	- -	- -
a5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.087 (.026) 3.361	- -	- -	- -
a6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.721 (.018) 41.168	- -	- -	- -
f1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.124 (.019) 60.037	- -	- -
f2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.041 (.023) 45.119	- -	- -
f3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.612 (.016) 38.865	- -	- -

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
f4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.428 (.018) 24.394	- -	- -
f5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.242 (.021) 59.689	- -	- -
f6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.248 (.019) 12.727	- -	- -
g1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.525 (.01) 51.951	- -
g2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.773 (.011) 68.887	- -
g3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.893 (.014) 62.75	- -
g4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.539 (.009) 57.385	- -
g5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.809 (.012) 64.916	- -
g6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.151 (.027) 5.67	- -
q1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.964 (.02) 48.989
q2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.032 (.021) 48.553
q3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.181 (.024) 48.527

Table 4.50 (continued)

PAPI-N measurement model unstandardised lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
q4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.937 (.019) 48.66
q5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.172 (.019) 61.346
q6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	1.113 (.024) 45.78

* The top value represents the estimate of loading X_j on latent variable ξ_i, i.e λ_{ij}, the second value in brackets the standard error of λ_{ij} and the third value $t=\lambda_{ij}/SE$. λ_{ij} is significant if $t\geq|1,96|$.

Diamantopoulos and Siguaw (2000), however, warn that there is a problem with solely relying on unstandardised loadings and their associated t-values. The problem is that unstandardised loadings retain scaling information of variables and thus can only be interpreted with reference to the scales of the variable in question. Unstandardised factor loadings can therefore not be compared across items. This problem can be avoided by examining the magnitude of the completely standardised solution, in which both latent and manifest variables have been standardised. The completely standardised factor loading values shown in Table 4.51 can be interpreted as the slope of regression of the standardised indicator variables on the standardised latent variables.

Spangenberg and Theron (2005) describe the completely standardised λ parameter estimates as reflecting the average change in standard deviation units in the manifest variable X (i.e., individual items), directly resulting from a one standard deviation change in an exogenous latent variable ξ to which it has been linked, holding the effect of all other variables constant. This study has set the critical requirement that standardised factor loadings should be .50 or higher. Interpreted in this sense, the completely standardised factor loading matrix is not without problems as 35 out of the 126 (27.77%) loadings are below the critical cut-off value of .50. This would suggest that the individual items generally (72.22%)[17] do represent the latent personality dimensions they were designed to reflect acceptably well, but that in a little bit more than a quarter of the items, less than 25% of the variance in the item was due to the latent variable it was designed to reflect. The latter inference agrees with the conclusion derived from the dimensionality and item analyses reported earlier. A finding of somewhat lower factor loadings is, on the other hand, to be expected given the broad nature of the personality dimensions and the fact that the responses to items are determined by the whole personality (Moyo, 2009).

The square of the completely standardised factor loadings indicate the proportion of indicator variance explained in terms of the latent variable it is meant to express (Diamantopoulos & Siguaw, 2000). Since each indicator only loads on a single latent variable the squared completely standardised loadings equal the R^2 values shown below in Table 4.52. The proportion of the variance in the observed variable that is explained by the latent variable

[17] This percentage compares very favorably with the results obtained in the earlier exploratory factor analyses performed on the subscales.

linked to it in the measurement model is indicated by the squared multiple correlations (R^2) for the observed indicator variables as shown in Table 4.52.

Table 4.51

PAPI-N measurement model completely standardised solution lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
p1	.389	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p2	.283	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p3	.599	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p4	.608	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p5	.587	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
p6	.558	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l1	- -	.741	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l2	- -	.487	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l3	- -	.742	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l4	- -	.690	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l5	- -	.546	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
l6	- -	.433	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c1	- -	- -	.449	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c2	- -	- -	.715	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c3	- -	- -	.646	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c4	- -	- -	.798	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c5	- -	- -	.635	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
c6	- -	- -	.710	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h1	- -	- -	- -	.472	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h2	- -	- -	- -	.759	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h3	- -	- -	- -	.693	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h4	- -	- -	- -	.780	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h5	- -	- -	- -	.784	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
h6	- -	- -	- -	.830	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d1	- -	- -	- -	- -	.392	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d2	- -	- -	- -	- -	.695	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d3	- -	- -	- -	- -	.683	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d4	- -	- -	- -	- -	.716	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d5	- -	- -	- -	- -	.754	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
d6	- -	- -	- -	- -	.752	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w1	- -	- -	- -	- -	- -	.503	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w2	- -	- -	- -	- -	- -	.622	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w3	- -	- -	- -	- -	- -	.642	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w4	- -	- -	- -	- -	- -	.736	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w5	- -	- -	- -	- -	- -	.658	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
w6	- -	- -	- -	- -	- -	.477	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r1	- -	- -	- -	- -	- -	- -	.252	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r2	- -	- -	- -	- -	- -	- -	.481	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r3	- -	- -	- -	- -	- -	- -	.515	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r4	- -	- -	- -	- -	- -	- -	.700	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r5	- -	- -	- -	- -	- -	- -	.725	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
r6	- -	- -	- -	- -	- -	- -	.359	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z1	- -	- -	- -	- -	- -	- -	- -	.648	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.51 (continued)

PAPI-N measurement model completely standardised solution lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
z2	- -	- -	- -	- -	- -	- -	- -	.671	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z3	- -	- -	- -	- -	- -	- -	- -	.754	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z4	- -	- -	- -	- -	- -	- -	- -	.589	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z5	- -	- -	- -	- -	- -	- -	- -	.413	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
z6	- -	- -	- -	- -	- -	- -	- -	.759	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n1	- -	- -	- -	- -	- -	- -	- -	- -	.325	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n2	- -	- -	- -	- -	- -	- -	- -	- -	.691	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n3	- -	- -	- -	- -	- -	- -	- -	- -	.705	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n4	- -	- -	- -	- -	- -	- -	- -	- -	.749	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n5	- -	- -	- -	- -	- -	- -	- -	- -	.749	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
n6	- -	- -	- -	- -	- -	- -	- -	- -	.754	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x1	- -	- -	- -	- -	- -	- -	- -	- -	- -	.431	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x2	- -	- -	- -	- -	- -	- -	- -	- -	- -	.671	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x3	- -	- -	- -	- -	- -	- -	- -	- -	- -	.722	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x4	- -	- -	- -	- -	- -	- -	- -	- -	- -	.726	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x5	- -	- -	- -	- -	- -	- -	- -	- -	- -	.779	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
x6	- -	- -	- -	- -	- -	- -	- -	- -	- -	.118	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.341	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.661	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.608	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.650	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.839	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
b6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.860	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
s1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.165	- -	- -	- -	- -	- -	- -	- -	- -	- -
s2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- .598	- -	- -	- -	- -	- -	- -	- -	- -	- -

Table 4.51 (continued)

PAPI-N measurement model completely standardised solution lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
s3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.548	- -	- -	- -	- -	- -	- -	- -	- -	- -
s4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.639	- -	- -	- -	- -	- -	- -	- -	- -	- -
s5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.693	- -	- -	- -	- -	- -	- -	- -	- -	- -
s6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.694	- -	- -	- -	- -	- -	- -	- -	- -	- -
01	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.718	- -	- -	- -	- -	- -	- -	- -	- -
02	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.772	- -	- -	- -	- -	- -	- -	- -	- -
03	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.118	- -	- -	- -	- -	- -	- -	- -	- -
04	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.682	- -	- -	- -	- -	- -	- -	- -	- -
05	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.584	- -	- -	- -	- -	- -	- -	- -	- -
06	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.378	- -	- -	- -	- -	- -	- -	- -	- -
i1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.716	- -	- -	- -	- -	- -	- -	- -
i2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.355	- -	- -	- -	- -	- -	- -	- -
i3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.211	- -	- -	- -	- -	- -	- -	- -
i4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.802	- -	- -	- -	- -	- -	- -	- -
i5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.771	- -	- -	- -	- -	- -	- -	- -
i6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.869	- -	- -	- -	- -	- -	- -	- -
t1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.460	- -	- -	- -	- -	- -	- -
t2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.626	- -	- -	- -	- -	- -	- -
t3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.752	- -	- -	- -	- -	- -	- -
t4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.801	- -	- -	- -	- -	- -	- -
t5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.826	- -	- -	- -	- -	- -	- -
t6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.783	- -	- -	- -	- -	- -	- -
k1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.591	- -	- -	- -	- -	- -
k2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.460	- -	- -	- -	- -	- -
k3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.152	- -	- -	- -	- -	- -
k4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.414	- -	- -	- -	- -	- -
k5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.357	- -	- -	- -	- -	- -
k6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.230	- -	- -	- -	- -	- -

Table 4.51 (continued)

PAPI-N measurement model completely standardised solution lambda-X matrix

	P	L	C	H	D	W	R	Z	N	X	B	S	O	I	T	K	E	A	F	G	SD
e1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.269	- -	- -	- -	- -
e2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.454	- -	- -	- -	- -
e3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.288	- -	- -	- -	- -
e4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.694	- -	- -	- -	- -
e5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.708	- -	- -	- -	- -
e6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.665	- -	- -	- -	- -
a1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.198	- -	- -	- -
a2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.649	- -	- -	- -
a3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.617	- -	- -	- -
a4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.686	- -	- -	- -
a5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.050	- -	- -	- -
a6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.531	- -	- -	- -
f1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.732	- -	- -
f2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.618	- -	- -
f3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.548	- -	- -
f4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.366	- -	- -
f5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.745	- -	- -
f6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.185	- -	- -
g1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.632	- -
g2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.762	- -
g3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.718	- -
g4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.665	- -
g5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.738	- -
g6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.086	- -
q1	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.621
q2	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.624
q3	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.629
q4	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.645
q5	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.695
q6	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	.611

A high R^2 value would be preferred as it would indicate that the variance for the concerned indictor, to a large degree, reflects variance in the latent variable to which it has been linked. The rest of the variance, not explained by the latent variable can then be attributed to systematic and random measurement error (Diamantopoulos & Siguaw, 2000).

