A MULTI-GROUP STRUCTURAL EQUATION MODELLING INVESTIGATION OF THE MEASUREMENT INVARIANCE OF THE CAMPBELL INTEREST AND SKILL SURVEY (CISS) ACROSS GENDER GROUPS IN SOUTH AFRICA

by

CLAYTON DONNELLY

Thesis presented in partial fulfilment of the requirements for the degree:

MAGISTER COMMERCII (INDUSTRIAL PSYCHOLOGY)

AT THE

UNIVERSITY OF STELLENBOSCH

SUPERVISOR: PROF CC THERON

CO-SUPERVISOR: DR G EKERMANS

DECEMBER 2009

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: ________________ Date: 31/08/2009

ABSTRACT

Clayton Donnelly, MComm (University of Stellenbosch)

Supervisors: Prof. CC Theron & Dr. Gina Ekermans

The choice of career path could create a stressful situation for many individuals. Researchers seem to agree that if a person is able to find *fit* between what they would like to do and what a job (work environment) involves then a person is likely to perform their chosen occupation well. Interest assessment is a method that assists in making personal and organisational career related decisions. The Campbell Interest and Skill Survey (CISS, Campbell, Hyne & Nilsen, 1992) is a well-known interest assessment instrument that can be used for such decisions. Even though interest assessment can assist, these instruments have been criticised for being gender biased and typically forcing people into stereotypical gendered type occupations. Bias is indicated as nuisance factors that threaten the validity of cross-group (cultural) comparisons (Van de Vijver & Leung, 1997). These nuisance factors could be due to construct bias, method bias and/or item bias. Therefore, due to the importance of the decisions made, it would seem essential that the information provided by test results apply equally across different reference groups – this would imply equivalent measurement. Equivalence is achieved at three levels: Configural, metric and scalar (Vandenberg & Lance, 2000; Vandenberg, 2002). Full measurement invariance (achieved when scalar invariance is found) implies the ability to compare observed scores directly. By making use of confirmatory factor analytic techniques suggested by Vandenberg and Lance (2000), increasing constraints of equivalence were proposed for the CISS measurement model. While adequate model fit was found for the CISS Basic scales, the sample size did not afford independent gender sample confirmatory factor analyses (CFAs) and consequent measurement invariance tests to be conducted on the Basic scales. The CISS Orientation scales were then subjected to CFA on the combined gender sample and then were subjected to independent CFAs on the separate gender samples. Unfortunately poor model fit was found at this global level of measurement in the CISS. This prevented the researcher from completing the necessary measurement invariance tests on the Orientation scales for the CISS. The implications of the results are discussed, limitations are indicated and areas for further research are highlighted.

OPSOMMING

Clayton Donnelly, MComm (Universiteit van Stellenbosch)

Studieleiers: Prof. CC Theron & Dr. G Ekermans

Die maak van ‘n loopbaankeuse kan spanning veroorsaak in baie mense. Dit wil voorkom of navorsers saamstem dat indien ‘n person se werklike beroep ooreenstem met dit wat hy/sy graag sou wou doen – dan sal die persoon waarskynlik goed presteer in die gekose beroep. Die benutting van belangstellingsvraelyste kan individue help om effektiewe persoonlike en beroepsgerigte keuses te maak. Die “Campbell Interest and Skill Survey” (CISS, Campbell, Hyne & Nilsen, 1992) is ‘n bekende belangstellingsvraelys wat gebruik kan word om ondersteuning te bied om bogenoemde keuses te maak. Alhoewel belangstellingsvraelyste oor die algemeen waardevolle hulpbronne is in die maak van beroepskeuses, is hierdie vraelyste al gekritiseer dat hulle sydig kan wees op grond van geslag en as sulks mense kan lei om geslagsgetipeerde beroepskeuses te maak. “Sydigheid” in toetse kan beskryf word as “lastige” faktore wat die geldigheid van kruis-kulturele vergelykings bedreig (Van de Vijver & Leung, 1997). Hierdie faktore kan veroorsaak word deur konstruksydigheid, metodesydigheid en/of itemsydigheid. Dit is dus noodsaaklik dat die informasie wat verskaf word deur die toetsresultate dieselfde betekenis moet hê oor al die verskillende verwysingsgroepe en dit noodsaak ekwivalente meting. Ekwivalensie kan bereik word op drie vlakke: konfiguraal, metries en skalêr (Vandenberg & Lance, 2000; Vandenberg, 2002). Volle invariansie van meting (wat bereik word wanneer skalêre invariansie bevind word) impliseer dat waargenome metings direk met mekaar vergelyk kan word. Deur gebruik te maak van bevestigende faktoranalitiese tegnieke voorgestel deur Vandenberg en Lance (2000), is toenemende ekwivalensiebeperkinge voorgestel vir die “CISS” metingsmodel. Alhoewel ’n bevredigende passing gevind is vir die “CISS Basic scales” model, het die grootte van die steekproef nie toegelaat dat die “CISS Basic scales” model onafhanklik op die twee geslagsgroepe gepas word nie en ook nie toegelaat dat die metingsinvariansie van die model oor die twee geslagsgroepe ondersoek word nie. Die “CISS Orientation scales” is toe blootgestel aan bevestigende faktorontleding op die gekombineerde geslagsteekproef en asook op die onderskeie geslagsgroepe. Op hierdie globale vlak kon daar egter nie bevredigende modelpassing gevind word nie. Die gebrekkige modelpassing het gevolglik

die navorser verhoed om enige verdere metingsvariansie toetse op die “Orientation scales” te doen. Die implikasies van die resultate word bespreek, beperkinge word aangedui en verdere moonlike navorsingsgebiede word uitgelig.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank those who were instrumental in making this dream come true.

Firstly, I would like to show my admiration and appreciation to both of my supervisors, Prof. Callie Theron and Dr. Gina Ekermans. Their unwavering support, empowering style and sense of humour have been vital in the completion of this study. Both of these individuals are scientists with true integrity and I am privileged to have worked with them.

I would also like to thank Dr. Nicola Taylor (Jopie van Rooyen & Partners) for authorizing and supporting this research. Her ongoing guidance, academic inspiration and comradery have been superb.

Veronika Macher deserves thanks for her assistance with a **vital** aspect of this study: data capturing. Amanda Orfanos, Kim Budge and Gillian Schultz are thanked for their assistance with proof reading and guidance on improving the quality of the manuscript. Besides for being very special people in my life, they have contributed enormously.

My colleagues at Psytech SA/International are thanked for their support during this important professional milestone. Nanette Tredoux has been a crucial person in shaping my professional growth "Thank you!"

Baie dankie aan Retha Kok. Her assistance with language translation (and sense of humour) was a great help.

I would also like to acknowledge the late Dr. Conrad Schmidt for inspiring me to become a psychologist. Thankfully, he was also able to open my mind to the world of psychometrics. I hope I will be able to inspire others as he did me.

Last but definitely not least, I would like to thank my family for being my loving support (as always) with this qualification. I am truly blessed to have a warm and caring family. I would also like to thank Joy Dijkman for walking this journey with me. I could not have asked for a better person to motivate me, and teach me discipline and perseverance – thank you.

This material is based upon work supported financially by the National Research Foundation (NRF). Opinions expressed in this thesis and conclusions arrived at are those of the author and are not to be attributed to the NRF.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1
INTRODUCTION AND OBJECTIVE OF THE STUDY

"The secret of success is making your vocation your vacation"
Mark Twain

This chapter aims to provide a systematic reasoned argument in terms of which the objective of the research can be justified. The chapter in essence argues that interest assessment plays an important role in ensuring that individuals are satisfied with their chosen careers and in ensuring that organisations are satisfied with the level of training and work performance their employees demonstrate. The chapter argues that lack of measurement equivalence could complicate the interpretation and use of interest assessments across gender groups and thereby impede the abovementioned objectives.

1.1 INTRODUCTION

"What are you going to be when you grow up?" This enigmatic question rears its head many times during a young person's life. It is a question that some can answer almost immediately, while creating feelings of worry in others. The choice of career path could create a stressful situation for many individuals. Lowman (1991), amongst others, indicates that if a person is able to find *fit* between what they would like to do and what a job involves then a person is likely to perform the job well. Nakamura and Csikszentmihalyi (2005, p. 89) declare that "a good life is one that is characterised by complete absorption in what one does". Yet, all of these statements are based on the presumption that the person has sufficient understanding of their career needs, wants and motives.

A way of assisting the individual in making career choices is to help the person discover what they would like to do, or what *interests* them most. Interest assessment is a method that assists in achieving these goals: life satisfaction and vocational productivity (Gregory, 2004). It also follows that decisions made on the basis of interest information will have a substantial impact not only on individuals but also organisations. For organisations, life satisfaction tends to have a positive effect on attitude and boost productivity at work. In terms of vocational productivity, the impact of individuals mismatched to work that they may be uninterested in what they do; this has a motivational effect which then negatively impacts on productivity. If work is interesting, personal experience of fulfilment is realised (Gregory, 2004).

This fundamental role that interests play in behaviour has major implications for the field of psychology. Interest information is essential for self-awareness (Greenhaus, Callanan & Godshalk, 2000) for the individual that wishes to make decisions relating to work. Research has been conducted into the effect of self-awareness on career decision making and associated outcomes (for example: Singh & Greenhaus, 2004; Sauermann, 2005). The research findings do indicate that higher levels of self-awareness as well as career self-efficacy can lead to suitable career choices.

Career seekers could make use of the Industrial/Organisational (I/O) psychologist's services for the purpose of increased career self-awareness and resultant career self-efficacy. I/O psychologists can help in understanding, measuring and predicting career development, occupational choices and work adjustment. The I/O psychologist is likely to make use of ability tests, personality questionnaires and interest inventories to form a full picture of the individual's skills, attitudes, interests and motivations (Lowman, 1991; Robitschek, 2004).

From an organisational perspective the I/O psychologist would be in a position to assist companies in making appropriate selection and development decisions. Once again, psychometric assessment of individuals applying for particular roles seems to be a possible way to ensure a decision that simultaneously optimises individual and organisational criteria. While general ability seems to demonstrate the highest levels of predictive validity across many occupational levels (Schmidt & Hunter, 1998), other influences, for example fit between interests and the job, could play a role in enabling a person to use their abilities. In the meta-analytic research conducted by Schmidt and Hunter (1998) it should be noted that while interest assessment yields a validity coefficient of 0.10 for selection decisions, it was found that interest assessment was a slightly stronger predictor of training performance (r=0.18). This only provides rather tenuous support of the hypothesis that those interested in a topic are likely to enjoy it and do well.

Considerable research has been conducted in the person-environment (P-E) and person-job (P-J) fit arenas (Spokane, Meir & Catalano, 2000). Hesketh (2000) indicates that when examining any fit theory four components are generally investigated, namely: (i) measurement of the person on a relevant array of dimensions covering both the competency and the motivational components (knowledge, skills, abilities, values, needs and interests), (ii) measurement of the environment on a commensurate relevant array of dimensions, (iii) measurement of an outcome and finally (iv) assessment of fit between the environment and organisational outcomes (as a function of the person's capabilities).

Therefore, when considering the linear-based findings of the Schmidt and Hunter (1998) article it would seem that the interactions as indicated by Hesketh (2000) have not been considered. This means that even though interest assessment seems to play a small role in performance, the actual role of interests and degrees of fit as determinants of performance have not been considered. Career interests may well be a moderator/mediator to true reflections of cognitive ability through a motivational relationship. Although the Schmidt and Hunter (1998) article indicate that cognitive ability may be the best predictor of job performance, the effect of ability on performance may be moderated by the individual's *interest* in the work. This moderating effect should be empirically examined in a further study.

Much of the interest assessment literature on P-E fit has focussed on the Holland (1985) RIASEC model of occupational interest. This model purports that interests can be measured on six categories of interest and a final three letter code indicates a "type" associated with a number of occupations. The majority of fit research tends to focus on the match between this three letter code and occupation. However, as pointed out by others (Fritzsche, Powell & Hoffman, 1999; Hesketh, 2000; Spokane et al., 2000), the research models do not take the full breadth of an individual's interest into consideration as they make use of only three areas of interest whereas the full six areas should be considered. Many other instruments have been developed using the Holland model as a basis for a further model, for example: the Campbell Interest and Skill Survey.

Even though it seems as if further research into the manner in which interest is structurally related to job performance is needed there does seem to be only a slightly stronger linkage, as per the Schmidt and Hunter (1998) article, between training performance and interest assessment. Nonetheless it would still seem justifiable to argue that interest assessment would be beneficial not only to the individual attending training programmes. Interest assessment would also mean that organisations are then in a position to align interests with preferred work tasks in training, thereby allowing for increased motivation and skills development. The skills shortage and the hunt for talent are cited frequently as contemporary human resource problems. Therefore, if existing employees could discover their personal work interests then the organisation would be in a position to capitalise on increased motivation in training and consequent skill improvement of current employees through judicious career planning and development. This would then aid in decreasing the emphasis on the external search for skills and could improve motivation. To discount interest inventories in organisational decision making might limit the approaches to training and development within organisations.

The history of interest assessment began in the early part of the 20th century and interest assessment has remained in use ever since (Kaplan & Saccuzzo, 2001). From very early on many interest questionnaires were developed, reflecting the assumed pivotal role of interests in career success. Major contributors in the interest assessment arena include: Kuder, Strong, Holland and Campbell (Campbell, 1995; Donnay, 1997; Holland, 1959; Kaplan & Saccuzzo, 2001). Each of these individuals contributed to making interest assessment what it is today. However, early interest inventories were based on typically male orientated roles, purely because men dominated the world of work. As a result many interest inventories where written with men in mind. This became problematic later on with many women's rights movements condemning interest inventories as being gender biased and typically forcing people into gendered type occupations (Campbell & Hansen, 1981; Kaplan & Saccuzzo, 2001; Murphy & Davidshofer, 2005).

The contentiousness of gender stereotyping in interest assessment can seriously jeopardize the objectives of career counselling/management and related decisions. The far-reaching consequences that the use of psychological assessment in decision-making could have on individuals and organisations has been indicated. Therefore, due to the importance of the decisions made, it would seem essential that the information provided by test results apply equally across different reference groups. The measurement models that underlie each test must be transferable across groups or the test is ultimately testing different latent variables (interest dimensions) across different groups – decision making is then based on two separate measurement models. Equivalent numbers of interest factors as well as equivalent factor loadings (configural invariance, Vandenberg & Lance, 2000) - although a necessary requirement - is however, not a sufficient condition to ensure that observed interest scores mean the same thing in terms of the underlying latent variable across gender groups. Even though the number of latent interest dimensions might be the same and the pattern of factor loadings might be the same across gender groups, the magnitude of measurement model parameters could still differ across gender groups and thereby affect observed score interpretation. To be able to confidently interpret observed score differences between genders as indicative of latent score differences, full measurement invariance needs to be indicated.

The measurement invariance issues associated with gender stereotypes could be termed bias. Van de Vijver and Leung (1997) describe *bias* as a generic term used for all the nuisance factors threatening the validity of cross-group (cultural) comparisons. These nuisance factors could be due to the construct being measured by the instrument not being identical across groups (construct bias). Bias arising from particular characteristics of the instrument or its associated administration (method bias) could be considered nuisance factors. Finally, item bias refers to undesirable measurement artefacts at the item (content coverage, inappropriate wording, ambiguities, idioms, comprehensibility etc) level.

Van de Vijver and Leung (1997) indicate that equivalence is the absence of bias. With the absence of bias the psychologist is then more confident about the validity of results and comparisons can be made between groups based on questionnaire/test results. Psychologists would want to compare candidates for selection or attendance at training programmes. However, without equivalence decisions would be based on comparing apples with pears. This would not be deemed appropriate when wanting to offer all an equal opportunity to present their talents to the organisation/occupation[1]. Van de Vijver and Leung (1997) indicate different hierarchical levels of equivalence that must be met

[1] Full measurement invariance, however, is no guarantee that discrimination in criterion-referenced selection cannot occur. Even though the latent predictor variable is measured without bias it should still, in principle, be possible that predictive bias could exist in the criterion inferences derived from the unbiased predictor measures. Predictive bias exists if the regression of the criterion on the predictor differs in terms of slope and/or intercept across protected and non-protected groups and this difference is not taken into account when deriving criterion estimates. This can easily happen even though no scale bias exists. This seems important since it would suggest that even if the Employment Equity Act (Republic of South Africa, 1998) would be successful in eradicating all forms of measurement bias it would thereby still not have succeeded in ensuring that selection decisions do not disadvantage members of specific groups (Theron, 2007).

prior to making direct comparisons between different groups. Only when meeting all the levels of equivalence can differences in mean observed scores of gender groups be compared, and interpreted to reflect true difference in the underlying latent variable.

This research study aims to address the issue of measurement equivalence across gender groups in career interest assessment. As previously discussed, appropriate career assessment impacts on the adjustment of individuals and is likely to affect organisations. Historically only men assumed the role of career holder in the home, but with the emancipation of woman this is certainly no longer the case. Outdated models of career assessment would also be deemed inappropriate in the current context. However, it should be stated that this study does not aim to investigate gender definitions of interests and resultant bias effects. The study purely aims to evaluate the gender equivalence of a well-known interest questionnaire (Gregory, 2004), namely the Campbell Interest and Skill Survey (CISS, Campbell, Hyne & Nilsen, 1992). Unlike previous measures of interest where separate forms were used for different gender groups, the CISS transcends gender archetypes and one single form is utilised.

The CISS attaches a specific connotative definition (Kerlinger & Lee, 2000) to the interest latent variable. Specific latent interest dimensions are distinguished in terms of this conceptualization. Specific items have been designed to serve as effect indicators (Hair, Black, Babin, Anderson & Tatham, 2006) of these latent interest dimensions. This design intention is reflected in the scoring key of the CISS. A very specific measurement model is moreover implied by the design intentions (and the scoring key) of the developers of the CISS. A critical question is whether the measurement model reflecting the design intentions of the developers fits data obtained from the instrument at least reasonably well. Evidence on the psychometric integrity of the instrument is reported in the test manual (Campbell et al., 1992). The validity and reliability analysis results reported in the manual, however, all originate from studies performed outside of South Africa. No South African studies that evaluated the reliability and construct validity of the CISS could be traced in the literature. Moreover, none of the studies on the psychometric integrity of the CISS evaluated the fit, through confirmatory factory analytic procedures, of the measurement model implied by the design intentions of the developers.