Table 4.52

PAPI-N measurement model squared multiple correlations for X-variables

p1	**l1**	**c1**	**h1**	**d1**	**w1**	**r1**	**z1**	**n1**	**x1**	**b1**	**s1**	**o1**	**i1**	**t1**	**k1**	**e1**	**a1**
.151	.549	.201	.223	.153	.253	.064	.420	.106	.185	.116	.027	.516	.513	.212	.350	.072	.039
p2	**l2**	**c2**	**h2**	**d2**	**w2**	**r2**	**z2**	**n2**	**x2**	**b2**	**s2**	**o2**	**i2**	**t2**	**k2**	**e2**	**a2**
.080	.237	.512	.577	.484	.387	.231	.450	.477	.450	.437	.357	.597	.126	.392	.212	.206	.421
p3	**l3**	**c3**	**h3**	**d3**	**w3**	**r3**	**z3**	**n3**	**x3**	**b3**	**s3**	**o3**	**i3**	**t3**	**k3**	**e3**	**a3**
.358	.551	.417	.481	.466	.413	.265	.569	.497	.522	.370	.300	.014	.045	.565	.023	.083	.380
p4	**l4**	**c4**	**h4**	**d4**	**w4**	**r4**	**z4**	**n4**	**x4**	**b4**	**s4**	**o4**	**i4**	**t4**	**k4**	**e4**	**a4**
.370	.476	.637	.609	.512	.542	.490	.347	.561	.528	.423	.409	.465	.643	.641	.172	.481	.471
p5	**l5**	**c5**	**h5**	**d5**	**w5**	**r5**	**z5**	**n5**	**x5**	**b5**	**s5**	**o5**	**i5**	**t5**	**k5**	**e5**	**a5**
.345	.298	.404	.614	.568	.432	.526	.171	.560	.607	.703	.480	.342	.594	.683	.128	.502	.003
p6	**l6**	**c6**	**h6**	**d6**	**w6**	**r6**	**z6**	**n6**	**x6**	**b6**	**s6**	**o6**	**i6**	**t6**	**k6**	**e6**	**a6**
.312	.188	.504	.689	.565	.228	.129	.576	.568	.014	.740	.482	.143	.755	.613	.053	.442	.282

f1	**g1**	**q1**
.536	.399	.386
f2	**g2**	**q2**
.382	.580	.389
f3	**g3**	**q3**
.301	.516	.396
f4	**g4**	**q4**
.134	.442	.416
f5	**g5**	**q5**
.555	.544	.483
f6	**g6**	**q6**
.034	.007	.373

The total variance in the i^{th} item (X_i), is the result of variance in the latent variable the individual item was designed to reflect (ξ_i), and/or variance due to variance in other systematic latent effects the individual item was not designed to reflect and/or lastly, variance due to random measurement error (Spangenberg & Theron, 2005). The latter two sources of variance in the indicator variable (in other words the individual item) are acknowledged in equation 1 through the measurement error term (δ_i). The measurement error terms (δ_i) thus do not differentiate between systematic and random sources of error or non-relevant variance. The square of the completely standardised factor loadings λ_{ij} (see Table 4.52) could be interpreted as the proportion systematic-relevant item variance given that each item load on one latent variable only. The unstandardised measurement error variances for individual items are reflected in Table 4.53 and the completely standardised theta-delta (Θ_δ) matrix that reflects the proportion of non-relevant variance in the items is shown in Table 4.54.

Table 4.53

PAPI-N measurement model unstandardised error variances

p1	l1	c1	h1	d1	w1	r1	z1	n1	x1	b1	s1	o1	i1	t1
2.454	.378	2.47	1.8	1.869	.738	2.108	.901	2.811	2.786	2.662	2.242	.847	.933	2.204
(.051)	(.012)	(.06)	(.042)	(.043)	(.017)	(.039)	(.023)	(.053)	(.06)	(.054)	(.042)	(.031)	(.027)	(.047)
48.521	31.651	40.839	42.395	43.854	42.453	53.56	38.837	53.036	46.791	49.471	53.174	27.542	35.059	47.238
p2	**l2**	**c2**	**h2**	**d2**	**w2**	**r2**	**z2**	**n2**	**x2**	**b2**	**s2**	**o2**	**i2**	**t2**
2.422	1.448	.612	.386	.495	.687	1.452	.541	.503	1.526	.605	.603	.617	2.471	1.153
(.045)	(.035)	(.016)	(.013)	(.014)	(.017)	(.034)	(.016)	(.014)	(.047)	(.019)	(.015)	(.021)	(.046)	(.031)
53.901	41.47	37.358	29.344	36.437	39.488	42.418	34.379	35.88	32.797	31.941	39.245	29.455	54.21	37.279
p3	**l3**	**c3**	**h3**	**d3**	**w3**	**r3**	**z3**	**n3**	**x3**	**b3**	**s3**	**o3**	**i3**	**t3**
1.557	.343	.528	.417	.565	.822	1.178	.388	.388	1.086	1.366	1.293	2.735	2.376	.515
(.04)	(.01)	-(.014)	(.012)	(.015)	(.022)	(.027)	(.014)	(.011)	(.032)	(.036)	(.032)	(.045)	(.043)	(.014)
38.913	35.009	38.812	35.46	37.439	37.149	43.769	28.506	35.928	34.464	38.116	40.763	60.123	55.55	36.046
p4	**l4**	**c4**	**h4**	**d4**	**w4**	**r4**	**z4**	**n4**	**x4**	**b4**	**s4**	**o4**	**i4**	**t4**
1.297	.526	.587	.3	.34	.63	.501	1.446	.449	1.271	.635	.643	.743	.684	.557
(.036)	-(.016)	(.019)	(.008)	(.009)	(.021)	(.014)	(.035)	(.013)	(.04)	(.018)	(.017)	(.022)	(.024)	(.02)
36.104	32.493	31.093	35.308	38.075	29.622	35.685	41.701	34.021	32.043	34.595	38.85	33.896	29.067	28.351
p5	**l5**	**c5**	**h5**	**d5**	**w5**	**r5**	**z5**	**n5**	**x5**	**b5**	**s5**	**o5**	**i5**	**t5**
1.424	1.531	.918	.34	.369	.774	.424	2.253	.408	1.134	.344	.846	.785	.699	.518
(.039)	(.04)	-(.025)	(.011)	(.014)	(.023)	(.015)	(.047)	(.013)	(.044)	(.016)	(.025)	(.019)	(.025)	(.018)
36.973	37.848	37.449	32.143	26.005	33.366	28.433	47.54	32.474	25.962	21.478	34.111	40.352	28.042	28.659
p6	**l6**	**c6**	**h6**	**d6**	**w6**	**r6**	**z6**	**n6**	**x6**	**b6**	**s6**	**o6**	**i6**	**t6**
1.248	1.79	.861	.344	.324	1.633	1.595	.436	0.3	1.213	.269	.642	2.088	.474	.675
(.031)	-(.039)	-(.026)	(.011)	(.01)	(.042)	(.033)	(.016)	(.009)	(.021)	(.014)	(.018)	(.042)	(.02)	(.021)
40.632	45.391	33.054	30.487	32.479	38.468	48.204	27.661	32.77	58.811	18.874	35.245	49.713	24.306	31.771

k1	e1	a1	f1	g1	q1
.894	3.17	3.46	1.093	.415	1.479
(.023)	(.055)	(.056)	(.034)	(.013)	(.039)
38.07	57.351	61.402	32.083	33.005	38.393
k2	**e2**	**a2**	**f2**	**g2**	**q2**
2.054	1.907	.412	1.75	.432	1.673
(.045)	(.043)	(.012)	(.046)	(.016)	(.042)
45.685	44.179	34.253	38.158	27.347	39.794
k3	**e3**	**a3**	**f3**	**g3**	**q3**
3.605	2.28	.348	.87	.749	2.13
(.059)	(.046)	(.009)	(.019)	(.025)	(.054)
61.183	49.594	36.801	46.898	29.861	39.545
k4	**e4**	**a4**	**f4**	**g4**	**q4**
1.659	1.251	0.6	1.182	.368	1.235
(.036)	(.041)	(.017)	(.023)	(.011)	(.035)
46.057	30.448	35.085	52.248	34.52	35.297
k5	**e5**	**a5**	**f5**	**g5**	**q5**
2.747	1.021	3.013	1.235	.549	1.471
(.049)	(.035)	(.048)	(.043)	(.019)	(.042)
55.965	29.082	62.855	28.528	28.917	35.249
k6	**e6**	**a6**	**f6**	**g6**	**q6**
2.504	.925	1.321	1.741	3.034	2.081
(.047)	(.027)	(.031)	(.03)	(.05)	(.054)
53.705	34.69	42.045	57.467	60.765	38.204

* The top value represents the unstandardised $\theta_{\delta j}$ estimate, the second value in brackets the standard error of $\theta_{\delta j}$ and the third value the test statistic z.

All the unstandardised measurement error variances shown in Table 4.53 are statistically significant ($p<.05$). H_{0i}: $\Theta_{\delta jj}=0$; i =129, 130,..., 254; j=1, 2.....126 was therefore rejected in favour of H_{ai}: $\Theta_{\delta jj} > 0$; i =129, 130,..., 254; j=1, 2.....126 for all i.

The completely standardised error variance of the i^{th} item ($\theta_{\delta ii}$) in Table 4.54 thus consists of systematic non-relevant variance and random error variance. The values given in Table 4.54 could be interpreted as indicator variable validity coefficients, $\rho(Xi,\xi j)$. Spangenberg and Theron (2005) therefore conclude that, since ($\lambda^2_{ij} + \theta_{\delta ii}$) are equal to unity in the completely standardised solution, the validity coefficients can be defined as follows:

$$\rho(Xi,\xi j) = \sigma^2_{systematic\text{-}relevant}/(\sigma^2_{systematic\text{-}relevant} + \sigma^2_{non\text{-}relevant})$$

$$= \lambda^2_{ij}/[\lambda^2_{ij} + \theta_{\delta ii}]$$

$$= 1- (\theta_{\delta ii}/[\lambda^2_{ij} + \theta_{\delta ii}])$$

$$= 1- \theta_{\delta ii}$$

$$= \lambda^2_{ij}$$

Table 4.54

PAPI-N measurement model completely standardised error variances

p1	**l1**	**c1**	**h1**	**d1**	**w1**	**r1**	**z1**	**n1**	**x1**	**b1**	**s1**	**o1**	**i1**	**t1**	**k1**	**e1**
.849	.451	.799	.777	.847	.747	.936	.580	.894	.815	.884	.973	.484	.487	.788	.650	.928
p2	**l2**	**c2**	**h2**	**d2**	**w2**	**r2**	**z2**	**n2**	**x2**	**b2**	**s2**	**o2**	**i2**	**t2**	**k2**	**e2**
.920	.763	.488	.423	.516	.613	.769	.550	.523	.550	.563	.643	.403	.874	.608	.788	.794
p3	**l3**	**c3**	**h3**	**d3**	**w3**	**r3**	**z3**	**n3**	**x3**	**b3**	**s3**	**o3**	**i3**	**t3**	**k3**	**e3**
.642	.449	.583	.519	.534	.587	.735	.431	.503	.478	.630	.700	.986	.955	.435	.977	.917
p4	**l4**	**c4**	**h4**	**d4**	**w4**	**r4**	**z4**	**n4**	**x4**	**b4**	**s4**	**o4**	**i4**	**t4**	**k4**	**e4**
.630	.524	.363	.391	.488	.458	.510	.653	.439	.472	.577	.591	.535	.357	.359	.828	.519
p5	**l5**	**c5**	**h5**	**d5**	**w5**	**r5**	**z5**	**n5**	**x5**	**b5**	**s5**	**o5**	**i5**	**t5**	**k5**	**e5**
.655	.702	.596	.386	.432	.568	.474	.829	.440	.393	.297	.520	.658	.406	.317	.872	.498
p6	**l6**	**c6**	**h6**	**d6**	**w6**	**r6**	**z6**	**n6**	**x6**	**b6**	**s6**	**o6**	**i6**	**t6**	**k6**	**e6**
.688	.812	.496	.311	.435	.772	.871	.424	.432	.986	.260	.518	.857	.245	.387	.947	.558

a1	**f1**	**g1**	**q1**
.961	.464	.601	.614
a2	**f2**	**g2**	**q2**
.579	.618	.420	.611
a3	**f3**	**g3**	**q3**
.620	.699	.484	.604
a4	**f4**	**g4**	**q4**
.529	.866	.558	.584
a5	**f5**	**g5**	**q5**
.997	.445	.456	.517
a6	**f6**	**g6**	**q6**
.718	.966	.993	.627

Since reliability can be defined as the extent to which the variance in items can be attributed to systematic sources, irrespective of whether the source of variance is relevant to the measurement intention or not, the values shown in Table 4.54 could, when squared, simultaneously be interpreted as lower bound estimates of the item reliabilities ρ_{ii}. The extent to which the true item reliabilities would be under-estimated would be determined by the extent to which δ_{ii} contains the effect of the systematic non-relevant latent influences.

In terms of the foregoing argument, the values of the squared multiple correlations for the items presented in Table 4.52, as well as the reciprocal values shown in Table 4.54, cause some concern as there are quite a substantial number of items (35 or 27.78%) that failed to adequately reflect variance in the latent variables they were meant to reflect, given the critical percentage set at the outset of the study ($R^2 \geq .25$). Table 4.52 and Table 4.54 further reveal that in the case of 34 out of 126 items (26.98%) the proportion of variance in the item that is explained by the latent personality dimension it has been designed to reflect in terms of the measurement model, exceeds the critical cut-off percentage set in this study (25%). This tends to erode confidence with which a definite, unqualified positive conclusion on the merits of the PAPI-N model will be possible.

The 20 latent variables constituting the personality domain as defined by PAPI-N are expected to correlate to some degree. Since the 20 latent personality subscales are qualitatively distinct, although related constructs of personality, they should not correlate very high with each other. The phi-matrix of correlations between the 20 latent personality dimensions is depicted in Table 4.55.

Ten (10) of the 210 inter-latent variable correlations were not statistically significant ($p<.05$). H_{0i}: $\phi_{jk}=0$; i =255, 257,..., 464; j=1, 2.....21; k=1, 2.....21 was therefore rejected in favour of H_{ai}: $\phi_{jk} > 0$; i =255, 257,..., 464; j=1, 2.....21; k=1, 2.....21 was therefore not rejected for $\emptyset_{10,3}$; $\emptyset_{10,4}$; $\emptyset_{10,8}$; $\emptyset_{10,9}$; $\emptyset_{11,10}$; $\emptyset_{17,10}$; $\emptyset_{20,10}$; $\emptyset_{21,1}$; $\emptyset_{21,10}$ and $\emptyset_{21,16}$. It is worthy of note that dimension 10 was involved in eight of the ten insignificant correlations. Dimension 10 is the *Need to be noticed* (X) scale.

Furthermore, inter-correlations can be considered very high if they exceed a value of .90 (Myburgh, 2013). Only one correlation ($\emptyset_{95}$, the correlation between *Attention to detail* and *Need to finish a task*) out of the 210 latent variable inter-correlations exceeded .80, but still fell below the critical cut-off value of .90. As the off-diagonal entries tend to contain relatively low to moderate correlations, the results tend to support the discriminant validity of the 20 first-order PAPI-N factors.

4.6.4 DISCRIMINANT VALIDITY

The absence of extremely high inter-correlations between latent variables in the phi matrix is, however, not enough evidence to confidently conclude discriminant validity (Myburgh, 2013). Discriminant validity can be examined by determining whether the correlations contained in the phi matrix are significantly different from one (Bagozzi et al., 1991). This can be achieved by calculating a 95% confidence interval for each phi estimate. An Excel macro developed by Scientific Software International (Mels, 2010) was used to calculate the 95% confidence interval for each sample estimate in **Φ**. Discriminant validity is achieved when the correlations between the factors are significantly less than 1.0 ($p<.05$). Discriminant validity will therefore exist if the 95% confidence intervals calculated for the 210 phi estimates do not contain the value 1. Should any confidence interval contain the value 1 it would imply that the null hypothesis H_{0i}: $p=1$ cannot be rejected. If discriminant validity has not been shown, confidence in the claims that the PAPI-N succeeded in measuring the latent variables comprising the personality construct as qualitatively distinct constructs would then be seriously compromised.

The 95% confidence intervals for the 210 latent variable inter-correlations in Table 4.56 indicate that none of the intervals included unity. The discriminant validity of the PAPI-N dimension measures was thereby confirmed.

The 20 latent personality dimensions correlated to some degree as depicted in Table 4.55. None of the correlations were considered extremely high. More specifically, it can be concluded with 95% confidence that none of the inter-latent correlations in the parameter were equal to the value of 1. Each of the latent personality dimensions are represented by six item statements. The average variance extracted (AVE) reflects the average proportion of variance in the indicator variables that is accounted for by the latent variable that the indicator variables were designed to represent (Diamantopoulos & Sigauw, 2000). According to Farrell (2010), the average variance extracted should be greater than .50 and should be greater than the squared correlation between the latent variables. The rationale offered for these two criteria is that, firstly, the latent variable should account for more variance in the indicator variables that were designed to reflect it than measurement variance. Farrell (2010), secondly, argues that the latent variables should account for more variance in the indicator variables that represent them than what they account for in each other.

Table 4.57 represents the squared correlations between the latent personality dimensions as well as the average variance extracted for each latent personality dimension. One hundred and sixty-seven of the squared inter-latent variable correlations (79.5%) were smaller than both the AVE values associated with the latent variable pairs being correlated. The latent variables involved in these 167 pairs therefore account for more variance in their indicator variables that were designed to reflect them, than they accounted for in each other. Twenty-two of the squared inter-latent variable correlations (10.5%) were smaller than one of the AVE values associated with one of the latent variables in the pair of variables being correlated, whereas twenty-one of the squared inter-latent variable correlations (10%) were larger than both the AVE values associated with the latent variable pairs being correlated. The AVE criterion provides psychological tests with an extremely stringent challenge that very few instruments that measure comprehensive multi-dimensional constructs comprising a sizeable number of latent dimensions manage to satisfy completely. Although the PAPI-N did demonstrate some difficulty in convincingly discriminating between some of the latent personality dimensions it nonetheless succeeded in doing so on the majority of the latent personality dimensions.

Table 4.55

Phi (**Φ**) *matrix*

	Control	Leader	Organise	Planner	Detail	Rules	Thinker	Change	Finish	Notice	Belong	Social	Relate	Decision	Workpace	Forceful	Emotion
Control	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Leader	.618 (.014) 44.137	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Organise	.197 (.018) 10.889	.518 (.015) 35.706	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Planner	.327 (.017) 19.157	.663 (.012) 56.226	.753 (.009) 80.549	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Detail	.247 (.018) 13.669	.540 (.014) 37.505	.727 (.011) 68.167	.791 (.009) 87.636	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Rules	.138 (.019) 7.302	.337 (.016) 21.052	.460 (.015) 31.660	.522 (.014) 38.482	.567 (.013) 42.804	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Thinker	.474 (.017) 27.395	.731 (.012) 62.819	.500 (.015) 32.639	.696 (.012) 56.386	.682 (.014) 50.461	.340 (.017) 19.569	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -	- -
Change	.261 (.017) 14.980	.582 (.013) 44.050	.438 (.015) 29.820	.574 (.013) 45.336	.558 (.014) 41.022	.378 (.016) 23.883	.719 (.011) 62.808	1.000	- -	- -	- -	- -	- -	- -	- -	- -	- -
Finish	.274 (.017) 15.843	.593 (.013) 44.939	.783 (.009) 82.505	.783 (.009) 89.468	.841 (.008) 103.491	.546 (.013) 41.658	.619 (.013) 46.798	.574 (.013) 43.285	1.000	- -	- -	- -	- -	- -	- -	- -	- -
Notice	.577 (.015) 39.362	.193 (.017) 11.539	.014 (.017) .834	.015 (.017) .871	-.035 (.017) -2.041	-.036 (.017) -2.064	.130 (.018) 7.189	.000 (.017) -.029	-.013 (.016) -.820	1.000	- -	- -	- -	- -	- -	- -	- -
Belong	.191 (.017) 11.058	.463 (.014) 32.804	.363 (.015) 24.752	.470 (.014) 34.537	.438 (.014) 31.019	.593 (.013) 44.762	.418 (.016) 26.598	.541 (.014) 39.729	.470 (.014) 33.605	-.004 (.017) -.267	1.000	- -	- -	- -	- -	- -	- -
Social	-.252 (.018) -13.633	-.563 (.014) 39.357	-.469 (.015) 31.295	- .549 (.014) -40.054	- .532 (.014) -37.329	-.454 (.015) -29.439	-.549 (.016) -34.646	-.480 (.014) -33.714	-.545 (.014) -38.919	-.107 (.018) -6.054	-.560 (.013) -42.788	1.000	- -	- -	- -	- -	- -