The current study will investigate the fit of the CISS measurement model on a gender diverse sample of South African respondents. If reasonable measurement model fit, along with significant ($p<0.05$) and reasonably high completely standardized factor loadings [at least 0.71 or higher (Hair et al., 2006)] would be found, that would permit the within gender group use of the CISS to measure the interest construct as constitutively defined. Cross-gender group comparisons would, however, thereby not be sanctioned. A further critical question is whether the measurement model parameters are the same across these groups in South Africa. Does the South African data, available for this study, fit the measurement model equivalently across the gender groups?

In order to answer these questions, the measurement model reflecting the design intentions of the CISS would need to be fitted simultaneously to both gender groups in a multi-group analysis in which model parameters are allowed to vary freely across groups and in which model parameters are constrained to be equal across groups. While Van de Vijver and Leung (1997) make use of an exploratory factor analytic approach to study measurement equivalence that is essentially a data-driven procedure, the current study favours a procedure that tests for measurement equivalence/invariance through a confirmatory factor analytic approach which allows for specific hypotheses to be tested.

The general question of invariance of measurement is whether or not, under different conditions of observing and studying phenomena, measurements yield comparable measures of the same attributes (Horn & McArdle, 1992, as cited in Vandenberg & Lance, 2000). This technique allows for an evaluation of model fit in one group versus another. This is particularly useful in determining whether the test measures the same latent variables across groups. By placing increasing constraints on the measurement model as specified by the test publisher, the researcher aims to identify which model parameters would be a likely antecedent of non-equivalence if applicable.

In order to conduct research on the CISS the local questionnaire sole distributor provided data on the instrument that included both genders in a format that ensures the anonymity of respondents. Written permission had been obtained from the South African distributor of the CISS to use the data for the purpose of the envisaged research. The test distributor welcomed the research, and indicated that research of this nature had not been conducted in South Africa (N. Taylor, personal communication, 11 August, 2007) on the questionnaire and expressed the opinion that the research would contribute towards the gender unbiased use of the CISS.

1.2 RESEARCH OBJECTIVE

The objective of this research study is to determine whether the CISS can be used in South Africa to derive valid inferences on the interest latent variable as it is constitutively defined by the test manual (Campbell et al., 1992) and whether the same latent interest dimension inference may be derived when the same observed scores are obtained on the instrument for matched male and female respondents. The objective of the research is to evaluate the fit of the measurement model of the CISS on a South African sample via confirmatory factor analysis (CFA) and to determine, if adequate model fit would be obtained on the total sample, whether significant differences in measurement model parameters exist between male and female subsamples.

1.3 OUTLINE OF THE STRUCTURE OF THE THESIS

The structure of this research document includes an introduction to the history, development and use of interest measurement. A discussion regarding the CISS will follow. A review of the measurement invariance literature and resultant research objectives are discussed in chapter 4. Thereafter, the proposed research methodology is described along with preliminary data analyses. A chapter on model fit and measurement invariance tests follows with a concluding chapter providing a discussion of results with limitations explored and recommendations made.

CHAPTER 2
LITERATURE REVIEW OF INTEREST ASSESSMENT

This chapter aims to provide a literature survey of the history, development and utility of the interest assessment. The chapter will also review the existing literature with regards to gender bias in interest assessment and its impact on inferences made from the measurement of interests.

2.1 HISTORY AND DEVELOPMENT OF INTEREST ASSESSMENT

Career counselling is broadly referred to as a process where a career counsellor or psychologist assists an individual in making a career-related decision with the overarching objective of ensuring a satisfying career path (de Bruin, 2001). A number of approaches can be taken to assist an individual in a life changing event such as career choice. These include a trait-and-factor approach, a developmental approach and postmodern approaches (de Bruin, 2001).

The focus of this review will be on the trait-and-factor approach. This approach is guided by the following principles, as identified by Parsons (1909, as cited by de Bruin, 2001): (a) the individual must know himself or herself, (b) the individual must know the world of work, and (c) the individual must find a fit between his or her characteristics and the world of work. This approach would posit that individuals have inherent preferences and these would determine the degree of fit between an individual's interest preferences and career route.

In order to gather information that allows thorough decision making, assessment of the individual's preferences, attitudes, interests, abilities and personality would be helpful. However prior to any ability-related assessment an understanding of what the client likes or dislikes in terms of careers is essential. The interest inventory is the key to this exploration. MacAleese (1984, as cited by de Bruin, 2001) suggests that interest inventories have three purposes, namely: (a) to identify interests that the client may not be aware of, (b) to confirm what the client is actually interested in and (c) to examine differences between the client's interest and their actual abilities or skills. Therefore, interest inventories are a crucial link in the career interest and occupational choice.

The first interest inventory was introduced in 1912 and named the Carnegie Interest Inventory (Kaplan & Saccuzzo, 2001). After a period of developments in the interest assessment arena, the publication of the first Mental Measurements Yearbook (1939) indicated that there were already 15 different interest measures in use at that stage. Of these instruments the most used included the Strong Vocational Interest Blank (SVIB) (1927) and the Kuder Preference Survey (1939) (Kaplan & Saccuzzo, 2001). Today there are many additional interest inventories, however the now evolved Strong Interest Inventory (SII) remains the most used interest inventory (Hansen & Campbell, 1985, Walsh & Betz, 1995, Zytowski & Warman, 1982; as cited in Kaplan & Saccuzzo, 2001).

Strong (1927) believed that interests are on a dimension of like versus dislike and this information could be used to predict the preferences that individuals would show for various occupational activities. Therefore, an analysis of the interests held by groups contrasted in terms of their preference for a specific occupation could assist in identifying relationships between interests and preferences for and satisfaction with occupations. This became the preferred empirical method to uncover the interest profile that differentiates those with a preference for a specific occupation from those that do not share the liking for the occupation (Donnay, 1997). The contrasted-groups approach resulted in various scales measuring a variety of interest items that best discriminated specific occupations from people in general. This approach was subsequently adopted by Hathaway and McKinley (1940) in the development of the Minnesota Multiphasic Personality Inventory (MMPI) – an aid in psychiatric diagnosis (Hathaway & McKinley, 1940, as cited in Gregory, 2004).

Kuder (1977a, as cited by Donnay, 1997) affirmed Strong's contributions to the field of interest measurement and suggested that "…it is almost impossible to do anything in the field that does not have some background in Strong's work." Nonetheless, as indicated previously, the Kuder Preference Survey was a prominent instrument in the pioneering days of interest measurement. Kuder originally opted for an ipsative approach to interest assessment, but later moved to assessing respondent preferences as compared to the preferences of specific occupational groups. Kuder made use of Cleman's lambda to statistically measure the individual's similarity to the occupational reference group (Dawis, 1991, Zytowski & Borgen, 1983, as cited in Donnay 1997). This approach allowed for the measurement of similarity with both the general and the unique interests of each occupational group, whereas Strong's contrasted-groups approach only measured similarity to relatively specific interests as endorsed by the occupational group as being important to the group. The two approaches, Strong's criterion-related measurement and Kuder's content-related measurement, are the founding stones of current vocational interest measurement (Donnay, 1997).

As Strong continues to play a dominant role in the field of interest measurement a further discussion on its development is warranted. The original version of the SVIB (Strong, 1927) contained 10 Occupational scales. The scales were constructed by comparing the interests of individuals with the interests of *men* employed in specific occupations (Donnay, 1997). This method of contrasted groups was initially used with professional men and then later extended to the interests of woman – this resulted in a women's version or form in 1933. The SVIB went through several revisions: the men's form in 1938 (Strong, 1938, as cited by Donnay, 1997) and 1966 (Campbell, 1966b, as cited by Donnay, 1997), and the women's form in 1946 (Strong, 1946, as cited by Donnay, 1997) and 1969 (Campbell, 1969, as cited by Donnay, 1997). These revisions expanded on the original Occupational scales that were constructed for the instrument. The revisions made during the 1960s include the further expansion of the Occupational scales as well as the addition of the Basic Interest scales. In the most recent version of the Strong Interest Inventory the number of Basic interests total 30 which are used to empirically derive scores on the 122 Occupational scales. The Basic scales were developed by investigating the intercorrelations of the SVIB items to determine clusters of activity that could be viewed as unique areas of interest (Donnay, 1997). This addition provided more detailed information beyond the broad occupational categories in the original version.

Reliability studies produced impressive results for the revised versions (1960s) of the SVIB. Test-retest correlation coefficients indicated values between the low 0.80s and the low 0.90s. Twenty year long test-retest reliability coefficients were in the 0.60s. In terms of validity, Strong placed much emphasis on the fact that interest inventories should be able to produce criterion-related evidence so that predictions can be confidently made in occupational choice (Donnay, 1997). Studies conducted looked at both concurrent and predictive validity. Strong (1935, as cited in Donnay, 1997) concluded that occupational choice (concurrent studies) and occupation engaged/satisfaction (predictive studies) are the key to determining the value of the instrument. Donnay (1997, p. 9) reports:

> Overall, the predictive accuracy of the heterogeneous Occupational scales on the Strong is well established, with follow-up studies ranging from 3 to 18 years in length reporting direct and indirect good hits at the rate of 32% to 69%. In all studies reported, the rate of good hits was well above chance.

Even though the SVIB attained great acceptance and use, some concerns began to come to the fore during the 1960s and early 1970s. Critics believed there to be a gender bias in the scales as separate tests were used for men and woman. Others felt that the test lacked some theoretical basis (Kaplan & Saccuzzo, 2001).

During the late 1950s Campbell, the developer of the CISS, began his career in interest assessment at the University of Minnesota. Due to a number of staff movements and Strong's illness at the time, Campbell become closely involved in the research and development of the SVIB. Due to some of the limitations of the SVIB, Campbell (1974) developed and published the Strong-Campbell Interest Inventory (SCII) (Donnay, 1997). In this version of the instrument, Campbell expanded the number of Occupational scales to 124, maintained the 23 Basic Interest scales and added 2 Special scales to measure academic comfort and introversion-extroversion[2]. The new revision included the development of a single-sex form as opposed to the previous version (Campbell, 1995).

The most significant addition made in the SCII's construction was the inclusion of Holland's (1959) hexagonal model of interest – Holland viewed interests as structured in a model analogous to personality "type". The six areas/dimensions that define the interest profile are: (a) realistic (e.g. mechanics, agriculture and sport), (b) investigative (e.g. science and scholarly pursuits), (c) artistic (e.g. visual and culinary arts, creative writing and drama), (d) social (e.g. teaching, counselling and other helping professions), (e) enterprising (e.g. selling products, services or ideas) and (f) conventional (e.g. typing, filing and accounting). The person's interest profile is determined by measuring the levels of interest on each of the six dimensions. Generally, the combination of the three highest scores provides an indication of occupation (Robitschek, 2003). The interest dimensions comprising the model are generally measured using another pioneer in interest

[2] The introversion-extroversion assessment was included so to determine an individual's interest in working independently versus with fairly continuous people contact. This could provide valuable information when wanting to narrow occupational choices in terms of personality preference.

assessment the Self Directed Search (SDS; Holland, 1985). In addition to Strong's work, the SDS is one of the best known interest inventories in the world (de Bruin, 2001). Therefore, in addition to the scales added to the SVIB, as well as the single-gender measurement model, Campbell was able to add a credible theoretical basis which the SVIB lacked (Kaplan & Saccuzzo, 2001).

During 1981 and 1985, the SCII was modified in a number of significant ways, most due to Hansen (Campbell & Hansen, 1981; Hansen & Campbell, 1985). The improvements were: (a) a more balanced gender composition of the general reference sample, (b) an expansion of the profile coverage to include more blue-collar occupations, (c) a concerted effort to provide both female and male scales for almost all of the occupations, and (d) an increase in the average size of the occupational samples (Campbell & Hansen, 1981; Hansen & Campbell, 1985).

From 1983 until 1988 a legal battle ensued between Campbell and Stanford University Press regarding the intellectual property rights of the SCII. Essentially the outcome of the legal confrontation was that Stanford gained all rights to the inventory and renamed it the Strong Interest Inventory (SII). However, Campbell retained the rights to use his name for the purposes of instrument development as it was no longer tied to the SII (Campbell, 1995).

The discussion regarding the SII ends at this point as this is the crucial junction where the CISS began its development, and due to the focus of this research it would seem more appropriate to continue the discussion with a focus on gender-related issues in interest measurement prior to leading into a discussion regarding the development of the CISS.

2.2 GENDER ISSUES IN INTEREST ASSESSMENT

Interest inventories and assessment has had its fair share of controversy. This is particularly true for advocates of women's rights whose thoughts on specific discriminating components in interest assessment were voiced publically and were believed to be valid (Brik, 1974, Campbell, 1995, Diamond, 1979, Peoples, 1975, Tittle, 1983, as cited in Kaplan & Saccuzzo, 2001; Watkins & Campbell, 2000).

The Commission on Sex Bias in Measurement (1973) concluded that interest inventories contributed to the policy of guiding young men and woman into gender-typed careers. The interest inventories tended to direct women into their traditional roles, for example: nursing, clerical service and primary school teaching. As discussed previously, the SVIB had two separate forms for men and women. According to the commission's report, careers on the women's form tended to be lower in status and generally commanded lower salaries (Harmon, Cole, Wysong & Zytowski, 1973; as cited in Kaplan & Saccuzzo, 2001).

Researchers that aimed to handle this disparity in the field of interest measurement felt that separate but equal inventories for the genders was the best approach. However, this solution was problematic as the inventories for men and woman were rarely equal (Murphy & Davidshofer, 2005). The SCII was revised to create a single form version. However, in the 1977 SCII manual, Campbell indicated that (Kaplan & Saccuzzo, 2001, p. 388):

> If Strong were alive, he may have felt that using the same norming tables for both men and woman would have harmed the validity of the test. A unisex interest inventory, according to Strong, ignores the social and statistical reality that men and woman have different interests. In other words, knowing the sex of the test taker tells us a lot about his or her interests.

Attempts have been made to reduce gender-related problems associated with interest assessment, but elimination does not seem probable. The SCII compares a respondent's Basic interest scores with those from a combined male/female reference group. The Occupational scales, however, are normed separately for men and woman. To simply compare the observed interest scores of both men and women with a combined group of men and woman seems unacceptable, given that research shows that "gender differences exist for between one quarter and one third of the items on the SII" (Harmon, Hansen, Borgen & Hammer, 1994, as cited in Murphy & Davidshofer, 2005, p. 362).

A strategy to solve the gender issue has been to develop both male and female criterion groups for each occupation – this has proven to be a popular approach. However, the difficulty with this approach is that in certain instances there are a few members of one gender in an occupation to form a representative sample which is both reliable and valid (Murphy & Davidshofer, 2005).

In the discussion of gender issues thus far, the proposed remedies have been to consolidate male and female assessment onto a single medium with the use of separate norms. A major concern is the specification of the measurement model across gender. The test publishers generally do not provide evidence of the equivalence of the underlying measurement model across the genders. Therefore, meaningful comparisons between gender groups of the type referred to earlier are not really possible. Evaluating the equivalence of the measurement models has to be the starting point prior to the actual comparison of the groups in each interest area. Without measurement model equivalence the same observed score obtained by a male and female respondent cannot be interpreted to reflect the same standing on the underlying latent interest dimension. The starting point in dissecting the gender bias issue prevalent in interest measurement would be to ascertain the equivalence of the measurement model of the chosen interest inventory across the genders before debating criterion reference groups or norms. Once measurement models are shown to be equivalent then robust comparisons can be made across the genders.

Establishing full measurement invariance for any specific interest questionnaire would allow the differences in observed scores to be interpreted to reflect corresponding differences in latent scores (metric invariance) and it would allow the observed score of equal magnitude to be interpreted as reflecting latent scores of equal magnitude (scalar invariance, Vandenberg & Lance, 2000; Hair et al., 2006).

Differences in latent means could exist between genders on specific latent interest dimensions. If full measurement invariance would exist, and if differences in mean interest dimension scores would be obtained between genders, the instrument should not be blamed for this. The instrument is simply accurately reflecting an existing fact. The question could be asked why differences between genders exist in specific interest fields and whether these differences in interest reflect an undesirable differentiation in the (culturally determined) socialisation of girls and boys. The answer to this question would reflect a value judgment.

How latent interest dimensions express themselves in behaviour could be seen as a function of culture. Whether the expression of specific latent interest dimensions would be sanctioned by society might be dependent on gender. This would result in gender differences in latent means and in observed means in the instruments displaying full measurement invariance. In addition, it might be possible that the denotations of a specific latent interest dimension might be different across genders due to culturally imposed gender stereotypes.

Whether interest questionnaires should measure the natural interest conceptions of a specific society is, however, debatable. In terms of psychological theory specific latent interest dimensions carrying specific constitutive definitions have been hypothesized (and have hopefully subsequently been shown) to be systematically related to specific criterion variables related to career success. The objective of Industrial/Organizational psychology is to measure those latent interest dimensions (carrying specific constitutive definitions) that have been shown to have utility in enhancing performance on the career criterion variables. If the focus would have been on the manner in which society naturally thinks about and conceptualizes interests and how this differs across the genders and cultures then it would have been more important to take cognisance of the connotative drift in the meaning of interests.

CHAPTER 3

REVIEW OF THE CAMPBELL INTEREST AND SKILL SURVEY (CISS)

The CISS is the focus of this chapter. A review of the history and development of the instrument is included and psychometric properties of the tool are elaborated. Previous research is also discussed.

3.1 HISTORY AND DEVELOPMENT OF THE CISS

In the previous section, the discussion focussed on the work of Strong in developing interest inventories. Strong created a legacy in the interest assessment arena – the SII is one of the most used interest inventories. With the combination of Holland's theory of occupational types, the instrument encompasses useful elements that would assist the psychologist and client in exploring career options.

The discussion on Strong's model and questionnaires ended with the introduction of Campbell's entrance and consequent revisions of the instrument. As mentioned, Campbell and Stanford University Press parted ways. However, Campbell had accumulated extensive knowledge years of test development. Later, Campbell then initiated the development of CISS (Campbell, 1995, p. 392).