Table 4.55 (continued)
Phi (Φ) matrix

	Control	Leader	Organise	Planner	Detail	Rules	Thinker	Change	Finish	Notice	Belong	Social	Relate	Decision	Workpace	Forceful	Emotion
Relate	.163 (.019) 8.693	.179 (.017) 10.637	.185 (.016) 11.247	.184 (.016) 11.367	.191 (.017) 11.480	.429 (.016) 27.479	.161 (.018) 8.963	.113 (.017) 6.725	.222 (.016) 13.847	.127 (.017) 7.418	.367 (.015) 23.976	-.440 (.016) -27.690	1.000	- -	- -	- -	- -
Decision	.430 (.017) 25.464	.487 (.015) 33.574	.294 (.016) 18.173	.322 (.016) 20.001	.285 (.017) 17.175	.113 (.017) 6.498	.501 (.016) 31.651	.329 (.015) 21.406	.329 (.015) 21.448	.218 (.017) 12.729	.141 (.016) 8.749	-.357 (.017) -21.321	.117 (.017) 6.830	1.000	- -	- -	- -
Workpace	.334 (.017) 19.326	.507 (.014) 36.168	.525 (.013) 39.898	.487 (.014) 35.434	.430 (.015) 28.596	.224 (.016) 13.560	.525 (.015) 33.970	.385 (.015) 25.916	.513 (.013) 38.853	.170 (.017) 10.073	.194 (.016) 12.414	-.412 (.016) -25.927	.147 (.017) 8.809	.715 (.011) 64.586	1.000	- -	- -
Forceful	.610 (.019) 32.143	.640 (.017) 38.537	.307 (.020) 15.214	.434 (.018) 23.851	.413 (.019) 22.028	.272 (.020) 13.558	.631 (.017) 36.086	.572 (.018) 32.168	.397 (.018) 21.765	.331 (.020) 16.423	.313 (.019) 16.412	-.311 (.021) -14.898	.118 (.021) 5.600	.436 (.019) 23.251	.369 (.019) 18.930	1.000	- -
Emotion	.270 (.019) 14.086	.322 (.017) 18.965	.360 (.016) 21.959	.401 (.016) 25.782	.388 (.016) 24.260	.235 (.018) 13.377	.354 (.018) 19.654	.308 (.017) 18.098	.366 (.016) 23.016	.011 (.019) .571	.260 (.017) 15.532	-.479 (.016) -29.531	.110 (.018) 6.029	.190 (.018) 10.554	.227 (.017) 13.096	-.049 (.022) -2.226	1.000
Achieve	.515 (.016) 32.814	.668 (.013) 52.614	.619 (.013) 46.705	.666 (.012) 55.965	.626 (.013) 48.191	.497 (.015) 32.887	.631 (.014) 45.535	.643 (.013) 48.904	.703 (.012) 59.272	.225 (.017) 13.617	.494 (.015) 33.340	-.501 (.015) -32.730	.289 (.017) 17.103	.330 (.016) 20.809	.458 (.015) 31.531	.528 (.019) 28.365	.328 (.017) 19.019
Support	.294 (.018) 15.953	.175 (.017) 10.089	.267 (.016) 16.342	.182 (.016) 11.139	.245 (.017) 14.869	.385 (.016) 23.847	.154 (.018) 8.410	.101 (.017) 5.935	.267 (.016) 16.790	.338 (.017) 20.038	.231 (.016) 14.424	-.350 (.016) -21.321	.467 (.015) 30.917	.172 (.017) 9.949	.269 (.017) 16.197	.095 (.021) 4.423	.201 (.018) 11.033
Work	.255 (.017) 14.768	.533 (.014) 38.516	.605 (.012) 50.130	.612 (.012) 52.661	.591 (.012) 48.436	.438 (.015) 30.119	.495 (.015) 33.324	.475 (.014) 33.378	.677 (.011) 60.137	.011 (.016) .677	.378 (.015) 25.243	-.469 (.015) -31.444	.239 (.016) 14.785	.299 (.015) 19.457	.502 (.013) 37.817	.329 (.020) 16.873	.321 (.016) 19.824
SD	-.011 (.019) -.596	.061 (.017) 3.651	.266 (.015) 17.171	.231 (.015) 15.228	.223 (.016) 14.040	.107 (.017) 6.378	.173 (.017) 10.064	.172 (.016) 10.608	.247 (.015) 16.157	.016 -0.017 0.904	.118 (.016) 7.517	-.223 (.016) -13.658	-.036 (.017) -2.123	.135 (.016) 8.194	.193 (.015) 12.444	.023 (.020) 1.136	.382 (.016) 24.160

	Achieve	Support	Work	SD
Achieve	1.000			
Support	.350 (.016) 21.568	1.000		
Work	.605 (.014) 44.418	.279 (.016) 17.836	1.000	
SD	.098 (.017) 5.777	.068 (.017) 4.036	.285 (.015) 18.904	1.000

Table 4.56

95% confidence interval for sample (Φ) estimates

	Control	Leader	Organise	Planner	Detail	Rules	Thinker	Change	Finish	Notice	Belong	Social	Relate
Control	-												
Leader	.590-.645	-											
Organise	.161-.232	.488-.547	-										
Planner	.293-.360	.639-.686	.735-.770	-									
Detail	.211-.282	.512-.567	.705-.748	.773-.808	-								
Rules	.101-.175	.305-.368	.430-.489	.494-.549	.541-.592	-							
Thinker	.440-.507	.707-.754	.470-.529	.672-.719	.654-.708	.306-.373	-						
Change	.227-.294	.556-.607	.408-.467	.548-.599	.530-.585	.346-.409	.697-.740	-					
Finish	.240-.307	.567-.618	.765-.800	.765-.800	.825-.856	.520-.571	.593-.644	.548-.599	-				
Notice	.547-.606	.159-.226	-.019-.047	-.018-.048	-.002- -.068	-.003- -.069	.095-.165	-.033- -.033	-.018- -.044	-			
Belong	.157-.224	.435-.490	.333-.392	.442-.497	.410	.567-.618	.386-.449	.513-.568	.442.497	-.029- -.037	-		
Social	-.216- -.287	-.535- -.590	-.439- -.498	-.521- -.576	-.504- -.559	-.424- -.483	-.517- -.580	-.452- -.507	-.517- -.572	-.072- -.142	-.534- -.585	-	
Relate	.126-.200	.145-.212	.153-.216	.152-.215	.157-.224	.397-.460	.126-.196	.080-.146	.190-.253	.094-.160	.337-.396	-.408- - .471	-
Decision	.396-.463	.457-.516	.262-.325	.290-.353	.251-.318	.080-.146	.469-.532	.299-.358	.299-.358	.184-.251	.110-.172	-.323- -.390	.084-.150
Workpace	.300-.367	.479-.534	.499-.550	.459-.514	.400-.459	.192-.255	.495-.554	.355-.414	.487-.538	.136-.203	.162-.225	-.380- -.443	.114-.180
Forceful	.571-.646	.605-.672	.267-.346	.398-.469	.375-.450	.232-.311	.597-.663	.536-.606	.361-.432	.291-.370	.275-.350	-.269- -.352	.077-.159
Emotion	.232-.307	.288-.355	.328-.391	.369-.432	.356-.419	.199-.270	.318-.389	.274-.341	.334-.397	-.026- .048	.226-.293	-.447- -.510	.075-.145
Achieve	.483-.546	.642-.693	.593-.644	.642-.689	.600-.651	.467-.526	.603-.658	.617-.668	.679-.726	.191-.258	.464-.523	-.471- -.530	.255-.322
Support	.258-.329	.141-.208	.235-.298	.150-.213	.211-.278	.353-.416	.119-.189	.068-.134	.235-.298	.304-.371	.199-.262	-.318- -.381	.437-.496
Work	.221-.288	.505-.560	.581-.628	.588-.635	.567-.614	.408-.467	.465-.524	.447-.502	.655-.698	-.020- .042	.348-.407	-.439- -.498	.207-.270
SD	-.026- -.048	.028-.094	.236-.295	.201-.260	.191-.254	.074-.140	.139-.206	.140-.203	.217-.276	-.017-.049	.087-.149	-.191- -.254	-.003- -.069

	Decision	Workpace	Forceful	Emotion	Achieve	Support	Work	SD
Decision	-							
Workpace	.693-.736	-						
Forceful	.398-.472	.331-.406	-					
Emotion	.154-.225	.193-.260	-.006- -.092	-				
Achieve	.298-.361	.428-.487	.490-.564	.294-.361	-			
Support	.138-.205	.235-.302	.054-.136	.165-.236	.318-.381	-		
Work	.269-.328	.476-.527	.289-.368	.289-.352	.577-.632	.247-.310	-	
SD	.104-.166	.163-.222	-.016-.062	.350-.413	.065-.131	.035-.101	.255-.314	-

Table 4.57

Squared sample phi estimates and average variance extracted per latent variable

	Control	Leader	Organise	Planner	Detail	Rules	Thinker	Change	Finish	Notice	Belong	Social	Relate	Decision	Workpace	Forceful	Emotion	AVE
Control	-																	.229315
Leader	.382	-																.363217
Organise	.039	.268	-															.405367
Planner	.107	.440	.567	-														.471292
Detail	.061	.292	.529	.626	-													.401805
Rules	.019	.114	.212	.272	.321	-												.351146
Thinker	.225	.534	.25	.484	.465	.116	-											.268403
Change	.068	.339	.192	.329	.311	.143	.517	-										.360704
Finish	.075	.352	.613	.613	.707	.298	.383	.329	-									.405315
Notice	.333	.037	.000	.000	.001	.001	.017	.000	.000	-								.382811
Belong	.036	.214	.132	.221	.192	.352	.175	.293	.221	.000	-							.389577
Social	.064	.317	.220	.301	.283	.206	.301	.230	.297	.011	.314	-						.285158
Relate	.027	.032	.034	.034	.036	.184	.026	.013	.049	.016	.135	.194	-					.32993
Decision	.185	.237	.086	.104	.081	.013	.251	.108	.108	.048	.020	.127	.014	-				.366235
Workpace	.112	.257	.276	.237	.185	.050	.276	.148	.263	.029	.038	.170	.022	.511	-			.462767
Forceful	.372	.410	.094	.188	.171	.074	.398	.327	.158	.110	.098	.097	.014	.190	.136	-		.148504
Emotion	.073	.104	.130	.161	.151	.055	.125	.095	.134	.000	.068	.229	.012	.036	.052	.002	-	.241857
Achieve	.265	.446	.383	.444	.392	.247	.398	.413	.494	.051	.244	.251	.084	.109	.210	.279	.108	.229826
Support	.086	.031	.071	.033	.060	.148	.024	.010	.071	.114	.053	.123	.218	.030	.072	.009	.040	.319702
Work	.065	.284	.366	.375	.349	.192	.245	.226	.458	.000	.143	.220	.057	.089	.252	.108	.103	.414126
SD	.000	.004	.071	.053	.050	.011	.030	.030	.061	.000	.014	.050	.001	.018	.037	.001	.146	.367836
AVE	.229315	.363217	.405367	.471292	.401805	.351146	.268403	.360704	.405315	.382811	.389577	.285158	.32993	.366235	.462767	.148504	.241857	

Table 4.57 (continued)

Squared sample phi estimates and average variance extracted per latent variable

	Achieve	Support	Work	SD	AVE
Achieve	-				.229826
Support	.123	-			.319702
Work	.366	.078	-		.414126
SD	.010	.005	.081	-	.367836
AVE	.229826	.319702	.414126	.367836	

○ AVE < r^2 for both latent variables ○ AVE < r^2 for one of the latent variables ○ AVE > r^2 for both latent variables

4.6.4 STATISTICAL POWER ASSESSMENT

Statistical power is essential when evaluating decisions regarding the rejection, or not, of statistical hypotheses about the fit of a model. In the context of this study statistical power refers to the conditional probability of rejecting the null hypothesis given that it is false, i.e. in the case of the exact fit null hypothesis P[reject H_{02}: P(Reject H_{01}: RMSEA=0| H_{01} false]. In the context of SEM, statistical power therefore refers to the probability of rejecting an incorrect model. Diamantopoulos and Siguaw (2000) explain the difference between Type I and Type II error in the context of structural equation modelling as follows:

> When we test a model's fit by, say, the chi-square test, we emphasize the probability of making a **Type I error**, i.e. rejecting a correct model; this probability is captured by the **significance level**, α which is usually set at 0,05. A significant chi-square result indicates that *if* the null hypothesis is true (i.e. the model is correct in the population), then the probability of incorrectly rejecting it is low (i.e. less than five times out of 100 if α= 0,05). However, another error that can occur is *not* to reject an incorrect model. This type of error is known as **Type II error** and the probability associated with it is denoted as β. The probability of avoiding a Type II error is, therefore, 1-β and it is this probability that indicates the power of our test; thus the power of the test tells us how likely it is that a false null hypothesis (i.e. incorrect model) will be rejected. (p. 93).

Unfortunately, this aspect is more often than not ignored, but it is important to understand that any model evaluation would be incomplete if statistical power considerations were neglected. The importance of conducting a power analysis stems from the critical role that sample size plays in the decisions made in model testing (Diamantopoulos & Siguaw, 2000). For instance, if the close fit null hypothesis (H_0: RMSEA≤.05) would not be rejected, the concern that arises is whether this result is due to a lack of statistical power or whether it accurately reflects the true state of affairs. This concern increases as sample size decreases. Specifically in small samples (i.e., low power conditions) the decision not to reject the null hypothesis causes ambiguity because it is not clear whether the decision was due to the accuracy of the model or the insensitivity of the test to detect specification errors in the model. Conversely, however, in large samples (i.e., high power conditions) the decision to reject the null hypothesis would create the fear that a reasonably accurate model was rejected because of relatively minor misspecifications rather than due to severe flaws in the model. When the chi-square test is applied only Type I errors are explicitly taken into account, thus, a power analysis must be undertaken to also account for the probability of Type II errors (Diamantopoulos & Siguaw, 2000).

Two types of power calculations can be performed. First, the power associated with a test of exact fit (H_{01}: RMSEA=0), as done by the Satorra-Bentler chi-square test, can be estimated. However, as argued before, this test is very limited since models are only approximations of reality and, therefore, rarely do they fit exactly in the population. Only the power associated with the test of close fit was therefore estimated. The null hypothesis of close fit states that the model has a close, but imperfect fit in the population. The statistical hypothesis of close fit makes use of the RMSEA statistic. If a model fits closely in the population the error due to approximation is set at .05 and the null hypothesis formulated earlier as H_{02a} is consequently tested against H_{a2a} (Diamantopoulos & Siguaw, 2000).

To determine the power of a test of the close fit hypothesis, at least one specific value for the parameter needs to be assumed under H_{a2} because there are as many power estimates as there are possible values for the parameter under H_{a2}. A value that makes good sense to use in this instance is RMSEA=.08, since RMSEA=.08 is the upper limit of reasonable fit. In this specific analysis two additional possible values for RMSEA under H_{a2} were also considered, namely .70 and .60.

The statistical power of the test of close fit becomes a function of the effect size, the significance level, the sample size N and the degrees of freedom (v) in the model (v=½[(p][p+1]-t)=8001-462=7539). Power tables compiled by MacCallum, Browne, and Sugawara (1996) only make provision for degrees of freedom ≤ 100 and N ≤ 500. A SPSS translation of the SAS syntax provided by MacCallum *et al.* (1996) was consequently used to derive power estimates for the tests of close fit. The effect size assumed above, a significance level (α) of .05 and a sample size of 5817 were used to calculate P[reject H_{01}: RMSEA≤.05| RMSEA=.08], P[reject H_{01}: RMSEA≤.05| RMSEA=.07] and P[reject H_{01}: RMSEA≤.05| RMSEA=.06]. The results of the power analysis are reported in Table 4.58.

Table 4.58

Statistical Power of the Null hypothesis of close fit under three different h_{a2} scenarios

H_0	H_a	N	ncp0	ncpa	cval	α	df	power
H_0: RMSEA≤.05	H_a: RMSEA=.08	5817	109617.060	280619.673	1001.000	0.05	7539	1.000
H_0: RMSEA≤.05	H_a: RMSEA=.07	5817	109617.060	214849.438	1001.000	0.05	7539	1.000
H_0: RMSEA≤.05	H_a: RMSEA=.06	5817	109617.060	157848.566	1001.000	0.05	7539	1.000

Table 4.58 indicates that the probability of rejecting the null hypothesis of close fit under the true condition of mediocre fit (i.e., RMSEA=.80) is unity. Furthermore, if it were assumed that the true model fit in the population was RMSEA=.70 or .60, the power of the test of close fit would still be very high (1.000). These power estimates, taken in conjunction with the fact that the close fit null hypothesis was not rejected, bolsters confidence in the merits of the model. It could therefore be concluded that the decision not to reject the close fit null hypothesis cannot be attributed to a lack of statistical power as the test is highly sensitive to misspecifications in the model.

4.6.5 SUMMARY

International research has explored the psychometric properties of the PAPI-N in various settings globally. This far no known study has been done on a relatively large South African sample. This chapter explored the psychometric properties of the PAPI-N. In addition, there is little empirical construct validity evidence available for the PAPI-N in South Africa. Consequently, there existed a need to investigate whether the PAPI-N provides a construct valid measure of personality within the South African work context.

The substantive hypothesis tested in this study was that the PAPI-N provides a construct valid and reliable measure of personality within a work context as it is defined by the instrument, amongst South African employees.

In operational terms the hypothesis tested in this study was that the measurement model implied by the scoring key and the design intention of the PAPI-N can closely reproduce the co-variances observed between the individual items comprising each of the PAPI-N scales. The operational hypothesis implied by the substantive research hypothesis further implies that the factor loadings of the individual items on their designed latent personality dimensions are significant and large, that the personality dimensions explain large proportions of the variance in the individual items and lastly that the latent personality dimensions correlate low to moderately with each other.

Multiple criteria fit indices were utilised to test the fit of the first-order measurement model of the PAPI-N. The model's overall fit was found to be acceptable. The null hypothesis of close fit was not rejected, and the basket of LISREL results were indicative of a close fit in the parameter and a very good model fit in the sample. However, a large percentage of standardised residuals, together with a high percentage of modification indices calculated for the $\mathbf{\Lambda}^{\mathbf{X}}$ matrix contradicted this relatively positive conclusion. On the other hand, a small percentage of the modification indices calculated for the Θ_δ matrix commented favourably on the fit of the measurement model.

The measurement of personality via the PAPI-N is not without problems, as factor loadings, although statistically significant, appears to be moderate. In addition, the measurement error variances was relatively large and the proportion variance explained in the linear item composites low. It should be noted that this finding was somewhat expected given the broad nature of the personality dimensions and the fact that responses to the items are determined by the whole personality.

The latent personality dimensions correlate low to moderate with each other in the sample. Although one correlation (ϕ_{95}, the correlation between attention to detail and need to finish a task) exceeded .80, none in the phi matrix were excessively high (<.90). Confidence intervals calculated for the phi estimates failed to include unity for any of the 210 correlations in **Φ**. Furthermore, in the case of 20.47% of the latent personality dimension pairs one or both latent variables account for less variance in their indicator variables that were designed to reflect them for which they account for in each other. This finding implied that PAPI-N, although with some difficulty, successfully discriminates between the unique aspects of the latent personality dimensions.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

5.1 INTRODUCTION

Since its inception, PAPI-N has become a leading work-related personality questionnaire used by HR professionals and line managers across the globe (e.g. France, Swedish, Turkish, Chinese, German) (Cubiks, 2012). It is also widely used by psychologists and psychometrists in personnel selection in South Africa. The confident utilisation of the PAPI-N in selection in South Africa requires (a) that a convincing argument be developed as to why and how personality (as interpreted by PAPI-N) should be related to job performance, (b) that a structural model derived from the foregoing argument fits empirical data, i.e. there is support for the performance hypothesis, (c) that evidence be available that the predictor and criterion constructs are validly and reliably measured in the various sub-groups typically comprising applicant groups in South Africa, (d) that evidence be available that (at least) race and gender group membership do not systematically affect the manner in which the predictor and criterion constructs express themselves in observed measures, (e) that evidence be available that the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures, (f) that evidence be available that criterion predictions derived (clinical or mechanical) from the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures and (g) that evidence be available whether (at least) race and gender group membership does explain variance in the criterion (either as a main effect or in-interaction with the criterion estimates derived from the PAPI-N) that is not explained by the criterion estimates derived from the PAPI-N. The objective of this research is to contribute to the available psychometric evidence with regards to the third aspect (c) mentioned above.

The PAPI-N is based on a specific interpretation of personality. The architecture of the instrument reflects a specific design intention. The design of the PAPI-N questionnaire reflects the intention to construct twenty essentially one-dimensional sets of six items each to reflect variance in each of the twenty latent personality domains collectively comprising the personality construct. In conjunction with its design and the connotative meaning the PAPI-N attaches to the personality construct, the scoring key denotes a specific measurement model which suggests that responses to specific items of the PAPI-N are a function of a specific underlying latent personality dimension.

Furthermore, the design intention of the PAPI-N is that the items load statistically significant ($p<.05$) and large ($\lambda_{ij}\leq.50$) on their designated latent personality dimensions, that the measurement error variance associated with each item is statistically significant ($p<.05$) but small, that the latent personality dimensions explain large proportions of the variance in the items that represent them ($\lambda^2_{ij}\geq.25$) and finally that the latent personality dimensions correlate low-moderate ($\phi_{ij}<.90$) with each other.

The objective of this study was to evaluate the fit of the (first-order) PAPI-N measurement model, as implied by the architecture of the instrument and the constitutive definition underlying its constructs, on a relatively large sample of the South African working population.

This chapter provides a summary of the results of the study, discusses the implications of the findings and makes recommendations for future research.

5.2 SUMMARY OF THE PRINCIPAL FINDINGS AND DISCUSSION

The measurement model was fitted by representing the twenty latent personality dimensions with individual items. Item analysis and exploratory factor analysis were performed on each of the PAPI-N subscales. In this study, these analyses served the purpose of evaluating whether the items comprising each subscale provides an internally consistent and essentially uni-dimensional measure of the latent personality dimension that it was intended to measure.

Chapter 4 outlined the detailed results of the study. The following conclusions are made with regards to the item analyses, the dimensionality analyses and the fit of the hypothesised measurement model.

5.2.1 ITEM ANALYSIS

The design of the PAPI-N questionnaire reflects the intention to construct twenty essentially one-dimensional sets of items to reflect variance in each of the twenty latent personality dimensions collectively comprising the personality construct. The PAPI-N items are meant to operate as stimulus sets to which respondents react with behaviour, that is primarily a relatively uncontaminated expression of a specific underlying latent personality dimension. If they do so successfully, the items of a subscale will correlate reasonably high, the item-total correlations will be high, the squared multiple correlations will be high when regressing each item on a weighted linear composite of the remaining items, the coefficient of internal consistency will be high for the subscale and will decrease when any item is deleted from the subscale.

A variety of item statistics were calculated to determine how well the items of a subscale reflect the content of the underlying personality dimension. The objective was to identify the presence of poor items. Poor items are items that fail to discriminate between different states of the latent variable they are meant to reflect as well as items that do not reflect a common latent variable. Two of the subscales reflected high alpha coefficients (i.e., $.90>\alpha\geq.80$). Eight subscales showed acceptable alpha coefficients (i.e. $.80>\alpha\geq.70$), while seven returned questionable alpha coefficients (i.e. $.70>\alpha \geq.60$). Two scales obtained poor alpha coefficients (i.e. $.60>\alpha\geq.50$) and only a single subscale returned an unacceptable reliability coefficient (i.e. $.50 > \alpha$). The reliability coefficient findings for the subscales are summarised in Table 5.1. The results on the descriptive statistics concluded that in about 50% the items of the subscales did not respond in unison to systematic differences in a single underlying latent variable. Internal consistency in these subscales was either questionable, poor or unacceptable. In addition, the inter-item correlation statistics and item-total statistics, predominantly the corrected item-total correlations, squared multiple correlations and increases in alpha if the item were deleted, further revealed a number of items that contributed towards lowering the homogeneity of the scales. This would have usually resulted in the deletion of these items, however, due to the nature of this study, all items were retained.