The CISS was published in 1992 (Campbell, Hyne & Nilsen, 1992; Campbell, 1993; Campbell, 1995). The CISS expands on earlier work of Campbell by adopting a more appropriate item pool[3] and a more flexible six-point Likert-type item response format[4]. In addition to the changes to the item pool, a new skills measurement model was added in parallel to interests – the first of its kind at the time. The skills measurement reflects an individual's self-assessed level of skill in a variety of activities. The authors felt that this type of skills measurement, even though not objective, may provide some idea of the person's propensity for having a certain skills set which could be valuable in making career

[3] Titles such as: "salesman, policeman have been replaced with gender neutral titles such as sales person, police officer. "Participating in a manhunt" or "charming members of the opposite sex" and other subtle vocabulary traps have been avoided. The use of American slang has been removed so to avoid unfamiliarly with the item content, for example: "Can pitch-hit in a variety of functions." Modernity was of important, therefore older items such as "Read the Literary Digest" are avoided as publications can go out of date quickly. The use of proper nouns was also avoided as these can date, for example: names of leading figures in specific careers (Campbell et al., 1992).

[4] The authors felt that a normative style six-point Likert scale would be preferred by the majority of candidates as apposed to an ipsative approach (Campbell et al., 1992).

decisions (Campbell et al., 1992). All of the scales were standardized on the same population and both genders are scored and norm-referenced in the same way. Both genders are compared to one combined gender norm. It is an instrument that has the advantage of benefitting from years of test development knowledge and modern technology (Campbell, 1995).

The questionnaire focuses on careers that require tertiary education. It would therefore be most appropriate to assess the interest profiles of individuals intending to enrol at university, are currently at university or have completed university. The instrument is also useful when adults wish to make career transitions or wish to understand specific current job dissatisfactions (Boggs, 1999). The American reading level is at the sixth grade and the instrument is therefore, in terms of reading proficiency, appropriate for adults and adolescents aged fifteen and older. It has, however, been used at younger ages in exceptional circumstances (Campbell et al., 1992). An objective assessment of the South African reading level has not been conducted (N. Taylor, personal communication, 2009), however, the local test distributor recommends grade 12 English proficiency.

Individuals who are assessed on this instrument are required to evaluate their own levels of interest on 200 academic and occupationally oriented items (85 occupations, 43 school subjects and 72 activities). The questionnaire also requires that individuals assess their own level of skill in 120 items based on occupational activities (Campbell et al., 1992).

For ease of use and interpretation, patterns of interest and skill scores are reported on the profile as: (a) Pursue – areas that are worthy of serious consideration as the interest and skill scale scores are both high (55[5] or above); (b) Develop – seek additional training to increase self-confidence or accept as hobbies, because interest scores are high (55 or above) and skill scores are lower (54 or below); (c) Explore – gain an understanding of why the area is not more appealing or consider applying the skills to another field because interests are lower (54 or below) and skills are high (55 or above); and (d) Avoid – activities not to consider, because interests and skills are both low (45 or below). If both interest and skill scores are in a mid-range or one is in mid-range and the other is lower, no pattern is reported (Boggs, 1999, p. 169).

---

[5] All scores are reported as T scores.

3.2 MEASUREMENT MODEL

The CISS provides the following measurement scales for the individual taking the assessment: Basic Interest and Skill scales, Orientation scales, Occupational scales, Special scales and Procedural checks.

3.2.1 Basic scales

The Basic scales provided the foundation of the CISS development and measurement model (Campbell et al., 1992). Hence, the Basic scales reflect the first-order factors measured by the CISS. These Basic scales, in turn, load onto seven second-order/global factors measured by the Orientation scales.

During the early stages of the development of the CISS, different, non-matching sets of interest and skill scales had been developed, working from item intercorrelations. The construction of the respective scales was based on the approach taken with the Strong series of questionnaires; that is, by examining the item intercorrelations to determine clusters that could be representative of unique interest/skill areas. The finding, however, was that it is virtually impossible for respondents to understand the interest and skill scales as functioning with different first-order factor structures (Campbell et al., 1992).Therefore, the interest and skill measurement models are viewed as being measured in a parallel fashion.

Table 3.1 depicts the manner in which the first-order interest and skill factors load onto the second-order orientation factors and at the same time defines the Basic scales in terms of the core activities that denote the latent interest and skill dimensions measured by the Basic scales.

Parallel latent interest and skill dimensions are assumed to exist. The latent first-order factors listed in column two of Table 3.2, therefore, should be interpreted consecutively as latent interest dimensions and as latent skill (or confidence) dimensions. The activities listed next to each first-order factor should likewise be interpreted consecutively as activities one would like to perform and one would feel confident to perform if the latent interest and skill dimension would be strongly developed. In essence, therefore one could have presented Table 3.1 as Table 3.1a and as Table 3.1b with slightly more explicit column headings that reflect the fact that either interests or skills are being measured.

Table 3.1.
Summary of the Second-Order Factor Structure of the CISS

Orientation	Basic scale	Activities
Influencing	Leadership	Acquire resources, inspire others to high performance.
	Law / Politics	Debate issues, be politically active, negotiate.
	Public Speaking	Give interviews to the media, deliver speeches, conduct training.
	Sales	Make sales calls, persuade others to purchase goods or services.
	Advertising / Marketing	Develop marketing strategies, design advertising campaigns.
Organizing	Supervision	Manage others, plan budgets, schedule work.
	Financial Services	Coordinate financial planning, investments, study economics.
	Office Practices	Perform secretarial duties and handle schedules, supplies and files.
Helping	Adult Development	Teach new skills to adults, work with students.
	Counseling	Counsel, help, advise, support people.
	Child Development	Teach classes, play with children, tell stories.
	Religious Activities	Conduct religious programs and services.
	Medical Practice	Provide healthcare services and first aid.
Creating	Art/Design	Draw, create works of art, design room layouts.
	Performing Arts	Play music, act, sing, dance direct plays.
	Writing	Research topics, write and edit materials.
	International Activities	Travel, work overseas, speak foreign languages.
	Fashion	Design fashions, buy and sell clothes and jewellery.
	Culinary Arts	Prepare gourmet meals, manage a restaurant.
Analyzing	Mathematics	Write computer programmes, analyse data, teach mathematics.
	Science	Perform lab research, work with scientific concepts and equipment.
Producing	Mechanical Crafts	Work with cars, machines and electrical systems.
	Woodworking	Do carpentry, build furniture and decks.
	Farming / Forestry	Raise crops, manage timber, care for livestock.
	Plants / Gardens	Design, plant and care for gardens.
	Animal Care	Care for pets, raise and train animals.

Adventuring	Athletics / Physical Fitness	Exercise, coach, compete, stay fit.
	Military / Law Enforcement	Use military strategies in challenging or dangerous situations.
	Risks / Adventure	Engage in high-risk, exciting, physically strenuous activities.

Adapted from Campbell et al., 1992

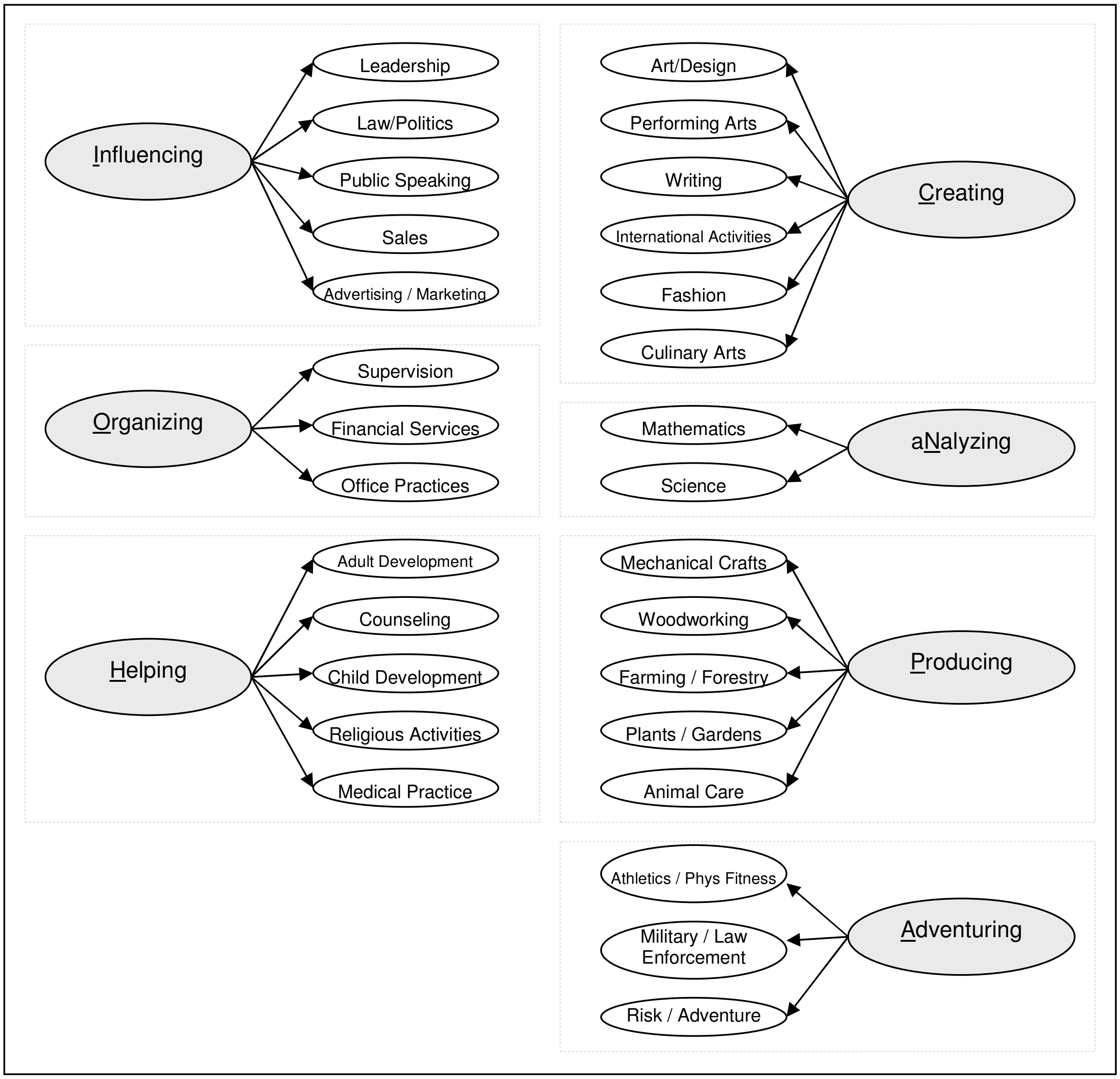

Figure 3.1. Path diagram of the second-order factor structure of the CISS.

The same activities are used to construct the interest and skill items that reflect the level of interest and the level of confidence in a specific activity area. Therefore, interest items load on the same specific Basic scales as do the corresponding skill items. Figure 3.1 provides a path diagram of the second-order factor structure of the CISS.

3.2.2 Orientation scales

The authors of the CISS constructed seven Orientation scales for both the interest and skill measurement. The Orientation scales serve as global factors which summarise the information gained through the primary Basic scales. The final CISS report presented to the client initially shows information at the global level, then proceeds to provide more detailed information at the primary level (Basic scales) and finally provides information derived from the Basic scales (Occupational and Special scales) (Campbell et al., 1992).

Although the interest orientations measured by the Orientation scales constitute the major organising structure of the CISS, the Basic scales (as discussed in the previous section) were constructed prior to the Orientation scales. In constructing the Orientation model, the developers' primary goal was to create between five and seven broad categories of interests in terms of which the Basic Interest and the Basic Skill scales could be summarized (Campbell et al., 1992).

Principle component analyses were conducted on the Basic interest scales. The appropriateness of a five, six and seven-component solution was considered. Based on the Kaisers criterion, inspection of the scree plots and the factor interpretability of the various solutions, the authors felt that the seven-component solution summarised the data best (Campbell et al., 1992). The seven interest components were defined above as the seven interest orientations. The seven interest orientations therefore, essentially, are seven second-order interest factors.

Little was known at the time regarding self-reported skill and whether the skill structure would reflect the same underlying structure as the interests (Campbell et al., 1992). Therefore, the research question arose as to whether a similar second-order skill factor structure would emerge from the first-order skill scales. This was hoped that the second-order interest and skill structures would be the same due to the authors' beliefs that interests and skills should be measured in a parallel fashion. The Basic skill scales were consequently also subjected to principal components analyses and according to Kaiser's criterion and an inspection of the scree plots, a six-component solution was indicated to fit the data best. However, the authors also chose a seven-component solution for the Skill scales, mostly due to the six-component solution not being easily interpretable (Campbell et al., 1992).

The seven-component solutions for both the interest and skill scales were rotated using both varimax and promax procedures to find the most interpretable solution. Promax rotation did not result in a cleaner, more sensible structure contrary to the authors expectations, therefore the varimax rotation solution which assumes orthogonal second-order factors was adopted (Campbell et al., 1992).

The authors report factor loadings from the principle components analyses of the Basic scales in the instrument's manual. In their opinion "the loading for the two sets of scales were quite similar and showed good support for the seven CISS Orientations" (Campbell et al., 1992, p. 66). In the instances where certain scales loaded on more than one component the scale content and theoretical reasoning provided judgment on their appropriate placement. Campbell et al. (1992), therefore, interpreted their factor analytic evidence to indicate that the same set of seven orientations underlie as second-order factors both the basic interest scales as well as the basic skill scales.

All items on the Basic scales contribute to the scores on their corresponding Orientation scales, with one exception. Although item 200, "Write a technical report" is part of the Writing Basic interest scale, it nonetheless contributes to the score on the Creating Orientation interest scale rather than the aNalyzing Orientation scale. At the item correlation level, this item clustered more closely with other writing items than with items in the science and mathematics area. It was therefore decided that the item would contribute to the aNalyzing Orientation scale rather than the Creating Orientation scale.

The instrument manual provides information regarding the correlations[6] between the Basic scales and the seven Orientations. In Campbell et al.'s (1992) opinion, this approach best covers the entire breadth of their domains. Mean correlations between Basic Interest scales and Orientations vary between 0.60 and 0.90 with a median of 0.72. Mean

[6] Since the 7 factor structure was orthogonally rotated to simple structure the factor loadings of the basic scales on the second-order factors can be interpreted as correlations (Tabachnick & Fidell, 2001).

correlations between Basic Skill scales and Orientations vary between 0.65 and 0.90 with a median of 0.77.

The following represents the seven broad (global level) areas of interest and self-reported skill include[7]:

i) Influencing – covers the general area of leading and influencing others; *influencing* others through leadership, politics, public speaking and marketing. Typical high-scoring occupations include company presidents, corporate managers and school superintendents.

ii) Organizing – activities that bring orderliness and planfulness to the work environment. *Organizing* the work of others, managing and monitoring financial performance. Typical high-scoring occupations include accountants, financial planners, office managers and administrative assistants.

iii) Helping – involves helping and developing others; *helping* others through teaching, healing and counselling. Typical high-scoring occupations include counsellors, teachers and religious leaders.

iv) Creating – includes artistic, literary and musical activities; *creating* artistic, literary or musical productions, and designing products or environments. High-scorers include artists, musicians, designers and writers.

v) aNalyzing – involves scientific, mathematical and statistical activities; *analyzing* data using mathematics and carrying out experiments. Typical high-scorers are scientists, medical researchers and statisticians. This interest orientation is labelled with the letter N as A is used for the Adventuring orientation (see below).

vi) Producing – covers practical, hands-on, productive activities. *Producing* products using "hands-on" skills in farming, construction and mechanical crafts. Typical high scoring occupations include: mechanics, veterinarians and landscape architects.

vii) Adventuring – includes activities in athletics, the police and military. *Adventuring*, competing and risk taking. Typical occupations: military officers, police officers and athletic coaches.

[7] The underlined letter is used to represent the seven interest orientations.

The relationship between the seven orientations is represented in Figure 3.2. Orientations closer to each other are regarded to be similar in the sense that individuals that would be attracted to - and would feel confident in a specific area - would tend to also be attracted to and feel confident in interest areas in close proximity to the focal area, albeit somewhat less so. Those orientations that are diagonally across are regarded to be dissimilar in the sense that individuals that would be attracted to and would feel confident in a specific area would tend to not be attracted to and feel confident in interest areas diagonally opposite the focal area. This is akin to the hexagonal relationships of interests as proposed by Holland (1985).

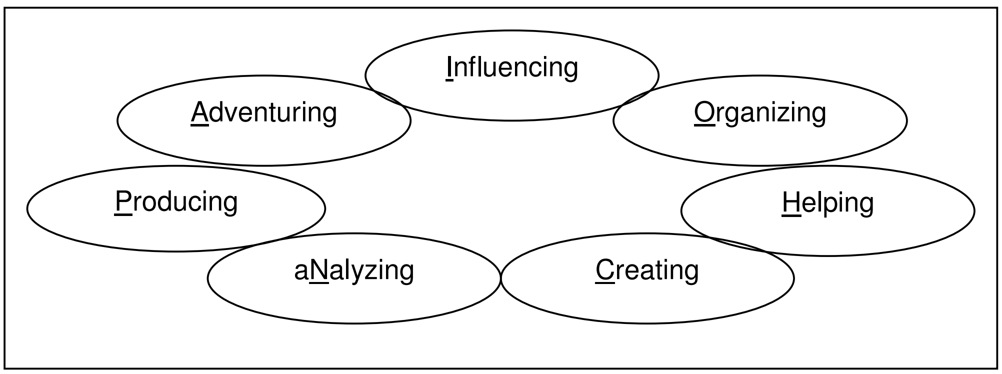

Figure 3.2. A graphic representation of the CISS Orientation scales.

Although similar to the Holland typology an additional category, Adventuring, has been added. The similarities and differences between the Holland typology and the CISS are represented in Table 3.2. The differences include: (a) the CISS Influencing orientation focuses more on leadership, whereas the Holland Enterprising type is tilted toward sales activities, (b) The CISS Organizing orientation focuses on management and financial services, whereas Holland's Conventional type looks at office and clerical work; (c) The CISS presents the Holland's Realistic type as two orientations, namely Producing and Adventuring.

Table 3.2.
Correspondence Between The CISS Orientations And Holland Typology

CISS Orientations	Corresponding Holland Themes
Influencing	Enterprising
Organizing	Conventional
Helping	Social
Creating	Artistic
Analyzing	Investigative
Producing	Realistic
Adventuring	Realistic

Adapted from Campbell *et al.* (1992, p. 56)

Norming of the Orientation scales was conducted by gathering the data from 65 occupational samples. A raw score mean was calculated on each Orientation scale for female respondents (*n*=1790) as well as for male respondents (*n*=3435). The two raw score *means* were averaged to create the effect of equal gender weighting; the unweighted means of means is then used in the raw-score-to-standard-score conversion formula. This would suggest that no significant distributional differences exist on the seven orientation scales between the two genders. Calculating a mean of means would not be a problem if the gender distributions coincide since the interpretation of a specific raw orientation score will remain the same irrespective of whether it is compared to the scores typically achieved by males or females. If, however, male and female distributions differ on any given orientation scale the interpretation should differ depending on which group it is compared to. The extent to which the orientation distributions coincide across gender is not indicated in the manual (Campbell et al., 1992). The authors did, however, aim to ensure that the standard deviations of all the scales are approximately equal across gender groups. Ultimately, respondents' Orientation scale raw scores are compared to *happily*[8] employed people spread over 65 equally represented occupational samples (Campbell et al., 1992). This norming process was also applied in generating norms for the Basic scales. No norms have as yet been created for the South African context (N. Taylor, personal communication, 31 October, 2008).