Table 5.1

Evaluation of the internal consistency reliability findings

Cronbach's alpha	**Evaluation**	**PAPI-N subscale**
$\alpha \geq .90$	Excellent	
$.90 > \alpha \geq .80$	Good	H, T
$.80 > \alpha \geq .70$	Acceptable	B, C, D, I, N, W, X, Z
$.70 > \alpha \geq .60$	Questionable	E, F, G, L, O, P, R
$.60 > \alpha \geq .50$	Poor	A, S
$.50 > \alpha$	Unacceptable	K

5.2.2 DIMENSIONALITY ANALYSIS

Unrestricted principal axis factor analysis with oblique rotation were performed on each of the PAPI-N subscales. The purpose was to confirm the uni-dimensionality of each subscale and to evaluate the success with which each item, along with the rest in that particular item set, measures the specific latent personality dimension it was intended to measure. The factor loading of each item was used to assess the degree to which the item reflected the latent variables underlying the subscales.

The eigenvalue-greater-than-unity rule of thumb along with the scree plot as discussed in Chapter 4 were used to determine the number of factors to be extracted. Eight out of the 20 subscales failed the uni-dimensionality test, namely subscales W (Need for Rules and Supervision), R (Conceptual Thinker), X (Need to be noticed), S (Social Harmoniser), T (Work pace), K (Need to be forceful), A (Need to achieve) and F (Need to be supportive). In all eight of the subscales, two factors had to be extracted to provide a satisfactory explanation of the observed correlation matrix. In seven of these cases the factor fission resulted in a meaningful theoretical interpretation of a common theme. In case of subscale K, the identity of the two underlying factors proved to be somewhat elusive. Subscale K also returned an unacceptably low internal consistency reliability (α=.424). the inability to extract meaningful factors can therefore in part be attributed to the large random error component in the subscale data. In seven of the cases the two-factor solution was suggested by the eigenvalue-greater-than-unity rule. In case of the T subscale the extraction of a second factor was suggested by the large percentage of large residual correlations.

Judging from the number of factors extracted and the magnitude of the factor loadings when a single factor is extracted, the present study seems to suggest that 90% of the items generally do systematically reflect their designated latent variables with reasonable success. To evaluate the success with which the items in those eight subscales reflect the original personality dimension that now in effect functions as a second-order personality factor the researcher forced SPSS to extract a single factor. When forcing the extraction of a single factor, in many of these cases, the factor loadings were generally satisfactory. Furthermore, differential skewness and differential kurtosis were inspected to determine whether the extracted factor structure could be described in terms of systematic differences in these item statistics. For all the subscales, with the exception of the A-subscale, neither differential skewness nor kurtosis was found to provide a credible explanation for the extracted factors.

Finally, the residuals correlations were further examined for both the two-factor and the one-factor solution. Of major importance is that the residuals calculated from the inter-item correlation matrix and the reproduced matrix showed that the initial solutions, prior to forcing a single factor, provide a much more convincing explanation for the observed inter-item correlation matrix than the single-factor solution. This is indicative of the fact that the personality dimensions measured by at least seven of the eight subscales should rather be interpreted as second-order personality factors on which two first-order factors load. Moreover, this makes it likely that the failure of the uni-dimensionality test on seven of the subscales could be explained by a splitting of the primary factors into narrower sub-factors.

The purpose of the dimensionality analyses was to provide a better understanding and insight into the functioning of the first-order factor structure of the PAPI-N scales. Further to this, the analyses assisted in gaining an understanding of the psychometric integrity of the indicator variables that were tasked to represent each of the latent personality variables. In conclusion, the results of the exploratory factor analysis for 12 of the 20 personality dimensions measures were compatible with the position that the items comprising these subscales measure what they were designed to measure. The extraction of a single factor, adequate factor loadings and a small percentage of large residual correlations, however, constitutes insufficient evidence to conclude with certainty that the latent variable being measured is the latent personality dimension the PAPI-N intended to measure. To achieve this certainty a structural model would have to be fitted in which the latent personality dimensions are individually modelled that reflects the manner in which the personality dimensions are causally related to each other and that reflects the manner in which the personality construct is embedded in a larger nomological network of latent variables according to the PAPI-N's constitutive definition of personality.

5.2.3 MEASUREMENT MODEL FIT

The measurement model's overall fit was acceptable. The null hypothesis of exact fit was rejected but the null hypothesis of close fit could not be rejected (p>.05). The fit indices reflected a close fit in the parameter and a very good model fit in the sample. A large percentage of standardised co-variance residuals and a high percentage of Λ^X modification indices contradicted this positive conclusion. At the same time, a small percentage of significant Θ_δ modification indices commented favourably on the fit of the measurement model.

In the case of the PAPI-N, a very specific stance is taken on the number and identity of latent personality dimensions underlying the observed inter-item co-variance matrix and the manner in which the items load on the personality factors. The fact that the measurement model fits the data closely suggests that the particular measurement model offers a valid (i.e. permissible) and credible description of the psychological process underlying the PAPI-N. Moreover, it suggests that the measurement model offers a valid and credible account of the process that produced the observed inter-item co-variance matrix as the pattern of co-variances observed between the combinations of items was satisfactorily explained by the model. The fact that the measurement model was able to closely reproduce the observed inter-item co-variance matrix does however not mean that the process depicted by the measurement model is in fact the only one operating to determine test takers' responses to the PAPI-N items. It only suggests that the process depicted by the measurement model provides one possible process that might have produced the observed inter-item covariance matrix.

Although the measurement model fitted the data closely, the factor loadings (although statistically significant) were generally of a moderate degree. Furthermore, the measurement error variances were relatively large and the proportion variance explained in the items were low. It should be noted that this finding was to some degree expected given the broad nature of the personality dimensions and the fact that responses to the items are determined by the whole personality.

5.3 LIMITATIONS

In this study the non-probability sampling method of convenience sampling was used which means that the findings of this study can only be generalised to the target population with great caution. More specifically, the extent to which observations can or may be generalised to the target population depends on the representativeness of the sample. The majority of the respondent's information did not include biographical information such as age, educational level, occupation, and management responsibility, which is certainly an unfortunate shortcoming in this study as it prevents the proper characterisation of the sample's age, educational level, occupation and managerial/non-managerial responsibility. A more accurate description of the research sample would have been desirable because these characteristics undoubtedly affect how respondents respond to the items comprising the PAPI-N.

Another possible limitation could be the language of the questionnaire, especially for those who are not fluent in English. Language could have played an important role in the way test takers respond to the items, especially the reflected items. The critical question is whether these single-group measurement models fit data obtained for each of the language groups. If at least a close fit is obtained the question in addition is whether the measurement model parameters can be considered satisfactory. If a single-group measurement model fit is achieved in the various language groups the question on the measurement invariance and equivalence of the different language versions of the South African PAPI-N also has to be examined.

A similar question exists with regards to gender and racioethnic sub-groups. Demonstrating that PAPI-N measurement model fit has been shown on a large South African sample would still beg the question whether the measurement model fit holds across the various gender-racioethnic sub-groups, and if so, whether the model parameters are the same across such groups. Therefore, a critical question is whether the measurement model underlying the PAPI-N succeeds in measuring the construct across different gender-racioethnic groups as it was constitutively defined and whether the inference that can be made about the state/level of the measured construct given a specific observed score is the same across groups.

The archival data on the PAPI-N was collected over a 6-year period from 2007 until 2012. This raises the question whether the data has to be considered outdated and no longer fit for psychometric analysis. The ideal methodological scenario would no doubt be to analyse the most recent data possible. This ideal, however, is in conflict with the need for large samples and the practical reality that the PAPI-N is only used on a limited scale in South Africa. It takes time to collate a sufficiently large sample that allows the use of single group confirmatory factor analysis to evaluate the fit of a measurement model in which the individual items serve as indicators. The question on the relevance of the data hinges on the extent to which the to-be-measured construct changes over time and the extent to which the behavioural denotations of the construct change over time. The researcher is of the opinion that personality is a relatively non-malleable construct and that the denotations of the construct change relatively slowly. The researcher is consequently of the opinion that the archival data used in this research is still sufficiently recent to allow the results to be interpreted and used with confidence.

Even though these limitations are noteworthy and should be taken cognisance of, the researcher is nevertheless convinced that the study will contribute to a greater understanding of the psychometric properties of the PAPI-N on samples different from the UK sample. Further to this, that the study will without doubt assist in eliciting the necessary further research needed to establish the psychometric credentials of the PAPI-N as a valuable assessment instrument in South Africa with confidence.

5.4 RECOMMENDATIONS FOR FUTURE RESEARCH

As mentioned in the previous discussion of the limitations of this research study, there are critical questions that still need to be addressed. The following proposals for future research on PAPI-N are suggested:

- The single-group measurement model should be fitted on each of the language groups. If at least a close fit is obtained the question in addition is whether the measurement model parameters can be considered satisfactory. If a single-group measurement model fit is achieved in the various language groups the question on the measurement invariance and equivalence of the different language versions of the South African PAPI-N also has to be examined;
- Future studies should also investigate the measurement invariance and equivalence of the PAPI-N for different racio-ethnic and gender groups via a multi-group confirmatory factor analysis;
- Ideal PAPI-N profiles should be developed for specific jobs or job families along with behavioural observation scale measures of performance. The predictive validity of a profile similarity measure calculated on the PAPI-N first-order factors in predicting job performance should be investigated;
- A structural model should be developed in which the latent personality dimensions are individually modelled and that explicates the manner in which the personality dimensions are causally related to each other and that explicates the manner in which the personality construct is embedded in a larger nomological network of latent variables according to the PAPI-N's constitutive definition of personality. The fit of this structural model should be evaluated. This would shed further light on the construct validity of the PAPI-N;
- If the aforementioned structural model would show at least close fit the structural invariance and equivalence should be investigated across different racio-ethnic and gender

groups via a multi-group confirmatory factor analysis. This would shed further light on the possibility of construct bias in the PAPI-N; and

- Finally, the validity, fairness and utility of criterion predictions derived (clinical or mechanical) from the measures of the PAPI-N should be investigated.