[8] The authors do not indicate their criteria for determining who is employed *happily*. It would be fair to rather describe these individuals as engaged in their respective occupations. It also should be noted that even though the individuals may be engaged in their respective occupations, their satisfaction with their career decision does not seem to be factored into the equation. The assumption that individuals that are engaged in occupations signals happiness in their roles would need to be researched further.

As with the Orientation scales, higher scores in the Basic scales indicate an attraction for the activities that the Basic scales measure. High scores on the Basic skill scales represent a self-perceived confidence in the activities that the Basic scales measure (Campbell et al., 1992).

3.2.3 Occupational scales

The Occupational scales of the CISS provide the individual seeking their ideal career with additional information regarding their interest in relation to other *happy* individuals in their chosen careers (Campbell et al., 1992). This is achieved through a derived set of interests and skills that match with a particular occupation. This is essentially the criterion referencing technique that Strong pioneered in his work on interest assessment. “They [Occupational scales] were developed empirically by contrasting the responses of occupational samples with those of a reference sample widely drawn from the working world” (Campbell et al., 1992, p. 138).

The Occupational scales assist in providing the career finder with occupational titles that may warrant further investigation because they suit his/her specific profile of occupational activity interests and skills. This scale also has the benefit of turning the assessment into commonly understood terms in contrast to the more abstract conceptual terms used in the Orientation and Basic scales.

Due to the CISS being developed using an item pool to cluster factors (Basic scales), it would not be essential in the present study to discuss the Occupational scales in detail. The Basic scales and Orientation scales are the naturally emerging structures from the items used in the construction of the instrument. This study focuses on these measurement models of the instrument and not on the empirically derived Occupational scales. It would seem essential to establish whether the Basic Interest and Skill measurement models fit at least reasonably well before embarking on tests of the validity of the Occupational scales. The validity of the Occupational scales will not be evaluated in the present study.

3.2.4 Special scales

The CISS has additional special scales that will further assist the career explorer. The three scales are: (a) an Academic Focus scale, (b) an Extraversion scale and (c) a Variety scale. As with the Occupational scales, these scales were derived by identifying items from the measurement model and comparing the interests and self-reported skills to individuals' scores at various levels in terms of academic, extraversion and need for variety preferences. The Academic Focus scales measure the respondent's interest and confidence in doing well in academic settings. Scores on the Extraversion scale reflect the strength of the respondent's interest and self-confidence in working with people. The Variety scale indicates an individual's preference for diverse interests/skills – however the authors are not clear on how this score should be used (Campbell et al., 1992). The authors (Campbell et al., 1992) do however, point out that the variety scale score does seem to indicate whether respondents opt for the extreme positive statements on many items.

As was the case with the Occupational scales, these scales will also not be investigated in this study as they are also derived empirically.

3.3 OFFICIAL PSYCHOMETRIC PROPERTIES OF THE CISS: TECHNICAL MANUAL

In this subsection, the reported psychometric properties, as per the technical manual (Campbell et al., 1992), (i.e., reliability and validity) of the CISS will be discussed. The discussion will be limited to the Orientation and Basic scales. Information presented in this subsection is based on data gathered in the United States. South African data is not reported in the official CISS manual.

3.3.1 Orientation scales

3.3.1.1 Reliability.

Reliabilities reported in the CISS test manual (Campbell et al., 1992) focus on internal consistency reliability coefficients (alpha) and test-retest correlations.

The median coefficient alpha for both the interest and skill scales was 0.87 at time of publishing. The alpha reported for the Creating Skill scale was slightly lower at 0.76 – which indicates that the scale is less homogenous than the others. The alpha coefficients were calculated on a diverse sample[9] of employed American adults (*N*=4842).

Test-retest reliabilities were calculated using a sample of 324 individuals (230 males, 94 females) who completed the CISS twice over approximately 90 days. Pearson correlation coefficients were calculated to determine test-retest reliability. The median interest correlation was reported as 0.87 and the median skill correlation was reported as 0.81. Both these figures are "high enough to demonstrate considerable stability" (Campbell et al., 1992, p. 84).

3.3.1.2 Validity.

Validity of the CISS, as reported in the manual (Campbell et al., 1992), is represented only by scale intercorrelations and by examining the scores of individuals engaged in occupations that would theoretically attract a score higher on specific scales. In the latter instance the interest and skill scores of individuals in occupations in which the specific interest and skills are theoretically expected to be related to satisfaction and success were correlated with these two criterion variables. If the relevant interest and skill distributions would differ significantly in terms of the mean across contrast groups created in the occupational sample in terms of their degree of engagement, it would constitute evidence that the interest and skill inferences derived from the scales are permissible (i.e., valid).

The authors of the CISS evaluated the construct validity by calculating Pearson correlations between the Interest Orientation scales and the Skill Orientation scales (in essence convergent validity in as far as the two scales are theoretically expected to correlate positively[10]) and they report correlations between 0.76 and 0.66 with a median of 0.70. This means that approximately 50% of their variance is in common (Campbell et al., 1992). Construct validity refers to the extent to which a measuring

[9] Unfortunately the descriptive statistics of the American sample are not indicated in the manual.
[10] Convergence would indicate that evidence gathered from different sources in different ways all indicate the same or similar meaning of the construct (Kerlinger & Lee, 2000), therefore a positive correlation would be expected between the gathered information.

instrument measures the theoretical construct it was designed to measure in accordance with its constitutive definition (Cronbach & Meehl, 1955, as cited in Kerlinger & Lee, 2000). The constitutive definition explicated the internal structure of the construct and the manner in which the construct and its dimensions are related to other constructs in an interconnected network of latent variables. Confidence in the construct validity of a measure is increased (but never fully established) if empirical confirmation is obtained for (a) the relationships the constitutive definition claims should exist between the relevant construct and other constructs contained in a nomological network through correlation and regression analysis or through structural equation modelling (SEM), and (b) the internal factor structure the constitutive definition claims should exist with regards to the dimensions comprising the construct through (confirmatory) factor analysis or SEM. The CISS manual does not report the results of any confirmatory factor analytic investigation of the factor structure underlying the CISS nor does it report the results of a SEM investigation of the nomological validity.

Concurrent and predictive validity was established by calculating mean scores for both interest and skill scales for 58 different occupational samples. The mean scores were then ranked from highest to lowest to show which types of occupations are occupied by people with strong interests and confidence in each Orientation scale. The top five occupations for each Orientation are as follows (Campbell et al., 1992):

- i.) Influencing: Media Executive, Marketing Director, Hotel Manager, Public Relations Director and Manufacturer's Representative.
- ii.) Organizing: Accountant, Bank Manager, Retail Store Manager, School Superintendent and Bookkeeper.
- iii.) Helping: Religious Leader, Guidance Counsellor, Child Care Worker, Nursing Administrator and Athletic Trainer.
- iv.) Creating: Fashion Designer, Liberal Arts Professor, Translator/Interpreter, Librarian and Commercial Artist.
- v.) Analyzing: Maths/Science Teacher, Medical Researcher, Chemist, Statistician and Veterinarian.
- vi.) Producing: Veterinarian, Agribusiness Manager, Carpenter, Landscape Architect and Test Pilot.

vii.) Adventuring: Police Officer, Athletic Coach, Military Officer, Test Pilot and Athletic Trainer.

Overall, it would seem as if the content of each Orientation scale relates to occupations fitting of these orientations. The evidence suggests that people tend to migrate to and remain in occupations that correspond to their interest in activities associated with those occupations as well as their confidence in performing activities related to those occupations. This provides some, albeit limited, support for the concurrent validity of the Orientation scales. Stronger, more convincing evidence of the criterion-related validity of the career-related inferences derived from the CISS Orientation scales would be obtained if appropriate criterion variables (for example: satisfaction and career success) would be regressed on linear composites of interest and skill scales that would theoretically be expected to explain variance in the criterion variable.

3.3.2 Basic scales

3.3.2.1 Reliability.

Reliabilities for the Basic scales are also focused on the calculation of an internal consistency reliability coefficient (alpha) and test-retest correlations.

The median coefficient alpha for the interest scales was 0.86 and 0.79 for the skill scales at time of publishing of the manual. The alpha reported for the Performing Arts Skill scale was lower at 0.62. The authors felt that the low internal consistency obtained for the Skill scales would suggest that confidence in one performing art (e.g., acting) activity might not be generalised to other specific performing art activities (e.g., singing, dancing etc). It should also be noted that the number of items loaded on each skill scale are less than those shown for the interest scales – this may be a reason for lower reliabilities (Kerlinger & Lee, 2000). The alphas were calculated on the sample of 4842 individuals (Campbell et al., 1992).

Test-retest reliabilities were calculated using a sample of 324 individuals (230 males, 94 females) who completed the CISS twice over approximately 90 days. Pearson correlation coefficients were calculated to determine test-retest reliability.

The median interest test-retest correlation was reported as 0.83 and the median skill test-retest correlation was reported as 0.79 (Campbell et al*.,* 1992, p. 84).

In the questionnaire authors' opinion, "data from both sources indicate that the CISS Orientation and Basic scales are measuring constructs that are homogenous and stable over short time periods" (Campbell et al*.,* 1992, p. 135).

3.3.2.2 Validity.

As with the Orientation scales, the validity of the Basic scales, as reported in the CISS test manual (Campbell et al*.,* 1992), is the nature and magnitude of the scale intercorrelations and by examining occupational samples that score higher on specific scales that should, theoretically, link to the said occupations. If this is found then a true positive or "good hit" would be indicated (i.e., those scoring higher on scales that match are employed in related occupations). In the latter instance the interest and skill scores of individuals in occupations in which the specific interest and skills are theoretically expected to be related to satisfaction and success were correlated with these two criterion variables. This was essentially the approach that Strong used in his questionnaire development.

The authors evaluated the construct validity of the Basic scales by calculating Pearson correlations between the Basic Interest scales and the Basic Skill scales (in essence convergent validity) and report correlations between 0.80 and 0.46 with a median correlation of 0.68 (Campbell et al*.,* 1992). Again it needs to be noted (given the argument presented in paragraph 3.3.1.2) that the CISS manual does not lead sufficient psychometric evidence to allow a confident positive verdict on the construct validity of the instrument

Concurrent and predictive validity was established by calculating mean scores for both interest and skill scales for 58 different occupational samples. The mean scores were then ranked from highest to lowest to show which types of occupations are occupied by people with strong interests and confidence in each of the Basic scales (Campbell et al*.,* 1992). Examination of the tables in the CISS test manual (Campbell et al*.,* 1992) reveals that the distribution of occupations across the Basic scale are generally the same as was found for the Orientation scales (see paragraph 3.3.1.2 above). It is somewhat disconcerting that regression analyses were not conducted in order to predict occupation choice (as well as additional relevant career criterion variables like career satisfaction and career choice) as a result of reported interest and skill. Such information would have certainly made for a stronger argument for the concurrent and predictive validity (depending on the validation design) of the career-related inferences typically derived from the CISS.

3.3.2.3 Gender differences.

Campbell et al. (1992) investigated the differences between the genders by calculating the mean scores obtained for each scale by using the occupational samples that were used for the validity studies. The mean differences between the genders were also calculated. With the exception of seven of the Basic scales, the mean differences were less than five points. Women score higher than men in Child Development and Fashion. Men score higher than woman in Financial Services, Mechanical Crafts, Woodworking, Military/Law Enforcement and Risk/Adventure. The manual does not report whether these differences were statistically significant.

The meaningful comparison of gender scores on the Basic Interest and Skill scales assumes that the manner in which the observed item responses on each item relates to the underlying latent variable is the same for the two gender groups. The assumption is, therefore, that the slope and intercept parameters that describe the regression of the item response on the latent variable are the same for men and women. If this assumption is met it would imply that a specific observed score (X) obtained by a man and a woman on any of the Basic Interest and Skill scales reflects the same standing on the underlying latent variable (ξ). Even though the CISS shows acceptable levels of reliability and sufficient validity, no credible psychometric research evidence is available on the equivalence of the measurement model parameters across the genders. Gender differences have been found in the observed scores on some of the Basic scales yet the slope and intercept parameters of the model have not been shown to be equivalent across gender. In conclusion, the lack of gender bias in the CISS has not yet been established. Without evidence on the measurement equivalence, however, the debate on gender differences in interests will not be able to proceed meaningfully.

The CISS has made progress in ensuring that the items in the instrument do not contain gender loaded terms and the instrument has consolidated scoring into one norm structure. A possible reason for the reduction in the magnitude of the gender differences in a number of scales that are reported in the manual (Campbell et al., 1992) could be due to the emergence of women in typically male-dominated careers. Additional factors that could have contributed to this reduction in interest profile differences between genders could be the possibility of males occupying stereotypically female occupations, and the fact that individuals of both genders have more knowledge about the contents of jobs that were previously unfamiliar to members of a specific gender due to the effect of gender stereotyping and thus not considered. Consequently, the narrowing of the gap between the scores on interest inventories is becoming a reality. The fundamental question, however, still remains whether equal differences in observed interest and skill scores across genders can be interpreted as indicative of corresponding differences in latent interest and skill scores (slope invariance) and whether the same observed interest and skill score across genders indicate the same standing on the latent interest and skill latent variable (intercept and slope invariance). The question is therefore whether the measurement model parameters characterizing the manner in which item responses relate to the underlying latent variables are the same across gender groups. No measurement invariance studies are reported in the CISS test manual (Campbell et al., 1992). Since this question has not been sufficiently investigated on the CISS it, thus, warrants further investigation.

3.4 PSYCHOMETRIC PROPERTIES OF THE CISS: INDEPENDENT RESEARCH

3.4.1 Construct validity of the interest scales

Sullivan and Hansen (2004) conducted research focussing on the construct validity of the interest scales of the CISS. The motivation for the study was based on the limitations of a similar study conducted by Savickas, Taber and Spokane (2002). The Savickas et al. (2003) study examined the convergent validity of the CISS with other instruments assessing interest. However, conclusions made were limited due to a homogenous sample (all respondents were either counselling professionals or professors). The sample size was too small to conduct separate gender analyses, an important omission given the well-documented gender differences in interest measurement. The Savickas et al. (2002) study also only evaluated convergent validity on the Orientation scales.

The Sullivan and Hansen (2004) study looked at all three scales of the CISS, Orientation, Basic and Occupational Interest scales. The correlational study compared CISS scales with the SII scales. In some cases it was not easy to identify the corresponding scale on the SII. Nevertheless the analysis on the matching scales was conducted. The study also aimed to evaluate the hexagonal relationship of the Orientation scales as suggested by Campbell (Campbell et al.*,* 1992) as a reflection of the Holland model.

The results of the analyses indicate that for the Orientation scales and the matched GOT scales (in the SII) the median correlation for females was 0.72 whereas the median correlation for non-matching scales was 0.17. Similar findings for the men are a median correlation of 0.69 for matching scales and 0.15 for non-matching scales. However, the Adventuring scale demonstrated the least evidence of convergent validity which is not surprising considering that this scale is not directly assessed by any other interest measure. The research also indicates that the Adventuring scale functions differently for the genders. Women tend to respond to the interpersonal aspects of the Adventuring scale whereas men tend to respond to the practical elements of the scale. Overall, the matching Orientation scales share about 50% of the variance, and non-matching scales share about 2% (Sullivan & Hansen, 2004). Findings in this study also support the hexagonal relationship.

The CISS Basic scales were matched with the SII BIS scales as was the case with the Orientation scales analyses. The median correlation for women was 0.75. The median correlation between non-matching scales was 0.15. For men, the matching median correlation was 0.74 and the non-matching median correlation 0.15.

Factor analyses were also conducted using both the SII BIS and CISS Basic scales to determine which scales from both instruments share variance. The purpose of this analysis was to confirm construct validity by using the well-established SII. Some gender differences were noted. For women the Investigative scales (CISS: Analyzing; SII:

Investigative) and two of the Realistic scales (CISS: Producing; SII: Realistic) loaded on a single factor. The CISS Adventuring scale did not load onto this factor. This suggests that females may not distinguish between these two categories of interest. However for females, the Artistic scales loaded on a different factor than the Social scales. Interestingly enough, for females the Adventuring scales loaded with the Social scales. Involvement of females in adventure activities seem to be motivated by a social interaction (as per the SII) need. This scale should be more closely aligned with the Realistic scales. For men, the Social (CISS: Helping; SII: Social) and Artistic (CISS: Creating; SII: Artistic) scales loaded on a single factor, suggesting that men do not distinguish between these two categories of interest (Sullivan & Hansen, 2004).

As for the Occupational scales, the median correlation for matching scales was 0.62 for women and 0.66 for men. Non-matching scales indicate a median correlation of 0.06 for woman and 0.05 for men.

From the reported studies it might appear as if there are gender differences in the structure of the interests. The fact that these conclusions were derived from observed interest and skill scale score differences without any credible research evidence on the equivalence of the CISS measurement model parameters caution against reaching a premature conclusion on gender differences in interests. The preceding argument would seem to suggest that the investigation of the measurement invariance/equivalence (Vandenberg and Lance, 2000) of the parameters of the measurement model underlying the CISS is an unavoidable necessity. No studies that investigated the invariance of the parameters of the CISS Interest and Skill measurement models across gender groups could be traced in the literature. In fact, no reference to any measurement invariance study on the CISS could be traced in the literature.

3.4.2 Construct validity of the skill scales

Hansen and Leuty (2007) conducted a further research study that aimed to clarify the skill construct in the CISS. The study aimed to firstly test the convergent and divergent validity of the CISS Skill scales on the CISS Interest scales. A second objective was to compare the CISS Skill scales and the SII Basic Interest scales to determine how the CISS Skill scales correlate with the SII BIS. This approach was necessary to evaluate the relationship between self-reported skill and interest independent of the similar item content as seen in the CISS. A third analysis comprised of correlating the CISS Skill and Interest scales with an independent measure of self-perceived abilities; the Minnesota Abilities Estimate Questionnaire (MAEQ, Desmond & Weiss, 1973, as cited by Hansen & Leuty, 2007).

The matching CISS Skill and Interest scales correlated between 0.46 and 0.71 for females, and between 0.62 and 0.72 for males. For the most divergent scales correlations varied between -0.1 and 0.00. For ease of interpretation of the Basic Interest and Skill scale correlations, the various matched Basic Skill and Interest scales were grouped into the global Orientation scales structure of the CISS and a median correlation was calculated for each Orientation. The median correlations for each group ranged between 0.22 and 0.56 for females and 0.24 and 0.55 for males. Similar findings were obtained for the Occupational scales (Hansen & Leuty, 2007).

The analyses of the CISS Orientation Skill scales and SII GOT scales (interests) resulted in correlations ranging between 0.25 and 0.52 for similar orientation scales for females and between 0.38 and 0.59 for males (Hansen & Leuty, 2007). Correlations were stronger for the CISS Orientation Interest scales and the SII GOT scales than for the CISS Orientation Skill scales. This finding could be expected based on the results of the earlier Sullivan and Hansen (2004) study. The CISS Basic Skills scales and the SII Basic Interest scales were also compared. Results indicated median correlations of between 0.19 and 0.47 for females and between 0.26 and 0.45 for males. Similar results were found for matching Occupational Skill scales and SII Occupational Interest scales. Based on these findings, it could be concluded that self-reported skills only share a relatively small proportion of common variance with interests and that the research evidence (Hansen & Leuty, 2007) suggests that self-reported interests in specific activities is a construct that should be distinguished from self-reported skill or confidence in those activities. Despite a moderate correlation between the two constructs, showing interest in specific activities should be seen as conceptually distinct from being skilful or confident in displaying those activities. Therefore, for the analyses in the current project, it is best to keep the interest and skill measurement models in the CISS separate.

The results of the third analysis: correlations between the Skills scales and the MAEQ findings reveal some significant relationships, however, the correlations are mostly modest and tend to lie between 0.02 and 0.49 for females and between 0.02 and 0.55 for males. A limitation in this analysis is that the MAEQ measures specific self-perceived abilities, namely verbal aptitude, numerical aptitude, spatial aptitude, form perception, clerical perception, finger dexterity and manual dexterity, whereas the CISS Skill scales tend to measure interest field self-efficacy perceptions. Therefore, it would seem as if these scales tend to measure different domains in the majority of cases. Hansen and Leuty (2007) indicate that self-perceived skill measurement may be more indicative of self-efficacy than of self-estimated abilities.

3.4.3 Concurrent validity for the skill scales

Two studies that focused on whether higher interest scores predicted choice of college majors were conducted by Hansen and Neuman (1999) and Pendergrass, Hansen, Neuman and Nutter (2003, as cited by Hansen & Leuty, 2007). The results indicate that the CISS interest measurement as well as the SII predict choice of college major. The major limitation of these studies is that the CISS is recommended for both university student and working adults that are currently pursuing or have pursued postsecondary studies. Both of these independent studies have not looked at the prediction of actual careers chosen after university. As far as concurrent prediction of the Skills scales, Hansen and Leuty (2007) conducted analyses using the McArthur Method[11] (1954) to determine the hit rates of higher skill scores with choice of college major. Results show that 57.7% of females were classified as having excellent or good hits for college major selection based on their CISS Occupational Skill scales scores. For males, an excellent or good hit was 69%. The differences were found to be not statistically significant.

[11] The McArthur Method is a common technique used to classify occupational interest scores for concurrent validity studies. The technique is synonymous with the Strong Interest Inventory. A set of decision rules are defined to determine whether scores on the interest questionnaire coincide with the intended occupational membership. Hit rates are categorized into poor, good and excellent. For the CISS, scores that were no more than half a standard deviation below the mean for the occupational criterion sample (choice of college major) were classified as excellent hits. CISS scores between 1.0 and .5 standard deviations below the mean were classified as good hits, and scores more than 1.0 standard deviation below the mean were categorized as poor hits (Hansen & Leuty, 2007).

CHAPTER 4
BIAS AND MEASUREMENT EQUIVALENCE/INVARIANCE

The previous discussion argued that the CISS attaches specific connotative interpretations (Kerlinger & Lee, 2000) to the interest and skill latent variables. Specific latent interest and skill dimensions are distinguished in terms of this conceptualization. Specific items have been designed to serve as effect indicators (Hair et al., 2006) of these latent interest and skill dimensions. This design intention is reflected in the scoring key of the CISS. Two very specific measurement models are moreover implied by the design intentions (and the scoring key) of the developers of the CISS. A critical question is whether these measurement models, reflecting the design intentions of the developers, fit data obtained from the instrument at least reasonably well. If the interest and skill measurement models would fit the data at least reasonably well the foregoing discussion moreover argued the necessity of investigating the question whether the measurement model parameters differ across gender groups. This chapter aims to critically review the literature on the methodology of measurement invariance with the purpose of describing and justifying a best practice SEM based procedure of investigating measurement invariance. The conclusions of chapter 4 will inform the procedure that will be used to investigate the measurement invariance of the CISS. The research methodology underpinning this study will be presented in chapter 5.

4.1 DEFINING MEASUREMENT

“Measurement in the broadest sense is defined as the assignment of numerals to objects or events according to rules” (Stevens, 1946, p. 677). Janda (1998) states that measurement is critical to all areas of psychology and the social sciences. Vandenberg and Lance (2000, p. 4) are more specific in stating that “measurement can be defined as the systematic assignment of numbers on variables to represent characteristics of persons, object or events.”

These definitions of measurement are, however, still not satisfactory in as far as they do not explicitly reflect the fact that psychological measurement is the indirect measurement of abstract constructs or latent variables through observable indicators in which the latent variables express themselves (reflective indicators)[12]. Latent psychological variables (like interest) are measured indirectly by eliciting a sample of observable behaviour through a sample of stimuli (items). Numerals are therefore, strictly speaking, not assigned to the latent variable characterizing a person, but rather to behavioural indicators of the characteristic (Murphy & Davidshofer, 2005).

In an attempt to ensure that variance in the assigned numerals only reflect variance in the underlying latent variable the assigned numerals are meant to reflect, extraneous variables that could cause non-relevant variance in the observed/allocated scores need to be controlled. Two broad sources of extraneous observed score variance exist, namely non-relevant latent variables and different variables characterizing the test conditions (in essence the stimulus set; Hair et al*.,* 2006). Through a process of item analysis, test item stimuli are identified (and removed if necessary) that do not primarily reflect the latent variable of interest (Murphy & Davidshofer, 2005). In addition, through item analysis unreliable, invalid and biased items are identified and deleted from the stimulus set. Moreover it is important to ensure that the rules, referred to by Stevens in terms of which numerals are assigned to the behavioural indicators of the latent variable, are applied consistently to ensure that variance in the observed measure only reflects variance in the (to be) measured characteristic of persons, objects or events. This is achieved by standardising the test procedure so that test stimuli, test instructions and test material remain the same across test takers and test administrators and therefore do not cause variance in the observed test scores (Theron, 1999). The processes of item analysis and standardization never fully succeed in controlling non-relevant systematic and random influences that cause responses to test stimuli to vary.

A more encompassing definition would be: measurement, in the social sciences, is the assignment of numerals to behavioural indicators of latent variables to represent characteristics or persons, object or events according to standardized rules. The more encompassing definition implies the following logic. A psychological measurement procedure elicits a sample of behaviour through a sample of standardized stimuli under standardized conditions. The stimulus sample is constructed so as to reflect the underlying construct of interest in the observable response of the testee to it. In addition,

[12] The possibility of formative indicators (Hair et al., 2006) is not considered here.

the stimulus sample is constructed so that the quality or nature of the behavioural response to the stimuli is dependent on the underlying construct. Procedures are finally formulated in terms of which the elicited behaviour is observed, recorded and quantified. Given the (assumed) dependence of the behavioural response to the stimulus sample on the construct of which quantitative information is desired, differences in observed scores obtained by *n* testees should therefore indicate differences in the construct of interest. Reflecting the foregoing logic, the Human Sciences Research Council (Owen & Taljaard, 1988) defines a psychological test as:

> … a purpose-specific evaluation and assessment procedure used to determine characteristics of people in areas of intellectual ability, aptitude, interests, personality profile and personality functioning. It consists of a collection of tasks, questions or items aimed at eliciting a certain type of behaviour under standard circumstances, from which scores with acceptable psychometric characteristics are inferred according to prescribed procedures.

Measurement occurs in the applied social sciences and more specifically applied Industrial Psychology to inform decisions. To make decisions, information on relevant latent variables characterizing the decision problem is required. Obtaining information in a quantitative format has distinct advantages. Information in a quantitative form is more precise and unambiguous[13]. Quantitative information also allows for the application of statistical procedures to evaluate the quality of the information (via reliability and validity analyses) as well as it allowing for (mathematical) argumentation that may not have been possible with information in, for example, a verbal format (Neuman, 2000). Measurement, therefore, provides information on one or more latent variables relevant to the decision that allows the decision-maker to determine a route of action based on observations of the variables of interest. In the case of this study, a key objective would be to enable the psychologist to provide comprehensive information to a client regarding his/her interests, this should then in turn, assist the individual in sound career selection and consequent life satisfaction.

Implicit in this argument is the assumption that the latent variables of interest are measured reliably, validly and without bias. In addition, it is also assumed that it is permissible to make inferences about the latent variables of interest and that these

[13] Strictly speaking, this is true only if measurement occurs on an interval or ratio scale.

inferences are not different for members of different groups. It would therefore be highly questionable to provide feedback to male and female clients, seeking assistance in making career decisions, based on measures of various interest and skill dimensions if those measures do not reflect the latent variables (i.e., interest and skill) in a reasonably accurate manner. Moreover, the process of assisting clients in making a decision on the appropriate career to pursue would be seriously complicated if the measures, obtained for males and females, would reflect different interest and skill constructs. Although probably less so, the process of assisting clients in making a decision on the appropriate career to pursue would also be seriously complicated if specific observed measures of various interest and skill dimensions would not signal the same standing on the underlying latent variables[14]. Therefore, the questions are whether valid inferences can be made for members of both gender groups about the career interest and career skill constructs as they are constitutively defined by the CISS and whether the nature of such inferences would be the same across the two gender groups. The former is a question of whether the manner in which responses of men and women to the CISS items relate to the underlying latent variables is the same. The latter relates to the question of whether the slope and the intercept of the regression of the observed item responses on the latent traits are the same across the two gender groups. A helpful taxonomy to address each of these issues is discussed later in this chapter.

The responses of men and women to the items of the CISS are quantified in accordance with a set of rules captured by the scoring mask and instructions of the instrument. These rules are standardized and are therefore applied in the same manner across the two genders. The questions raised above are really about whether this should be the case. Ideally the rule in terms of which behavioural responses to items are quantified should be based on the nature of the relationship between the observed responses and the latent variable represented by the item[15]. If the slope and/or intercept of the regression of the behavioural response on the latent trait differ for men and women, one could argue that the rule in terms of which behavioural responses to items are quantified should reflect this

[14] It could be argued that if uniform and/or non-uniform gender bias would exist in specific CISS items and the nature of the bias would be such that it results in observed scale score differences between the genders that cannot be explained in terms of the latent variable, this situation still need not unavoidably result in the inferences on career success and/or career satisfaction being biased. This line of reasoning, however, presupposes the development of an actuarial inference rule that would include a gender main effect and/or gender x predictor interaction effects. The development of such actuarial inference rules would undeniably pose severe *practical*, technical and logistical challenges to the psychologist.

[15] Essentially, this is what is achieved if a Rasch model (Rasch, 1980) is fitted to item data.

difference. The rules governing item scoring do, however, not typically reflect the specific nature of the relationship between the observed responses (X) and the latent variable (ξ) represented by the item (i.e., the regression of X on ξ). Nor do they typically acknowledge slope and intercept differences of the regression of the behavioural response on the latent trait. Rather, items displaying uniform or non-uniform bias are frowned upon and considered for deletion[16].

Although, not expressed in so many words, the developers of the CISS have nonetheless, by implication, taken the position that the measurement model implied by the scoring key is invariant across the two gender groups. The position of the publishers presents itself in the discussion regarding gender differences in career-related interests and confidence in career-related skills observed during the research stages of the instrument's development. Without appropriate empirical research evidence, however, the question regarding the equivalence of the parameters of the measurement model underlying the CISS across the genders remains unanswered and therefore high on the research agenda because measurement invariance is a necessary prerequisite for any meaningful cross-gender comparisons. The latent mean score comparisons made thus far would remain ambiguous unless it can be shown that the hypothesised measurement model suggested by Campbell et al. (1992) actually fits the data gathered and unless it can be shown that measurement model parameters remain invariant across gender groups. To be fair, it should be conceded that measurement invariance only recently started receiving increased research attention (Vandenberg & Lance, 2000; Vandenberg 2002). The developers of the CISS are therefore not accused of negligence or in any way blamed for the lack of gender-related measurement invariance research on the instrument. What is true for the CISS is probably also true for most other psychological tests being used in South Africa today.

[16] From a structural equation modelling perspective the question could be raised whether it would not be possible to estimate latent scores for men and women via a well fitting model that acknowledges differences in slope and intercept of the regression of the behavioural response on the latent trait where they exist. The algorithm underlying the estimation of the latent scores then becomes the rule in terms of which numerals are assigned to behavioural indicators of latent variables.

4.2 BIAS IN MEASUREMENT

Van de Vijver and Leung (1997) regard the term *bias*[17] as a generic term used for all the systematic nuisance factors threatening the validity of cross-cultural (group) comparisons. Bias in this sense refers to all factors that could account for variance in observed test scores that cannot be accounted for in terms of the latent variable of interest. A specific latent variable or construct is indirectly measured by requesting testees to respond to a sample of test stimuli under standardized conditions under the assumption that the responses would be largely governed by the construct of interest. The response to the test stimuli is, however, not solely an expression of the construct of interest. Other systematic but non-relevant factors and non-systematic, random factors also play a role in determining the response to the test stimulus set. These systematic nuisance factors essentially refer to any systematic source of unique variance in the test scores that cannot be explained in terms of variance in the latent variable of interest. When referring to bias in measurement not all sources of systematic error are of equal interest. In the analysis of measurement bias the focus typically falls on sources of systematic measurement error that are systematically related to (cultural, language, gender) group membership.

Despite the fact that measurement bias constitutes systematic measurement error and therefore sources of variance that the test developer would prefer to avoid. Berry, Poortinga, Segall and Dasan (2002) nonetheless hold the important position that when cultural bias is uncovered the bias should be interpreted as systematic information about cross-cultural differences and should not be simply dismissed as measurement error. Berry et al. (2002) would then seem to attribute a “cultural meaning” to the bias. Whether this approach could be applied in a cross-gender study warrants discussion. It could be argued that this sentiment should apply to all measurement equivalence studies, irrespective of the nature of the groups being compared. If differences exist in the manner in which people from different groups (irrespective of whether it would be language, gender, cultural or age groups) respond to the same test stimuli, this is presumably because of one or more latent variables that systematically differ across the groups in question. Exploring the reasons for measurement bias (or more broadly, lack of

[17] Van de Vijver and Tanzer (1997) note that bias has to do with the characteristics of an instrument in a specific cross-cultural / group comparison, not with the instrument’s intrinsic properties. That is, the question as to whether an instrument is biased cannot be answered in general terms. The question of bias is examined and addressed in a specific comparison (e.g. across gender groups).

measurement invariance) should therefore be encouraged as a way of gaining greater understanding of group differences. In a similar vein, Cheung and Rensvold (2002) argue that invariance studies should not be conducted simply to justify the direct comparison of observed scores and observed means across groups but also to shed light on the manner in which the groups being compared differ in the manner in which they respond to test stimuli. Investigations of any form of measurement equivalence should be seen as a source of potentially interesting and valuable information about how different groups view the world.

At the same time, however, the earlier argument should be kept in mind that in an applied context, information on a specific latent variable carrying a specific constitutive definition is required for decision-making. Measurement bias would from this perspective be considered undesirable.

Group-related systematic measurement error could, for example, be due to poor item sampling in the respective cultures/groups, disparate domain sampling across different groups, or type and consistency of administration procedures, amongst other factors. Van de Vijver and Poortinga (1997) developed a taxonomy to describe different types of bias that should be identified prior to making cross-cultural comparisons. The taxonomy had originally been developed to provide a meaningful framework in which cross-cultural researchers could classify the different forms of measurement bias that they encountered in their studies (Van de Vijver & Leung, 1997, Van de Vijver & Poortinga, 1997, Van de Vijver & Tanzer, 1997). The types of bias distinguished in the taxonomy, however, effectively generalises to groups of a different nature (i.e., gender) as well. These types of bias are called; construct, method, and item bias.

4.2.1 Construct bias

Construct bias occurs when the construct measured by the instrument is not identical across groups (Van de Vijver & Leung, 1997). The number of underlying latent variables (or factors) that needs to be assumed to satisfactorily account for the covariance in the test responses of different groups, the degree to which the latent variables are inter-related and/or the specific latent variables that underlie the observed responses to each test stimulus (or item) therefore differ across groups. Stated more concisely, construct bias exists if the factor structure that is required to closely reproduce the observed covariance matrix differs across groups in terms of number of factors, correlation between factors and/or loading pattern. As argued in the previous paragraph, an important question to consider is why different (factor analytic) explanations would be required to account for the manner in which people from different groups respond to the, objectively speaking, same set of test stimuli. Essentially, this means that behaviours that serve as denotations of a specific construct in one group do not do so in another group. Inadequate domain sampling of the behaviours in different groups is another potential source of construct bias. A further cause may be due to an instrument not capturing complete coverage of a construct's sub-domains (Van de Vijver & Poortinga, 1997). Upon review of the literature related to the CISS, the researcher was unable to source construct bias research on the CISS.

4.2.2 Method bias

Whilst it is important to ensure that construct bias is eliminated, other nuisance factors may still threaten the validity of the inferences derived from the observed test scores or the instrument's measurement properties. Four common sources of method bias include (i) differential social desirability (DSD), (ii) differential response styles (DRS), (iii) differential stimulus familiarity (DSF) and the (iv) lack of comparability of samples (Berry et al., 2002; Byrne & Watkins, 2003).