5.5 CONCLUDING SUMMARY

PAPI-N has become a leading work-related personality questionnaire used by more than 5000 professionals across the globe (Cubiks, 2012). It is also widely used by psychologists and psychometrists in South Africa. The confident use of the PAPI-N in selection in South Africa requires (a) that a convincing argument be developed as to why and how personality (as interpreted by PAPI-N) should be related to job performance, (b) that a structural model derived from the foregoing argument fits empirical data, i.e. there is support for the performance hypothesis, (c) that evidence be available that the predictor and criterion constructs are validly and reliably measured in the various sub-groups typically comprising applicant groups in South Africa, (d) that evidence be available that (at least) race and gender group membership do not systematically affect the manner in which the predictor and criterion constructs express themselves in observed measures (i.e. that the predictor and criterion measures are unbiased), (e) that evidence be available that the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures, (f) that evidence be available that criterion predictions derived (clinical or mechanical) from the measures of the PAPI-N correlate statistically significantly ($p<.05$) with construct valid criterion measures and (g) that evidence be available whether (at least) race and gender group membership does explain variance in the criterion (either as a main effect or in-interaction with the criterion estimates derived from the PAPI-N) that is not explained by the criterion estimates derived from the PAPI-N (i.e. whether the criterion inferences derived from the PAPI-N can be considered predictively unbiased).

The objective of this research was to contribute to the available psychometric evidence with regards to the third aspect (c) mentioned above. The confident utilisation of the PAPI-N in specific personnel selection procedures aimed at filling posts in specific positions in specific organisations would, however, in addition to the above also require credible evidence on the fairness and utility (Guion, 1998) of the selection procedure.

The data used in this study was attained from the data archives of Cubiks (Pty) Ltd, with written permission of the test distributor to utilise the sample data for the purpose of this research. The initial South African PAPI-N database comprised all respondents who were assessed by Work Dynamics in the period 2007 to 2012 for various assessment purposes as requested by their client organisations across different industries and occupations or jobs. The database contained the scale scores and individual raw item scores for each of the items comprising the PAPI-N, and self-reported information on each respondent's gender, nationality, ethnicity, and mother language. However, records of the sample have been provided in an anonymous format which can at no point be traced back to an individual's score or profile as anonymity was a condition set by Cubiks and the Research ethics Committee (Humanities) of Stellenbosch University to gain access to the archival data base.

The objective of this research study was mainly to examine the first-order factor structure of the PAPI-N through a factor analytic investigation. This was done through confirmatory factor analysis. Prior to this analysis, item and dimensionality analysis were used to determine the extent to which each one of the dedicated items of each of the PAPI-N subscales satisfactorily reflect the underlying personality dimension it was intended to represent. The PAPI-N measurement model was fitted using individual items that reflects the structural design intention of the PAPI-N.

The results suggests that while the intention of the PAPI-N to have sets of items replicating specific primary personality factors succeeded, the subscale measures of the primary personality factors mostly hold a sizable amount of systematic and random error. The evidence resulting from this study hence produces a certain degree of uneasiness about the use of the PAPI-N in personnel selection in South Africa since the available evidence does not allow the conclusion that the PAPI-N provides a highly reliable and valid measure of personality within the multi-cultural South African work context.

Psychological tests as measures of psychological constructs should not be psychometric evaluated in absolute terms only but also be psychometric evaluated comparatively. This is not without difficulty even when the focus is restricted to a specific construct like personality for the simple reason that different instruments attach different connotative meaning to the personality construct. Despite this it should still be possible to compare how successful various instruments meet specific criteria set out to determine how well the instrument measures its designated target construct. Very few South African personality measures have been confronted with SEM-based confirmatory factor analysis. The 15FQ+ single-group study on Black managers by Moyo (2009) is the only South African study that could be traced in the literature. The results obtained by Moyo (2009) were roughly similar to those obtained in this study in terms of subscale reliability, violations of the uni-dimensionality assumption, the magnitude of the factor loadings and the magnitude of the error variances.

To pronounce a verdict on the validity of the PAPI-N measures with confidence, there is a need to broaden the scope of this study by appropriately attending to the proposed recommendations mentioned above and questions that would themselves generate future research.

REFERENCES

Allport, G.W. (1937). *Personality: A Psychological Interpretation*. New York: Holt, Rinehart & Winston.

Allport, G.W. (1961). *Pattern and Growth in Personality*. New York: Holt, Rinehart & Winston.

Allport, G.W., & Odbert, H.S. (1936). Trait-names: a psycho-lexical study. *Psychological Monographs*, *47*, 1-171.

Akrani, G. (2009). *Human Resource Management and Development – HRM Notes*. Retrieved February 24, 2012: http://kalyan-city.blogspot.com/2009/12/human-resource-management-human.html.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing. (7th ed.).* London, UK: Prentice Hall International.

Anderson, P., & Lewis, C. (1998). *PAPI Technical Manual*. London, PA Consulting/Cubiks.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290.

Ashton, M.C. (1998). Personality and job performance: The importance of narrow traits. *Journal of Organisational Behaviour, 19*, 289-303.

Austin, J.T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, *77*, 836-874.

Babbie, E,. & Mouton, J. (2001). *The Practice of Social Research*. Cape Town: Oxford University Press.

Bagozzi, R.P., Yi, Y., & Phillips, L.W. (1991). Assessing Construct Validity in Organization Research. *Administrative Science Quarterly, 36* (3), 421-458.

Bandalos, D. (2002). The effects of item parcelling on goodness-of-fit and parameter estimate bias in structural equation modelling. *Structural Equation Modeling, 9* (1), 78-102.

Bandalos, D.L., & Finney, S.J. (2001). Item parcelling issues in structural equation modelling. In G.A Marcoulides & R.E Schumacker (Eds.), *Advanced structural equation modelling: New developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bandura, A. (1977). *Social Learning Theory*. New York: General Learning Press.

Bartram, D., Baron, H. & Kurz, R. (2003). *Let's turn validation on its head*. Occupational Psychology Conference of the British Psychological Society Bournemouth.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the Criterion Domain to Include Elements of Contextual Performance. In N. Schmidt, W. C. Borman, A. Howard, A. Kraut, D. Ilgen,

B. Schneider, & S. Zedeck (Eds.), *Personnel Selection in Organizations* (pp 71-98). San Francisco: Jossey-Bass.

Binning, J.F,. & Barrett, G.V. (1989). Validity of personnel decisions: a conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74* (3), 478-494.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.

Bollen, K.A., & Long, J.S. (1993). *Testing Structural Equation Models.* Newbury Park: Sage Publications, Inc.

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

Byrne, B.M. (1989). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models*. New York: Springer Verlag.

Byrne, B.M. (1998). *Structural Equation Modelling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications, and Programming*. New Jersey: Lawrence Erlbaum.

Cascio, W.F. (1998). *Applied psychology in personnel management* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Cattell, R.B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, *38*, 476-506.

Cattell, R.B. (1950). *Personality: a Systematic, Theoretical and Factual Study*. New York: McGraw-Hill.

Cattell, R.B. (1965). *The Scientific Analysis of Personality*. Baltimore: Penguin

Cattell, R.B. (1979). Are culture fair tests possible and necessary? *Journal of Research and Development in Education, 12*, 1-13.

Cattell, R.B. (1979).*Personality and Learning Theory: Vol. 1. The Structure of Personality in its Environment*. New York: Springer.

Cattell, R.B. (1990). Advances in Cattellian personality theory. In L.A. Pervin (Ed.). *Handbook of Personality*: *Theory and Research (pp.101-110).* New York: Guilford.

Cattell, R.B., Eber, H.W., & Tatsuoka, M. (1970). *Handbook of the Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality & Ability Testing.

Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption-Inovation Inventory using multiple-group mean and co-variance structure analysis. Multivariate Behavioral Research, *35* (2), 169-199.

Cherry, K. (n.d.). Cattell's 16 Personality Factors. About.com Psychology. Retrieved June 7, 2012, from http://psychology.about.com/od/trait-theories-personality/a/16-personality-factors.htm.

Chong, E., & Hughes, D. (2004). Defining the work-based personality traits and competencies: A Comparative Study of British and Belgian Interim Managers, *18th Annual Conference of the Australian and New Zealand Academy of Management*, 2004.

Clark, L.A., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.

Cleary, T.A. (1968). Test bias: prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, *5* (2), 115-124.

Conn, S.R., & Rieke, M.L. (1994). *16PF Fifth Edition Technical Manual*. Champaign, IL, Institute for Personality and Ability Testing, Inc.

Cronbach, L.J., & Gleser, G.C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana, Ill: University of Illinois Press.

Costa, P.T., & McCrae, R.R. (1985a). *Manual for the NEO Personality Inventory*. Odessa, FL: Psychological Assessment Resources.

Costa, P.T., & McCrae, R.R. (1985b). *The NEO I Personality Inventory and Five-factor Model Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P.T., & McCrae, R.R. (1989). *NEO PI/FFI Manual Supplement*. Odessa, FL: Psychological Assessment Resources.

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

Costa, P.T. & McCrae, R.R. (1995). Domains and facets: hierarchical personality assessment using the Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 64*, 21-50.

Cubiks. (2003). Sales advisors in Dutch insurance organisation: Relating PAPI-N^{TM} personality dimensions to sales performance. *Cubiks Assessment Reliability and Validity Series*, V04.12. [Online]. Retrieved June 28, 2012: http://www.cubiks.com/InformationCentre/Pages/validity_stuidies.aspx

Cubiks. (2004). Identifying the relationship between PAPI^{TM} and the personality variables measured by FIRO Element B^{TM}. *Cubiks Assessment Reliability and Validity Series*, V04.13. [Online]. Retrieved June 28, 2012: http://www.cubiks.com/InformationCentre/Pages/validity_studies.aspx

Cubiks. (2006). Investigating the relationship between PAPI-N™ and the lexical Big Five personality factors. *Cubiks Assessment Reliability and Validity Series*, V06.14. [Online]. Retrieved June 30, 2012: http://www.cubiks.com/InformationCentre/Pages/validity_studies.aspx

Cubiks. (2007). *Personality and P Inventory PAPI-N (Normative) for South Africa: Supplement User Guide*. Unpublished guide.

Cubiks. (2012). PAPI™. [Brochure]. Retrieved from: http://www.cubiks.com/SiteCollectionDocuments/Files%20ENG/Flyers%20ENG/PAPI_Brochure_ENG.pdf

Davidson, M.C.G. (2000). *Organisational climate and its influence upon performance: A study of Australian hotels in South East Questions*. Unpublished doctoral thesis, Griffith University.

De Goede, (2007). *An investigation into the learning structure of the learning potential construct as measured by the APIL test battery.* Unpublished master's thesis, University of Stellenbosch, Stellenbosch.

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

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

Du Toit, S.H.C. & Mels, G. (2002). *Supplementary notes on multiple imputation*. http://www.ssicentral.com/other/impute.htm.

Dunbar-Isaacson, H. (2006). An investigation into the measurement invariance of the Performance Index. Unpublished master's thesis. Stellenbosch University.

Einhorn, H.J., & Bass, A.R. (1971). Methodological considerations relevant to discrimination in employment testing. *Psychological Bulletin*, *75* (4), 261-269.

Ellis, M.V., & Blustein, D.L. (1991). Developing and using educational and psychological tests and measures: the unificationist perspective. *Journal of Counselling and Development, 69*, 550-555.