It is possible that members of one group, therefore, tend to systematically respond in a more socially desirable manner to test stimuli than members of another group (i.e., DSD), could be more (or less) familiar with the test stimuli than members of another group (i.e., DSF), tend to favour certain response alternatives more (or less) than members of another group (i.e., DRS) or tend to be systematically different to members of another group on (non-relevant) characteristics that are related to test responses, For example: acquiescence or extreme rating[18] may be more prevalent in particular (cultural) groups which would threaten the validity of results (Van de Vijver & Leung, 1997). This is because such response styles may be driven by culture (for example: courtesy bias identified by Skeran (1983) in Asian cultures). Social desirability refers to the tendency to want to

[18] Acquiescence Response Style is also known as agreement bias, regardless of item content (Johnson, Kulesa, Cho, & Shavitt, 2005). Extreme Response Styles is the tendency to use the extreme ends of a rating scale (Cheung & Rensvold, 2002).

present a favourable impression of oneself when responding to questionnaire items in terms of prevailing norms (Edwards, 1957; Edwards 1970). In interest questionnaires there may be the possibility that males may respond negatively to “female” careers and females may respond negatively to “male” careers – not because of disinterest but for fear of displaying their own interest in an opposite gender’s stereotypical career. Therefore, this could be viewed as socially desirable responding in each gender group.

A major source of nuisance would be a group’s familiarity with a stimulus that is used to assess a particular domain. For example: one group is familiar with written comprehension type exercises versus another group that is au fait with demonstrating comprehension through dialogue or pictures. If both groups were presented with a written exercise to demonstrate comprehension of a story, the group familiar with written comprehension exercises is likely to perform better than the other group due to stimulus familiarity (Van de Vijver & Leung, 1997). With regards to typical performance measures/questionnaires that require degrees of preference to be indicated perhaps through a Likert-type scale, differential stimulus familiarity may occur in groups that have not been exposed to a Likert-type scale (Berry et al., 2002). This is more likely for groups that have not participated in psychological assessment or survey completion. Due to the high literacy level required for the CISS it would then be assumed that respondents to the CISS, which makes use of a 6 point Likert scale, have been exposed to some form of Likert-type response scales.

Differences in physical conditions during the administration session could lead to method bias. This would imply that differences in administration sessions for different groups may have a bearing on the results. Examples may include inadequate lighting, paper-and pencil versus computerised administration or unstandardized testing conditions (Van de Vijver & Leung, 1997).

A further source of bias may be due to the interaction between test administrator and the respondent. This could be due to communication problems in the case of a test administrator’s use of language, understanding of cultural norms or personal bias against the group being tested (Van de Vijver & Leung, 1997). The previous two cases could then also be viewed as DSF as all candidates do not have the same knowledge of the testing experience when administration sessions are varied in application.

While the discussion in the previous two paragraphs may be relevant when addressing method bias, it is assumed that the testing conditions would have been in accordance with the standardized conditions specified in the administration manual as trained professionals[19] are the custodians of the assessment process. As far as the delivery approach is concerned the CISS data gathered for this study was through a paper-and-pencil method (N. Taylor, personal communication, 19 September, 2007).

An additional source of bias could be attributed to the lack of comparability of the samples on other factors than the construct being assessed (i.e., biographical and demographic variables). Ideally the samples used in the analyses should be reasonably comparable in terms of the biographical and demographic characteristics of the samples (Berry et al., 2002; Byrne & Watkins, 2003). The nature of the samples utilised in this study is discussed in the next chapter.

4.2.3 Item bias

Item bias refers to undesirable measurement artefacts at the item level (Van de Vijver & Leung, 1997). This is often referred to as differential item functioning (DIF). DIF can be said to exist if group membership explains variance in the observed item response (either as a main effect and/or in interaction with the latent variable being measured) that is not explained by the latent variable being measured. If removing biased items eliminates group differences on the scale, the groups may have differed because of DIF rather than from inherent group differences (Van de Vijver & Leung, 1997).

From a different perspective, item bias could be said to exist if the probability of achieving a specific observed score on the item would differ across groups for individuals with the same standing on the latent variable being measured. Item bias would exist if the regression of the observed response on the latent variable being measured would differ across groups in terms of slope and/or intercept. The former situation is referred to as non-uniform bias (Van de Vijver & Leung, 1997) and would imply a group x latent variable

[19] In South Africa, a trained professional is an individual that is registered with the Health Professionals Council of South Africa (Professional Board for Psychology) as being permitted to carry out psychological assessment. These registered individuals are required to complete academic, supervision and examination requirements prior to registration and consequent access to psychological tests/questionnaires (Health Professions Act, 1974). The local test distributor adheres to this legal requirement by selling the CISS to registered professionals that may assess psychological constructs only (N. Taylor, personal communication, 19 September, 2007).

interaction effect on the observed item response, whereas the latter situation is known as uniform bias and would imply a group main effect on the observed item response (Van de Vijver & Leung, 1997). As argued in paragraph 4.2 above an important question to consider is why the regression of the observed test response on the latent variable would differ across groups. Sources of item bias include poor translation of items, inadequate item formulation (using complex wording, double-negatives, idiomatic expressions), items that tap into other constructs and appropriateness of item content for the target group (understanding of item content for the testing context). As indicated previously, the authors of the CISS state that they have taken steps to ensure that items, particularly occupational titles, are gender neutral. However, the researcher was not able to source literature on previous studies that investigated the presence of DIF in the CISS items.

4.3 EQUIVALENCE OR INVARIANCE IN MEASUREMENT

Bias influences the comparability of scores across cultures/groups: that is, the measurement implications of bias for comparability are addressed in the concept of equivalence (Van de Vijver, 2003). Construct and item bias express themselves in differences in measurement model parameters across groups. These two forms of bias in essence are defined in terms of the manner in which the measurement model underlying the test differs across groups. Measurement equivalence or invariance (or rather the lack thereof) therefore presents a different perspective on errors in measurement. Measurement invariance articulates error in measurement in different terms but essentially refers to the same issue as that discussed above under the headings of construct and item bias.

A somewhat contentious issue would be the question of how method bias relates to measurement equivalence/invariance. It could be argued that method bias does not translate to unique problems with measurement model characteristics that are not already covered by the concepts of construct and item bias. Unlike construct and item bias, method bias cannot be defined in terms of differences in measurement model characteristics[20]. Rather method bias, when uncovered, embodies the plea of Berry et al. (2002) and Cheung and Rensvold (2002) that attempts should be made to understand

[20] Reference is made here of measurement model characteristics rather than measurement model parameters so as to accommodate differences in the number of latent variables and loading patterns and not only differences in Λ, Φ or Θ_δ.

construct and item bias (i.e., differences in measurement model characteristics) when it occurs.

By establishing the absence of bias, the I/O psychologist may place more confidence in the validity of results and comparisons/inferences that are made based on questionnaire/test results. While different types of bias have been indicated in the previous section, the measurement level issues have not been fully dissected. Equivalence or invariance is the key issue of investigation when wanting to understand bias in measurement. Equivalence evidence indicates the absence of factors that challenge the validity of cross-group comparisons.

Van de Vijver and Leung (1997) indicate different hierarchical levels of equivalence that must be met prior to making direct comparisons between different groups, in this study, a comparison between genders. These levels are construct, measurement unit (or metric), and scalar (or scale origin) equivalence.

4.3.1 Construct equivalence

Construct bias (Van de Vijver & Leung, 1997) refers to differences in the number of factors required to satisfactorily account for the observed covariance matrix (i.e., Λ), the relationship between the factors (i.e., Φ) and the nature of the loading pattern (i.e., Λ). Van Herk, Poortinga and Verhallen (2004) note that evidence for the presence of structural equivalence (i.e., the same factor structure satisfactorily accounts for the observed covariance matrix cross-culturally – which could also to apply to different groups such as gender) indicates the absence of construct bias in the scores. Cheung and Rensvold (2000, see p.192) do not necessarily agree with this statement. According to them, this type of non-invariance (i.e., construct bias) cannot be detected statistically. This is because structural equivalence may be obtained even if a particular set of items is conceptually adequate in one group, but not completely adequate in another. Vandenberg (2002) also supports this view when he admits that if it is not known to what extent a strong method artefact as a characteristic of a given study this influences the viability of tests of equivalence. At this level of measurement, the influence of construct bias is difficult to establish.

Construct equivalence/structural equivalence is present when the same number of factors is required to satisfactorily account for the observed covariance matrix, the relationships between the factors are the same and the nature of the loading pattern is the same across groups. However, the construct may not necessarily be operationalized equivalently (Van de Vijver & Leung, 1997). For example: different groups may attach disparate meanings to different test stimuli, therefore the test stimuli evoke different conceptual frames of reference in each of the comparison groups (Riordan & Vandenberg, 1994; Vandenberg & Lance, 2000).

A hypothetical example in this study: both males and females may understand the construct of medical interests. However, females may see nursing and doctor-type activities as being one interest, whereas men may view nursing as a sub-element of medical interests with doctor-type activities operationalizing the interest better. Yet both groups associate these two professions within the medical arena. Van de Vijver and Leung (1997) recommend that the nomological network of the instrument should be compared across the two groups to gain an understanding of the structure of the construct.

Construct/structural equivalence is synonymous with the term configural invariance used by Vandenberg and Lance (2000).

4.3.2 Measurement unit equivalence

The next (higher) level of equivalence implies equivalence of the measurement unit used (Van de Vijver & Leung, 1997). This would then imply that the scale is calibrated in the same manner across groups. Fixed increases in the latent variable result in the same increases in item and scale scores across groups. This level of equivalence exists if the slope of the regression of the observed item response on the latent variable measured by the item is the same across groups (Vandenberg & Lance, 2000). However, this level of equivalence does not imply that the measurement scales have the same origin (or starting point on the scale). This would imply possible inequality of intercepts. Measurement unit equivalence, within the Van de Vijver and Leung (1997) framework is similar to metric equivalence (Hair et al., 2006; Vandenberg & Lance, 2000) which tests for the equality of scaling units. Lack of metric equivalence therefore would imply that the measures of people with the same standing on the latent variable differ across groups but that they are a linear function of each other. The linear function, however, would have to specify both intercept and slope.

Scaling units would be the same (i.e., metric invariance would be found) if the of the slope of regression of the observed item response on the latent variable measured by the item is the same across groups. Equal differences in the latent variable would then translate into equal differences in the observed item score across groups even though the magnitude of the item scores would be different across groups. An example provided by Van de Vijver and Leung (1997) describes that both the Celsius and Kelvin scales measure temperature with the degrees unit. Three different objects differing in temperature with a fixed amount would result in three different Kelvin and three different Celsius scores. The differences in the three Kelvin and the three Celsius scores would, however, be the same. However, the origin of the scales differs – for Celsius water's freezing point is 0 where as with the Kelvin scale it is 273. The temperature readings of the three objects would therefore differ on the two scales. By making an adjustment to either of the scales a conversion is possible (Van de Vijver & Leung, 1997) via a linear function that does not require fixing the intercept to zero.

The question arises as to why the relationship between item response and latent variable would be different for members of different (gender, language, cultural) groups? An answer need not necessarily be readily available. However, as mentioned in the section on method bias, the manner in which test takers respond to test items may be affected by (i) DSD, (ii) DRS, (iii) DSF and (iv) the lack of comparability of samples. More concisely, any of these effects may be the cause of measurement unit equivalence. Van Herk et al. (2004) emphasise this point by indicating that lack of metric invariance could be due to group-related differences in acquiescence response style (ARS) or extreme response style (ERS) or both. Cheung and Rensvold (2000) are specific in suggesting that lack of metric invariance is due to group differences in ERS. It was stated previously that ARS and ERS are specific forms of DRS.

The possibility should be considered that some of the slopes of regression of the observed item response on the latent variable measured by the item are the same across groups while others may not be. Byrne, Shavelson and Muthén (1989) proposed the concept of partial measurement/metric invariance to make provision for this eventuality. In their research they found that it is in fact possible to find that the slope may be different for some items in certain groups but not in others. While “full” metric invariance is ideal, practically partial metric invariance could be considered sufficient when wanting to compare groups in terms of the equivalence of structural model parameters (but not when wanting to compare latent means across groups). However, this implies a relaxation of the constraints placed on the model being tested for equivalence. Vandenberg and Lance (2000) do indicate that partial metric invariance may be controversial as the statistical criteria for relaxing constraints are not consistent in the literature and a lack of consensus on their application has been found. Hair et al. (2006) do indicate that partial metric invariance exists when at least two items loading on each factor have equivalent factor loadings/slopes across groups. Vandenberg and Lance (2000) do, however, indicate that scalar invariance tests could be investigated for those indicators that demonstrate partial metric invariance.

4.3.3 Scalar equivalence or full score comparability

Van de Vijver and Leung (1997) indicate that the highest level of equivalence is achieved when the measurement instrument is on the same ratio scale in each (cultural) group. This type of equivalence can only be achieved when the regression of the item response on the underlying latent variable has the same slope across groups as well as the same origin (Vandenberg & Lance, 2000). Only when the regression of the item response on the latent variable coincides in terms of slope and intercept will the same standing on the latent variable of individuals in different groups result in the same expected observed scores on the item. Only if regression of the item response on the latent variable coincides in terms of slope and intercept are equal then equal observed scores of individuals from different groups can be interpreted as indicating an equal standing on the latent variable. Only if regression of the item response on the latent variable coincides in terms of slope and intercept are equal can equality of observed score means across groups (or differences) be interpreted as indicating equality in latent means (or differences; Vandenberg & Lance, 2000). The same argument applies for observed scale scores.

The question could be asked whether scalar equivalence should not possibly be defined in terms of a correspondence in the intercepts of the regression of the item response on the latent variable only. A correspondence in intercepts but a difference in slope would again mean that a conversion from one scale to the other would be possible provided that the linear transformation formula would be known in which the slope would be the critical parameter. For example: if one wishes to determine the height of a number of people, yet different units are used (metric versus imperial), a direct comparison could not be made. However it is known that height is being measured and it would be known that the two scales share the same origin. The term scalar invariance could be used to refer purely to the fact that the regression of the item response on the latent variable coincides in terms of intercept. The fact, however, remains that only if the regression of the item response on the latent variable coincides in terms of both intercept and slope (whatever term would be used to indicate this situation) can differences in item response across groups be interpreted to reflect differences in latent variable standing.

Van de Vijver and Leung (1997) also remark that scalar equivalence is sometimes claimed when *only* construct equivalence has been established. Construct equivalence may be satisfied, however, the relationship between item response and the underlying latent variable may not be the same across groups.

4.4 INVESTIGATING MEASUREMENT INVARIANCE/EQUIVALENCE

While Van de Vijver and others have dissected the problem of measurement bias and provided a useful taxonomy for the cross-cultural/group psychology, issues regarding equivalence in an organisational setting have been attended to by Vandenberg and Lance (2000). In their thorough review of the measurement equivalence/invariance (ME/I) literature, Vandenberg and Lance (2000) report that there seems to be inconsistency in the terminology and methodology used in evaluating ME/I. Therefore, a goal in their article is to consolidate the knowledge of the ME/I topic and propose a methodological framework that can be used consistently for the investigation of ME/I in the organisational sciences.

Their argument begins by stating that historically the evaluation of measurement quality has arisen through applying the classical measurement theory's (Crocker & Algina, 1986; Lord & Novick, 1968; Nunnally & Bernstein, 1994) distinction between true and error scores in the observed score of the variable of interest to the researcher. This paradigm has been used for a great part of the 20th century through reliability studies that determine the consistency of respondents answering a given questionnaire. However, simple reliability studies tend to ignore the issue of equivalent models of measurement – a group can consistently respond in a biased fashion to items aiming to measure a variable that "supposedly" applies across groups.

Van de Vijver and Leung (1997) indicate that the most frequently applied technique for addressing the various facets of measurement equivalence (i.e., construct, measurement unit and scalar) is exploratory factor analysis (EFA), which is then followed by target rotations and the computation of an index of factorial agreement across groups. Factorial agreement could be assessed through Tucker's Phi (Tucker, 1951). In simpler terms, factor-analytical solutions are derived from the testing of each group separately and scrutinising similarities in the solutions. This technique would seem to limit the researcher's ability to test the hypothesized underlying model of measurement due to the exploratory nature of the technique. EFAs would only investigate the degree to which items load similarly or "hang together" – it tests the degree to which a particular group may identify with the construct that is measured; it does not directly test the model as specified by the test publisher. As Van de Vijver and Leung (1997) have indicated, an investigation into the nomological network of the construct would need to be conducted to establish construct equivalence[21].

Confirmatory factor analysis, which is an application of the SEM approach, is unlike EFA in that it aims to cast factor analysis in the tradition of hypothesis testing. It explicitly tests the overall quality of the factor solution and the specific parameters composing a model (Kelloway, 1998). CFA allows for the testing of an *a priori* model, as developed by a test publisher, by fitting data to the model and establishing levels of fit. CFA analysis can be described mathematically in terms of equation 1 below:

$$X = \tau + \Lambda\xi + \delta \text{--} (1)$$

where X is a q x 1 column vector of observed item responses, τ (tau), is a q x 1 column vector of intercept terms, Λ is a q x n^{th} factor loading matrix, ξ is a n x 1 column vector of latent constructs, and δ is a q x 1 column vector of measurement error terms (Meade & Lautenschlager, 2004; Vandenberg & Lance, 2000). Equation 1, however does not fully

[21] This remains true for any procedure used to investigate measurement invariance.

capture the measurement model in that it fails to specify the manner in which the latent variables are related and the manner in which the measurement error terms are related. A statement on the nature of the Φ and Θ_δ matrices is therefore also required. Essentially the argument in ME/I through the CFA paradigm would be to evaluate the equivalence of the above equation (as well as the Φ and Θ_δ matrices) across different groups. This would mean testing whether (a) the conceptual equivalence of the underlying theoretical variable (the latent ξ) in each group, (b) equivalent associations (λ and τ) between operationalizations (X) and ξ across groups and (c) the extent to which the Xs are influenced to the same degree and perhaps by the same unique factors (δ) across groups.

If by assuming $E\,(\xi, \delta_k) = 0$, then the covariance equation that follows from Equation (1), is:

$$\Sigma^g = \Lambda^g{}_X \Phi^g \Lambda^{g'} + \Theta^g{}_\delta \text{---} \quad (2)$$

where Σ^g is the matrix of variances and covariances among the q items in the g^{th} population (group), $\Lambda^g{}_X$ is the matrix of items' factor loadings on ξ^g, Φ^g contains variances and covariances among the ξ^g, and $\Theta^g{}_\delta$ is the diagonal matrix of unique variances. The inclusion of the τ vector in Equation 1 is really an extension of the basic model in which the intercepts are assumed to be zero and are thus not estimated (Vandenberg & Lance, 2000).

Vandenberg and Lance (2000) propose that to operationalize the hypotheses suggested by the above equations the following hypotheses should be tested to establish ME/I:

- $\xi^g = \xi^{g'}$, that is, that the set of q items evokes the same conceptual framework in defining the latent construct (ξ) in each comparison group.
- $\Lambda^g = \Lambda^{g'}$, that is, that the regression slope linking the manifest measures $X^g{}_k$ to the underlying construct(s) are invariant across groups.
- $\Theta^g{}_\delta = \Theta^g{}_\delta$, that is, that unique variances for like $X^g{}_k$s are invariant across groups.
- $\Phi^g = \Phi^{g'}$, that is, that variances and covariances among the latent variables are invariant across groups.

The question should however be asked whether the hypothesis that $\tau^{g} = \tau^{g'}$ should not also be explicitly tested. The discussion on uniform item bias and the previous discussion on scalar invariance would suggest that it would be imperative to also examine differences in the tau vectors across groups (in conjunction with differences in factor loadings across groups).

After conducting a thorough literature review of studies relating to the ME/I issue, Vandenberg and Lance (2000) report that most researchers use multi-sample applications of CFA to establish ME/I. A step-by-step process emerged from the review of previous studies. Researchers commonly perform an omnibus test of covariance equality ($\Sigma^{g} = \Sigma^{g'}$) first. If the omnibus test indicates that there are no differences in covariance matrices between the data sets then full ME/I holds for the data. Meade and Lautenschlager (2004) report that the efficacy of this test has been questioned by authors (Jöreskog, 1971; Raju, Lafitte & Byrne, 2002; Rock, Werts & Flaugher, 1978). They argue that the omnibus test would indicate that ME/I is reasonably tenable when a more specific test of ME/I may find otherwise. Regardless of the results of the omnibus test, a series of nested model chi-square difference tests can be performed to determine possible sources of differences (Meade & Lautenschlager, 2004; Vandenberg & Lance, 2000).

To perform these nested model tests, both data sets (in this case male and female) are tested simultaneously holding only the pattern of factor loading invariant, while other parameter estimates are allowed to vary between groups – this test is referred to as a test of “configural invariance” or “weak factorial invariance” as described by Horn and McArdle (1992). For subsequent tests to be meaningful, evidence of configural invariance must be established. In the Van de Vijver and Leung (1997) framework this would imply construct (structural) equivalence. This model becomes a baseline model for which a foundation chi-square value is calculated – this chi-squared value is then compared to the chi-squared value attained when fitting models on which more stringent parameter constraints have been imposed in subsequent tests of ME/I. The next test is that of factor loading invariance across groups ($\Lambda^{g}_{k} = \Lambda^{g'}_{k}$), also known as a test of “metric invariance” (Horn & McArdle, 1992). This could be viewed as measurement unit equivalence as postulated by Van de Vijver and Leung (1997). The difference in the baseline and the more restricted model is expressed as a chi-squared statistic with degrees of freedom equal to the number of freed parameters. Next, Vandenberg and Lance (2000) recommend a test of the equality of item intercepts ($\tau^{g}=\tau^{g'}$), also known as a test of "scalar invariance" (Meredith, 1993; Steenkamp & Baumgarter, 1998). The test of scalar invariance would only be practically meaningful if the previous test of metric invariance has shown the slope of the regression of X on ξ to be the same across groups. This would imply equivalent scale origins described as scalar equivalence by Van de Vijver and Leung (1997). Other tests that could be conducted include tests of latent means ($\kappa^{g}=\kappa^{g'}$), tests of equal item uniqueness terms and tests of factor variances and covariances – each of which are constrained in a specific model under a specific measurement invariance null hypothesis. This again illustrates the fact that measurement invariance/equivalence is a much broader concept than measurement bias but one that can successfully accommodate all conceptualizations of measurement bias.

4.4 RESEARCH QUESTIONS

Horn and McArdle (1992) state (as cited in Vandenberg & Lance, 2000, p. 9) that "the general question of invariance of measurement is one of whether or not, under different conditions or observing and studying phenomena, measurements yield measures of the same attributes" – this is a necessary prerequisite to making any comparisons between groups. This discussion has argued that the unaddressed issue of gender invariance in interest measurement may impede the practical utility of such assessments in the workplace. The developers of the CISS hold that the instrument has made great strides in improving the overall measurement of interest in the most unbiased way possible. This was done by using modern items with careful elimination of stereotype meaning in the wording. Yet measurement invariance hypothesis testing, in the purest sense of CFA, has not been investigated in South Africa for this instrument. By using a multi-group CFA SEM approach, the researcher aims to answer the following central research question:

"Do the measurement models, implied by the design intentions of the developers of the CISS, fit data obtained for a South African sample on the instrument and are the measurement model parameters equivalent across gender subsamples?"

In order to answer this broad question, the following specific research questions should be answered in the order presented:

1. When fitting the interest and skill measurement models to the combined sample, do the models fit the data adequately?
2. When independently fitting the interest and skill measurement models to separate gender samples, do the models fit the data adequately?
3. When fitting the measurement models to the separate gender samples simultaneously, do the models fit adequately when holding only the pattern of factor loadings invariant, while all measurement model parameter estimates are allowed to vary between groups (configural invariance)?
4. When fitting the measurement models to the separate gender samples simultaneously, with all measurement model parameters constrained to be equal across groups, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (omnibus test of measurement invariance)?
5. When fitting the measurement model to the separate gender samples simultaneously with only the factor loadings constrained to be equal across groups but all other model parameters estimated freely, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (lack of metric invariance)?
6. When fitting the measurement model to the separate gender samples simultaneously, with the items (or item parcels) intercepts constrained to be equal across groups but all other model parameters estimated freely, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (lack of scalar invariance)?

The relevance of questions 3 to 5 is dependent on the answer obtained to the question that immediately precedes them. If the interest and skill measurement models fit the total sample, the assumption would be that the models should also fit within the separate gender groups. It would, however, be important not to simply assume this but rather to empirically test this assumption. If the interest and skill measurement models do not fit the total sample the question arises whether this could be attributed to the fact that the models fit the data of one gender but not the other or whether it is due to the fact that the models fit neither gender. If the interest and skill measurement models independently fit the separate gender samples it would be necessary to formally confirm that the same number of factors required to satisfactorily account for the observed covariance matrix and the nature of the loading pattern are the same across groups (question 3)[22]. If configural invariance would be shown, a legitimate question to ask would be whether any measurement model parameters differ across the two gender groups (question 4). Only if the omnibus test of measurement invariance would indicate a lack of measurement equivalence does it make sense to pose the question whether the slope of the regression of the items on the latent interest/skill dimension they are meant to represent differs across gender groups (question 5). The last question is aimed at establishing whether it would be permissible to interpret equal observed scores of individuals from different groups as indicating an equal standing on the latent variable. This would only be permissible if the regression of items on latent variables would coincide in terms of slope and intercept. It therefore would make sense to proceed with the testing of intercept differences (question 6) only if metric equivalence would be indicated.

The next chapter aims to describe how these research questions will be operationalized and practically answered.

[22] Question 3 could have directly followed on a finding of close model fit on the total sample.

CHAPTER 5
RESEARCH METHODOLOGY AND PRELIMINARY DATA ANALYSES

5.1 INTRODUCTION

The fundamental hypothesis being tested in this investigation is that the CISS measures the interest construct as constitutively defined and that the construct is measured in the same manner across gender groups. To determine the validity of this hypothesis requires a series of confirmatory factor analyses (CFAs) in which the fit of the implied measurement model is evaluated. The validity and credibility of the verdict on the legitimacy of these claims depends on the methodology used to arrive at the verdict. The methodology serves the epistemic ideal of science (Babbie & Mouton, 2001). If the methodology would be flawed the chances of the arriving at a valid conclusion on the measurement invariance of the CISS would be jeopardized. The credibility of the verdict on the appropriateness of using the CISS across gender samples in South Africa would thereby suffer.

To ensure that the epistemic ideal of science is met, the method of inquiry used in a study should be subjected to critical inspection by knowledgeable members of the scientific community in which the research is being performed (via publication and conference presentations). In this sense, science could be said to be rational (Babbie & Mouton, 2001). Scientific rationality can, however, only serve the epistemic ideal of science if the method used in the scientific inquiry is comprehensively described, and if the methodological choices that have been made are thoroughly motivated. In this chapter the proposed research methodology is therefore explained and motivated. Specific attention is focussed on the research design, statistical hypotheses, statistical analysis techniques, and the nature of the sample.

5.2 RESEARCH HYPOTHESES

The substantive hypothesis tested in this study is that the CISS provides a valid and reliable measure of the interest construct as defined by the instrument, and that the construct is measured in the same manner across gender groups.

The substantive hypothesis would ideally translate to the following specific operational hypotheses:

- The first-order Basic scales interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels[23] (formed from the items) comprising each of the Basic scales in the combined sample.
- The second-order interest measurement model[24] implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels (formed from the items) comprising each of the Basic scales in the combined sample.
- The first-order Basic scales interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels (formed from the items) comprising each of the Basic scales in the separate gender samples.
- The second-order interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels (formed from the items) comprising each of the Basic scales in the separate gender samples.
- The first-order Basic scales interest measurement model implied by the scoring key of the CISS displays configural invariance across the two gender samples.
- The first-order Basic scales interest measurement model implied by the scoring key of the CISS displays full measurement invariance across the two gender samples.
- The first-order Basic scales interest measurement model implied by the scoring key of the CISS displays metric invariance across the two gender samples

A corresponding substantive hypothesis and operational hypotheses could be formulated with regards to the self-reported skill construct. Due to time and logistical constraints these hypotheses will, however, not be investigated as part of the current study.

It would be important to highlight an important limitation with regards to sample size for this study. Due to the complexity of the CISS measurement models, a CFA can be conducted

---

[23] The formation of item parcels will be motivated and explained in sections 5.7.1 & 5.7.2.

[24] The second-order measurement model maps the latent Orientation interests as higher-order factors on the first-order latent Basic interests.

on the Basic Interest scales using the combined sample, however the ratio of parameters to be estimated versus sample size would not yield credible solutions when conducting independent CFAs of the Basic Interest scales model for each gender sample separately. This is due to the number of parameters to be estimated exceeding the number of observations in each gender sample. Therefore, the possibility of conducting the subsequent independent sample CFAs and measurement invariance tests is not possible with the present sample size. Further detail of the samples is provided in section 5.5.

An alternative strategy for the present study would be to work with the Orientation scales models as the ratio of parameters to be estimated versus sample size would yield a credible solution [this is described in more detail in the model specification & identification sections (5.6.1 to 5.6.4) of this chapter. Therefore a CFA would need to be conducted on the Orientation scales of the CISS for the combined sample, and subsequent independent sample CFAs could then be conducted on the separate gender samples. Should these analyses show adequate fit, then measurement invariance tests would be conducted on the Orientation scales.

As the Basic scales are the primary factors of the CISS models, the combined sample CFA would be conducted to validate the Basic scales measurement models. Should adequate fit be found at this level of measurement, further data fit analyses would be conducted at the global factor level as the Basic scales provide the basis of measurement at the Orientation scale level.

Given the above argument regarding sample size constraints, the foregoing operational hypotheses were revised as follows:

- The first-order Basic scales interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels[25] formed from the items comprising each of the Basic scales in the combined sample.

[25] The formation of item parcels will be motivated and explained in sections 5.7.1 & 5.7.2.

- The Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the combined sample.
- The Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the separate gender samples.
- The Orientation scale interest measurement model implied by the scoring key of the CISS displays configural invariance across the two gender samples.
- The Orientation scale interest measurement model implied by the scoring key of the CISS displays full measurement invariance across the two gender samples.
- The first-order Orientation scales interest measurement model implied by the scoring key of the CISS displays metric invariance across the two gender samples

5.3 RESEARCH DESIGN

The hypotheses formulated under 5.2 make specific claims with regards to the CISS interest measurement model. The interest measurement model implied by the scoring key of the CISS hypothesizes specific measurement relations between the items comprising the instrument and the interest dimensions measured by the instrument. More specifically, the measurement model assumes that the slope of the regression of specific indicator variables (X) on the specific latent variable (ξ) the indicator variable is meant to represent is positive and significantly greater than zero. In addition, the measurement model makes assumptions about the covariance between the latent variables and the covariance between the measurement error terms. The hypotheses formulated under 5.2, moreover, assumes that the measurement model parameters remain invariant across the two gender groups.

To empirically test the merit of the assumptions made by the measurement model requires a plan or strategy that will guide the gathering of empirical evidence to test the operational hypotheses. The research design constitutes this plan or strategy (Kerlinger & Lee, 2000). The function of the research design is to attempt to ensure empirical evidence that can be interpreted unambiguously for or against the operational hypotheses.

This study will use a correlational *ex post facto* research design. In terms of the logic of the *ex post facto* correlational design, the researcher observes the observed variables[26] and calculates the covariance between the observed variables (Kerlinger & Lee, 2000). Estimates for the freed measurement model parameters are obtained in an iterative fashion with the purpose of reproducing the observed covariance matrix as accurately as possible (Diamantapoulos & Siguaw, 2000). If the fitted model fails to accurately reproduce the observed covariance matrix (Byrne, 1989; Kelloway, 1998), the conclusion would inevitably follow that the measurement model underlying the CISS does not provide an acceptable explanation for the observed covariance matrix. Therefore this would then signal that the CISS does not measure the interest domain as intended in the South African sample. The converse, however, is not true. If the covariance matrix derived from the estimated model parameters closely corresponds to the observed covariance matrix it would not imply that the processes postulated by the measurement model necessarily produced the observed covariance matrix, and that the CISS therefore measures the interest domain as intended. A high degree of fit between the observed and estimated covariance matrices would only imply that the processes portrayed in the measurement model provide one plausible explanation for the observed covariance matrix.

Essentially the same line of reasoning would also apply with regards to the measurement invariance hypotheses. If the model fitted simultaneously to the two gender samples with all parameters estimated freely would fail to accurately reproduce the observed covariance matrices, the conclusion would inevitably follow that the measurement model underlying the CISS does not provide an acceptable explanation for both of the observed covariance matrices. Consequently that the CISS does not measure the interest in the same manner in the two South African gender samples.

5.4 STATISTICAL HYPOTHESES

The nature of the statistical analyses that will be used to test the operational hypotheses will necessarily affect the decision as to whether statistical hypotheses should be formulated and the format in which they will be formulated. One possibility would have been to use an unrestricted, exploratory factor analytic approach in which no *a priori* stance is taken on the number of factors underlying the observed covariance matrix, their

[26] These could be individual items or item parcels as linear composites of individual items.

identity and the manner in which the items load on the factors (Ferrando & Lorenzo-Seva, 2000). If this option would have been chosen, no statistical hypotheses would have been formulated. This option, however, seems inappropriate in as far as it ignores the design intentions of the developers of the CISS.

In the case of the CISS, a very specific stance is taken on the number of interest factors underlying the observed covariance matrix, their identity and the manner in which the items load on the interest factors. Interest items were explicitly and intentionally developed to reflect specific dimensions of interest construct. Specific CISS items were written to function as stimulus sets to which test takers would respond with behaviour which would be behavioural expressions of specific latent interest dimensions. The scoring key of the CISS reflect these design intentions (Campbell et al.,1992).

Therefore, it seems more reasonable towards the developers of the instrument to first evaluate the the question whether their intentional instrument design did succeed in providing a comprehensive and relatively uncontaminated empirical grasp on the interest construct as the CISS manual defines it. A hypothesis testing, restricted, confirmatory factor analytic approach should rather be followed. In terms of this approach specific structural assumptions are made with regards to the number of latent variables underlying the CISS, 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). Specific assumptions are, moreover, made on how these structural assumptions apply across the two gender groups.

Structural equation modelling utilizing LISREL (Jöreskog & Sörbom, 1996b) will be used to test the operational hypotheses listed in paragraph 5.2. Therefore, the following statistical hypotheses will be tested:

Operational hypothesis 1:
The first-order Basic scales interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the combined sample:

$$H_{01}: \text{RMSEA} \leq 0.05$$

$$H_{a1}: RMSEA>0.05$$

If H_{01} would not be rejected (i.e., close model fit would be found) or if at least reasonable model fit would be obtained (as indicated by the basket of fit indices produced by LISREL) the following 58 null hypotheses on the slope of the regression of item parcel j on latent Basic interest dimension k will be tested:

$$H_{0i}: \lambda_{jk}=0; i=2, 4, \ldots, 59; j=1, 2, \ldots, 58; k=1, 2, \ldots, 29$$
$$H_{ai}: \lambda_{jk}\neq 0; i=2, 4, \ldots, 59; j=1, 2, \ldots, 58; k=1, 2, \ldots, 29$$

These 59 hypotheses will form the basis for examining the merits of the claim made by the developers that the CISS successfully measures the 29 Basic interest dimensions it intends to measure and in the manner that it intends to do according to the scoring key.

Operational hypothesis 2:

The Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the combined sample.

$$H_{060}: RMSEA\leq 0.05$$
$$H_{a60}: RMSEA>0.05$$

If H_{060} would not be rejected (i.e., close model fit would be found) or if at least reasonable model fit would be obtained (as indicated by the basket of fit indices produced by LISREL) the following 58 null hypotheses on the slope of the regression of item parcel j on latent Orientation interest dimension k will be tested:

$$H_{0i}: \lambda_{jk}=0; i=61, 4, \ldots, 118; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$
$$H_{ai}: \lambda_{jk}\neq 0; i=61, 4, \ldots, 118; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$

These 15 hypotheses will form the basis for examining the merits of the claim made by the developers that the CISS successfully measures the seven global interest dimensions it intends to measure via the Orientation scales and in the manner that it intends to do according to the scoring key.