Elkington, J. (1998). *Cannibals with forks: The triple bottom line of 21st century business (2nd ed.)*. Gabriola Island, BC Canada: New Society Publishers.

Farrell, A.M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty and Shu. Business Research, 63, 324-327.

Felthan, R., & Hughes, D. (1999). Interim managers: Distinguishing personality characteristics of managers on short-term contracts. *International Journal of Selection and Assessment, 7*, 209-214.

Ferrando, P.J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multi-dimensional test items: some aspects of the problem and some suggestions. *Psicológica, 21*, 301-323.

Foxcroft, C., Paterson, H., Le Roux, N., & Herbst, D. (2004). *Psychological assessment in South Africa: A needs analysis. The test use patterns and needs of psychological practitioners* (Human Sciences Research Rep.). Pretoria: HSRC.

Foxcroft, C. & Roodt, G. (Eds.). (2005). *An introduction to psychological assessment in the South African* context (2nd.ed.). Cape Town: Oxford University press.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.).* Boston: Allyn & Bacon.

Ghiselli, E.E., Campbell, J.P. & Zedeck, S. (1981). *Measurement theory for the behavioural sciences*. San Francisco, CA: Freeman and Company.

Goodstein, L.D., & Lanyon, R.I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business and Psychology, 13*, 291-322.

Guion, R.M. (1998). Assessment, measurement and prediction for personnel decisions. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Guion, R. M., Gottier, R. F. (1965). Validity of personality measures in personnel selection.
Personnel Psychology, 18, 135-164.

Hackett, P. (1985). *Success in Management: Personnel (2nd ed.)*. London: Murray Publishers.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis (7th ed.)*. Upper Saddle River, NJ: Prentice Hall.

Habeeb, F. (2009). Human Resource Management (Competitive advantage). Retrieved April 15, 2012: http://www.slideshare.net/fathima_sy/human-resource-management-competitive advantage-1928241.

Heffner, C.L. (2011, November 29). *Raymond Cattell and the 16PF*. Retrieved from http://allpsych.com/personalitysynopsis/cattell.html.

Holtzkamp, J. (2013). *Measurement Invariance of the second edition of the Fifteen Factor Personality Questionnaire (15FQ+) over different ethnic groups in South Africa.* Unpublished master's thesis. Stellenbosch: Stellenbosch University, Stellenbosch.

Hooper, D., Coughlan, J. and Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60.th

HR-Guide to Internet: Personnel Selection: Overview. (2001). Retrieved 25 February, 2012, http://www.hr-guide.com/data/G300.htm

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response theory: Applications to psychological measurement.* Homewood, IL: Dow Jones-Irvin.

Hu, L., & Bentler, P.M. (1999). Cut-off criteria in fix indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, *6*(1), 1-55.

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

Implicit Personality Theory. [n.d.] [Online]. Retrieved June 24, 2012: http://en.wikipedia.org/wiki/Implicit_personality_theory.

Jackson, D.N. (1984). *Personality research form manual (3rd ed.)*. Port Huron, MI: Research Psychologists Press.

John, O.P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement and theoretical perspectives. In L.A. Pervin and O.P. John (Eds.), *Handbook of personality*: *Theory and Research (2nd ed.,* pp. 102- 138). New York: Guilford Press.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modelling with the SIMPLIS Command Language.* United States of America: Scientific Software International, Inc.

Jöreskog, K.G., & Sörbom, D. (1996a). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1996b). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Jöreskog, K.G., & Sörbom, D. (1998). *Structural equation modelling with SIMPLIS command language.* Chicago: Scientific Software International.

Jöreskog, K.G. & Sörbom, D. (2003). *LISREL 8: New Statistical Features (Third Printing with Revisions*. Lincolnwood, IL: Software International, Inc.

Jung, C.G. (1921, 1971). *Psychological Types*. Routledge: London.

Kaplan, D. (2000). *Structural Equation Modelling: Foundations and Extensions*. Thousand Oaks, California: Sage Publications.

Kaplan, R.M., & Saccuzzo, D.P. (2001). *Psychological Testing: Principle, Applications and Issues (5th Edition)*, Belmont, CA: Wadsworth.

Kelloway, E.K. (1998). *Using LISREL for structural equation modelling: a researcher's guide.* Thousand Oaks: Sage Publications.

Kerlinger, F.N. (1973). *Foundations of behavioural research. (2nd ed.).* New York: Holt, Renehart and Winston, Inc.

Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of Behavioural Research. (4th ed.)*. Fort Worth: Harcourt College.

Kim, S., & Hagtvet, K.A. (2003). The impact of misspecified item parcelling on representing latent variable in co-variance structure modelling: a simulation study. *Structural Equation Modeling, 10* (1), 101-127.

Laher, S., (2010). Using exploratory factor analysis in personality research: Best-practice recommendations, SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 36(1)

Landy, F.J. (1986). Stamp collecting versus science; validation as hypothesis testing.
American Psychologist, *41* (11), 1183-1192.

Little, T.D., Cunningham, W.A., Shahar, G. & Widaman, K.F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151-173.

MacCallum, R.C. (1995). Model specification: procedures, strategies and related issues. In R.H. Hoyle (Ed.). *Structural equation modelling; concepts, issues and applications*. Thousend Oaks, California: Sage Publications.

MacCallum, R.C., Browne, M.V. and Sugawara, H.M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, *1* (2), 130-149

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

Mels, G. (2010). *Structural equation modelling with LISREL9 for Windows.* Chicago: Scientific Software International.

Messick, S. (1989). Validity. In R.L. Linn (Ed.). *Educational measurement (3rd ed.*, pp. 13-103). New York: Macmillan.

Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.

Mischel, W. (2004). Towards an integrative science of the person. *Annual Review of Psychology*, *55*, 1-22.

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

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

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

Mulaik, S.A., & Millsap, R.E. (2000). Doing the four-step right. *Structural Equation Modeling, 7*, 36-73.

Murphy, K.R., & Davidshofer, C.O. (2001). *Psychological testing: principles and applications (5th ed)*. Englewood Cliffs, New Jersey: Prentice Hall.

Murphy, K.R., & Davidshofer, C.O. (2005). *Psychological Testing: Principles and Applications (6th Edition)*, Upper Saddle River, N.J.: Pearson Education C2005.

Murray, H.A. (1938). *Explorations in personality*. New York: Oxford University Press.

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

Myburg, H.M. (2013). *The development and evaluations of a generic individual non-managerial performance measure*. Unpublished master's thesis. Stellenbosch: Stellenbosch University, Stellenbosch.

Neuman, W.L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches 4th Edition*. Allyn & Bacon. Needham Heights, MA.

Nunnally, J.C. (1978). *Psychometric Theory (2nd ed.)*. New York: McGraw-Hill.

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

Pervin, L.A., Cervone, D., & John, O.P. (2005). *Personality: Theory and Research (9th ed)*. New York: John Wiley & Sons.

Plumbley, P.R. (1985). *Recruitment and Selection (4th ed.)*. London: Institute of Personnel Management.

Popper, K.R. (1972). *Conjectures and refutations: the growth of scientific knowledge*. London: Routledge and Paul.

Psychology Dictionary – *Dictionary of Psychology Terms*. Retrieved March 20, 2012: http://dictionary-psychology.com/index.php?a=term&d=Dictionary+of+psychology&t=Industrial+psychology)

Republic of South Africa, (1998). *Employment Equity Act*. *Government Gazette*, 400(19370), Cape Town, 19 October.

Sanz, J., Gil, F., Barrasa, A., & García-Vera, M.P. (2006). Self-assessment of needs and behaviour patterns at work: Psychometric properties of the Personality and Preference Inventory-Normative (PAPI-N). *Personality and Individual Differences, 41*, 837-847.

Sass, D.A., & Smith, P.L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modelling. *Structural Equation Modeling, 13* (4), 566-586.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Sixteen Personality Factors. (n.d.). In *Wikipedia, the free encyclopedia*. Retrieved June 22, 2012, from http://en.wikipedia.org/wiki/16_Personality_Factors.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman and Hall/ CRC.

Smit, G.J. (1996). *Psigometrika; aspekte van toetsgebruik*. Pretoria: HAUM Uitgewers.

Smith, M., & Smith, P. (2005). *Testing people at work: competencies in psychometric testing*. Malden, MA: BPS Blackwell, 2005.

Society for Industrial and Organisational Psychology (SIOP). (2003).*Principles for the Validations and Use of Personnel Selection Procedures (4th ed.)*. College Park, MD: Author

Spangenberg, H.H., & Theron, C.C. (2004). Development of a performance measurement questionnaire for assessing organisational work unit effectiveness. *South African Journal of Psychology, 30* (1), 19-28.

Spangenberg, H.H., & Theron, C.C. (2005). *Promoting ethical follower behaviour through leadership ethics: The development of the ethical leadership inventory (ELI). South African Journal of Business Management, 36* (2)*,* 1-17.

Steiger, J. H. (1990). *Structural model evaluation and modification: an interval estimation approach*. Multivariate Behavioural Research, 25, 173-180.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using Multivariate statistics (2nd ed.).* New York: Harper Collins Publishers.

Tabachnick, B.G., & Fidell, L.S. (1996). *Using Multivariate statistics (3rd ed.).* New York: Harper Collins.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using Multivariate Statistic s(4th ed.)*. Boston: Allyn and Bacon.

Theron, C.C. (2007). Confessions, scapegoats and flying pigs: psychometric testing and the law. *SA Journal of Industrial Psychology*, *33* (1), 102-117.

Theron, C.C. (2009a). *Research Methodology. Unpublished class notes Industrial Psychology 778*. University of Stellenbosch.

Theron, C.C. (2009b). *Psychometrics. Unpublished class notes (Industrial Psychology 778).* University of Stellenbosch.

Theron, C.C., & Spangenberg, H.H. (2005). Towards a comprehensive leadership-unit performance structural model: The development of second-order factors for the Leadership Behaviour Inventory (LBI). *Management Dynamics, 14* (1), 35-50.

Twigge, L. (2003). *A psychometric investigation into the use of an adaptation of the Ghiselli Predictability Index in personnel selection*. Unpublished Thesis. University of Stellenbosch, Stellenbosch.

Ullman, J. (1996). Structural equation modelling. In M.C.G Davidson, (2000). *Organisational climate and its influence upon performance: A study of Australian hotels in South East Questions*. Unpublished doctoral thesis, Griffith University.

Van der Merwe, G. (2005). *The construct validity of the first view for selection purposes in South Africa*. Unpublished master's thesis. University of Pretoria, Pretoria, South Africa.

Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, *81*, 525-531.

Van Zyl, C.J.J., & Taylor, N. (2012). Evaluating the MBTI® Form M in a South African context. *SA Journal of Industrial Psychology, 38* (1), 1-15

Walter, V. (2002). *16PF personal career development profile: Technical and interpretative manual. Champaign*: Institute for Personality and Ability Testing.

Winter, D.G., John, O.P., Steward, A.J., Klohnen, E.C., & Duncan, L.E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review, 105*, 230-250.