Operational hypothesis 3:

The Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the separate gender samples.

$$H_{0119M}: RMSEA \leq 0.05$$
$$H_{0119F}: RMSEA \leq 0.05$$
$$H_{a119M}: RMSEA > 0.05$$
$$H_{a119F}: RMSEA > 0.05$$

If H_{0119M} would not be rejected (i.e., close model fit would be found) or if at least reasonable model fit would be obtained (as indicated by the basket of fit indices produced by LISREL), the following 58 null hypotheses on the slope of the regression of item parcel j on latent Orientation interest dimension k will be tested:

$$H_{0Mi}: \lambda_{jk}=0; i=120, 4, \ldots, 177; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$
$$H_{aMi}: \lambda_{jk}\neq 0; i=120\ 4, \ldots, 177; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$

If H_{0119F} would not be rejected (i.e., close model fit would be found) or if at least reasonable model fit would be obtained (as indicated by the basket of fit indices produced by LISREL), the following 58 null hypotheses on the slope of the regression of item parcel j on latent Orientation interest dimension k will be tested:

$$H_{0Fi}: \lambda_{jk}=0; i=120, 4, \ldots, 177; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$
$$H_{aFi}: \lambda_{jk}\neq 0; i=120, 4, \ldots, 177; j=1, 2, \ldots, 58; k=1, 2, \ldots, 7$$

Operational hypothesis 4:

The Orientation scale interest measurement model implied by the scoring key of the CISS displays configural invariance across the two gender samples:

H_{0178}: RMSEA≤0.05

H_{a178}: RMSEA>0.05

Operational hypothesis 5:

The Orientation scale interest measurement model implied by the scoring key of the CISS displays measurement invariance across the two gender samples:

H_{0179}: $SB\chi^2_{diff}=0$

H_{a179}: $SB\chi^2_{diff}>0$

Operational hypothesis 6:

The first-order Orientation scales interest measurement model implied by the scoring key of the CISS displays metric invariance across the two gender samples:

H_{0180}: $SB\chi^2_{diff}=0$

H_{a180}: $SB\chi^2_{diff}>0$

5.5 SAMPLING

The following section describes the nature, details and limitations of the samples used in the present study. It also aims to provide information in support of the decision to conduct the CFAs at the global (Orientation Interest) level of measurement.

The sample could be considered a non-probability sample of respondents comprising both genders from the population of South African test takers who have completed the CISS. The use of a non-probability sampling procedure means that the findings of this study can only be generalized to the target population with great circumspection. Records of the sample have been provided by the sole local questionnaire distributor in an anonymous format[27]. The majority of the records did not include biographical information for example: race, educational level, current occupation or first language. Much of the respondent information was not provided by questionnaire respondents upon completing the test session. The lack of biographical information is rather unfortunate as it prevents the proper characterization of the study sample. This is certainly a regrettable shortcoming in this study. As this research aims to determine the equivalence of the measurement model

[27] Written institutional permission had been obtained from the test distributor to utilize the data of the sample for the purpose of this research.

of the CISS across gender, the sample will be considered suitable for the proposed purpose as gender information has been provided by the questionnaire distributor for all selected cases.

The total sample consists of 810 respondents of which 408 (50.4%) are male and 402 (49.6%) are female. The purpose of assessment in most cases would have been career guidance, organisational succession planning, selection and other developmental actions (N. Taylor, personal communication, 19 September 2007).

As the sample information excludes other demographic details, it would be difficult to explore complexities regarding the subsets of each male/female sample. It would therefore be difficult to gain further insight into the impact that race, educational background and current occupation has on the data. However, the gender details are the primary concern in this study and will remain the focus of how the data will be handled. Also, as indicated previously, the sample sizes are ineffective in creating a unique solution when wanting to estimate the Basic Interest scales. This was described in the discussion in section 5.2.

5.6 PREPARATORY PROCEDURES

The operational hypotheses listed in paragraph 5.2 were tested by testing the statistical hypotheses presented in paragraph 5.4. The 180 null hypotheses were tested by utilizing SEM by means of LISREL (Jöreskog & Sörbom, 1996b). The following section aims to describe and motivate the initial procedures undertaken prior to conducting the SEM analyses. The section begins by specifying the respective models that would be subjected to confirmatory factor analyses. Thereafter, the identification of the measurement models is evaluated. The approach used in handling missing values is indicated. Finally, the necessity of performing item and dimensionality analyses is explained and the procedures described.

5.6.1 Specification of the Basic scales measurement models

The detailed specification of the measurement models, in SEM notation, is required to determine whether the relevant measurement models are identified. The specification provides a clear understanding of the model complexity as well as number of parameters to be estimated.

5.6.1.1 Basic interest measurement model specification.

$X = \Lambda^{x}\xi + \delta$---(3)

Where:

- X is a 58x1 column vector of observable *interest* indicator scores[28];
- Λ^{x} is a 58x29 matrix of factor loadings;
- ξ is a 29x1 column vector of first-order latent Basic Interest factors;
- δ is a 58x1 column vector of unique/measurement errors components comprising the combined effect on X of systematic non-relevant influences and random measurement error (Jöreskog & Sörbom, 1996b).

The foregoing measurement model implies two additional matrices. The first is a symmetrical 29x29 Φ matrix. This matrix describes covariance/correlations between each latent variable. The second matrix θ_{δ} is a 58x58 diagonal matrix - which would imply that the measurement error terms are assumed to be uncorrelated across the indicator variables. By freeing off-diagonals in this matrix, it would then imply that that error terms may be correlated indicating the possibility of additional common factors. Due to the confirmatory nature of this study, freeing the off-diagonals would be impossible to justify in terms of the design intentions of the developers of the instrument.

5.6.2 Basic scales model identification

In evaluating the identification of the model, the researcher is determining whether sufficient information is available to obtain a unique solution for the parameters to be estimated in the measurement model (Diamantopoulos & Siguaw, 2000).

Diamantopoulos, Siguaw (2000) and MacCallum (1995) make recommendations regarding an approach to model identification. An initial recommendation is that each latent variable

[28] The formation of item parcels will be described and motivated in paragraph 5.7.1

should be allocated a definite scale. The second recommendation is that model parameters to be estimated may not exceed the number of unique variance/covariance terms in the sample observed covariance matrix. For the interest model, the latent variables will be treated as a (0;1) standard variable (MacCallum, 1995), thereby satisfying the first recommendation.

5.6.2.1 Basic Interest model: parameter estimates vs. unique variance/covariance terms.

The number of model parameters set free to be estimated equal (t=522) which is less than the nonredundant elements in the observed sample covariance matrix [(p)(p+1)]/2= 1711] (Diamantooulous & Siguaw, 2000; MacCallum 1995, 1996). Degrees of freedom are then calculated by deducting the parameters to be estimated from the number of elements in the covariance matrix, resulting in 1189.

A CFA can be conducted on the Basic Interest scales using the combined sample. This is due to the size of the combined sample (*N*=810) exceeding the number of parameters to be estimated (t=522). However, this is not the case for the independent samples. The size of the male (*n*=408) and female (*n*=402) samples do not exceed the number of parameters to be estimated. Therefore, independent CFAs will not be conducted for the Basic Interest model. Although the sample size would not present a problem to fit a multi-group measurement model on the Basic Interest scales in which all parameters are constrained to be equal across gender groups (the situation would be essentially the same as in the single group analysis), it would present problems in the multi-group analysis in which all model parameters are estimated freely (i.e., the conditions under which configural invariance would be investigated).

A solution to the problem was sought by investigating whether a similar problem arises when the measurement model is fitted at the global level of measurement, the Orientation Interest scales. This is addressed next.

5.6.3 Specification of the Orientation scales measurement models

The item parcels that were constructed to represent the 29 Basic Interest scales would then be used as indicators of the seven Orientation scales. This is regarded as permissible due to the fact that the Orientation scales were constructed in order to summarize the data obtained through the Basic scales. It is, therefore, assumed that the corresponding Basic scale item parcels would be suitable indicator variables for the seven Orientation scales.

5.6.3.1 Interest measurement model specification.

$X = \Lambda^x\xi + \delta$-- (4)

Where:

- X is a 58x1 column vector of observable *interest* indicator scores;
- Λ^x is a 58x7 matrix of factor loadings;
- ξ is a 7x1 column vector of first-order latent Basic Interest factors;
- δ is a 58x1 column vector of unique/measurement errors components comprising the combined effect on X of systematic non-relevant influences and random measurement error (Jöreskog & Sörbom, 1996b).

The foregoing measurement model implies two additional matrices. The first is a symmetrical 7x7 Φ matrix. This matrix describes covariance/correlations between each latent variable. The off-diagonals, as with the model specified previously, would not be freed in the 58x58 θ_δ matrix.

5.6.4 Orientation scales model identification

As with the Basic scales, the Orientation scales model identification is a necessary prerequisite to performing CFAs.

5.9.4.1 Orientation Interest model: parameter estimates vs. unique variance/covariance terms

The number of model parameters set free to be estimated equal (t=137) which is less than the nonredundant elements in the observed sample covariance matrix [(p)(p+1)]/2= 1711] (Diamantooulous & Siguaw; MacCallum 1995, 1996). Degrees of freedom are then calculated by deducting the parameters to be estimated from the covariance matrix, resulting in 1574.

Fortunately, a combined sample CFAs as well as independent and multi-group CFAs can be conducted on the Orientation Interest model. This is due to the combined sample (*N*=810) exceeding the number of parameters to be estimated (t=137). For the independent CFAs the male (*n*=408) and female (*n*=402) samples also exceed the number of parameters to be estimated. Therefore, independent CFAs can be conducted for the Orientation Interest measurement model. In the case of the multi-group analysis in which all parameters are estimated freely across gender groups, the combined sample (*N*=810) still exceeds the number of parameters to be estimated (t=274). Therefore, combined and independent CFAs will be conducted on the Orientation scales measurement model. Should suitable model fit be found with each independent sample then the measurement invariance tests will commence. These results will be presented in the chapter 6.

5.6.5 Treatment of missing values

Missing values would need to be identified and handled to ensure the completeness of the data prior to conducting analyses. The missing values analysis was conducted using the SPSS Missing Values Analysis Procedure (SPSS Version 16, 2007). Results of the analysis are presented in Table 5.1.

Table 5.1.
Summary of Missing Values per Dimension: Interest Model

	Male Sample	Female Sample	Combined Sample
Basic scale	N 408	402	810
Adult Development	0.5	1.0	0.7
Advertising	1.7	2.2	2.0
Animal Care	1.2	1.2	1.0
Art/Design	2.9	2.5	2.7
Athletics	2.7	1.7	2.2
Child Development	2.9	1.7	2.3
Counseling	2.5	2.2	2.3
Culinary Arts	5.6	3.2	4.4
Farming	2.2	1.0	1.6
Fashion	1.2	1.7	1.5
Financial Services	2.7	2.0	2.3
Int'l Activities	2.2	2.2	2.2
Law/Politics	3.2	3.2	3.2

Leadership	1.7	2.0	1.9
Mathematics	1.5	1.7	1.6
Mechanical Crafts	1.0	2.0	1.5
Medical Practice	5.6	2.7	4.2
Military	1.7	1.2	1.5
Office Practices	1.5	2.5	2.0
Performing Arts	2.9	2.0	2.5
Plants	0.2	0.5	0.4
Public Speaking	0.7	1.5	1.1
Religious Activities	2.2	1.5	1.9
Risk	1.2	1.2	1.2
Sales	3.9	3.7	3.8
Science	2.9	3.0	3.0
Supervision	4.7	4.7	4.7
Woodworking	1.5	1.5	1.5
Writing	7.6	6.7	7.2
Total number of missing values	294	260	554

Note. The above values should be interpreted as percentages of missing values per scale.

The analyses reveal a missing values problem that would need to be attended to prior to further analyses. Missing values can be handled in numerous ways, these include: (1) listwise deletion, (2) pairwise deletion, (3) mean substitution, (4) group mean substitution, (5) imputation by regression, (6) structural equation modelling approach, (7) hot-deck imputation, (8) expectation maximization, (9) full information maximum likelihood and (10) multiple imputation.

5.6.5.1 Listwise deletion.

Listwise deletion simply involves the removal of each case that contains a missing value. This approach could be problematic as this could drastically reduce sample size which could reduce the power of the statistical analysis as standard errors and subsequent *t*-tests are a function of sample size (Olinsky, Chen, & Harlow, 2003). Listwise deletion may also cause further problems when nonignorable missing data is concerned. Nonignorable missing data is “where the pattern of data missingness is non-random and is not predictable from other variables in the dataset” (Olinsky et al., 2003, p. 56). The listwise technique is likely to be unbiased when the data is considered missing completely at random (MCAR). “If missing values are MCAR, cases with missing values are indistinguishable from cases with complete data" (Dunbar-Isaacson, 2006, p. 25).

5.6.5.2 Pairwise deletion.

Pairwise deletion is the systematic exclusion of a case when it is missing a value that is required for a particular part of an analysis. The pairwise approach could be employed, however a particular difficulty associated with pairwise deletion is that when a correlation matrix is constructed; it may not be positive definite – a condition required to invert the correlation matrix (Jöreskog & Sörbom, 1996; Kim & Curry, 1997, Malhotra, 1987, as cited in Olinsky et al., 2003). Hair et al. (2006) also indicate that effective sample sizes for pairwise deletion to be effective in SEM are uncertain. They also indicate that the approach is not well known in the SEM methodology.

5.6.5.3 Mean substitution.

This approach would indicate inserting the mean value for the variable under analysis based on the values of the total sample with complete values for the variable. For the unstandardized parameter estimates to be unbiased, the data would need to be approximately normally distributed and missing values assumed MCAR (Olinsky et al., 2003). However, this approach would "effectively wash out most of the structure that exists in the data" (Theron & Spangenberg, 2004, p. 23).

5.6.5.4 Group mean substitution.

In group mean substitution, the mean is calculated from an alternative demographic (race, educational level etc) group within the sample. The alternative group should display a relatively homogeneous response pattern (excluding the missing values pattern). The mean is calculated in the alternative group and then used as a value to replace the missing values in the group with the missing values (Olinsky et al., 2003). Unfortunately, due to the incomplete information regarding other demographics it would not be possible to replace missing values using this approach. Also, the limitation of mean imputation as described above may also apply.

5.6.5.5 Imputation by regression.

With this method, each variable is regressed on all of the other variables in the set via multiple regression. The unstandardized solution allows for prediction of the missing value. There is, however, some degree of error in the estimation of the imputed value as it would lie on the regression plane. As with regression, estimation of the error terms is available and can become a correcting factor for the actual imputed value (Olinsky et al., 2003). However, this process would seem to make the identification and corrections necessary for the value to be imputed tedious and difficult, considering the size of the measurement models and samples in this study.

5.6.5.6 Structural equation modelling approach.

Olinsky et al. (2003), indicate that this approach is the most elegant solution to handling missing values. This approach makes use of patterns of missing information in the data - the patterns serve as criteria in which to group cases. The initial pattern identified would be of variables that have no missing values. These cases are then treated as a group of cases. Independent groups are then compiled per variable containing those cases that have missing values on the specific variable. A final group of cases is identified that have missing values on all the variables. The researcher would use structural equation modelling to fit the measurement or structural model to all the groups simultaneously in a multi-group analysis in which all model parameters are constrained to be equal. The parameter estimates for variables for groups that have missing data on that variable are constrained to be equal to groups that contain data on the said variable. Imposing these equality constraints allow the estimation of model parameters for groups that do not have values on the variables involved. This approach is only feasible when there are few missing values in a large dataset. Moreover, it is an extremely complex technique and tedious to implement - it is not really practical in most realistic applications (Olinsky et al., 2003).

5.6.5.7 Hot-deck imputation/Imputation by matching.

Hot-deck imputation is an approach that tends to be popular in survey research (Ford, 1983; Rizvi, 1983, as cited by Olinsky et al., 2003). The method replaces a missing value with an actual value from one or more similar cases in the current dataset (Kline 2004; Olinsky et al., 2003). The technique separates complete from incomplete cases, then sorts both sets of records so that cases with similar profiles on classification (or matching) variables[29] (determined by the researcher) are grouped together. In the case of hot deck imputation the incomplete record is then randomly included among the complete records, and then replaces missing scores with those on the same variable from the nearest complete record. This process continues until the case contains no missing data (Kline, 2005). In the case of imputation by matching (Jöreskog & Sörbom, 1996a); the imputation of a missing value on variable y_a for a specific case a with no missing values on a set of p matching variables $x_1, x_2, \ldots, x_p$ involves the following procedure:

- All cases b_i; i=1, 2, …, n are identified with no missing values on either y_{bi} or on the set of matching variables for which $W = \Sigma(zb_i - za_i)^2$; i=1, 2, …, n is a minimum.
- If only n=1 case exists for which W is a minimum, then y_a is simply replaced by y_b.
- If, however W is a minimum for n>1 cases, with y values $y_1^{(m)}, y_2^{(m)}, \ldots, y_n^{(m)}$, the mean $E(y^m)=(1/n)\Sigma y_i^{(m)}$ and variance $s^2_m = (1/[n-1])\Sigma(y_1^{(m)} - E(y^m))$ of the y-values of the matching cases will be calculated.
- If $s^2_m/s^2_y < v$, where the variance ratio v was set equal to 0.50, y_a is replaced by $E(y^m)$. If the variance ratio does not pass the critical value, no imputation is done (Jöreskog & Sörbom, 1996a).

Supporters of these approaches indicate that the values preserve the distributional characteristics of the data as opposed to a mean substitution. A pitfall to the approach is that when dealing with large datasets many classification variables would need to be specified, therefore, making the decisions regarding classification cumbersome. It is also ideal if matching variables are used that are not included in the actual data analysis (Dunbar-Isaacson, 2006; Olinsky et al., 2003).

5.6.5.8 Expectation maximization.

Expectation maximization (EM) makes use of a correlation/covariance matrix that assumes the shape of a distribution (usually a *normal* distribution) for the partially missing data and bases inferences about the missing values on the likelihood under the estimated distribution (Tabachnick & Fidell, 2007). The expectation step finds

[29] Classification variables can take the form of a demographic variable i.e., gender or occupation (Kline, 2005).

the conditional expectation of the missing values given the observed data and the estimated parameters. The maximization step involves maximum likelihood estimation as if there were no missing data. These expectations are then substituted for the missing values (Little & Rubin, 1987, as cited in Olinsky et al., 2003). A disadvantage of this method is that standards errors are deemed as invalid. A further concern is that effective sample sizes are uncertain for this technique (Hair et al., 2006).

5.6.5.9 Full information maximum likelihood.

Full information maximum likelihood (FIML) is a related to EM. The FIML approach involves minimizing the determinant of the covariance matrix associated with residuals of the reduced form of the structured equations. The approach assumes that errors are normally distributed (Olinsky et al., 2003). The main advantage of FIML is that it allows the size of datasets to remain the same despite missing data. This is especially helpful as larger sample sizes are more suited to SEM analyses. However, as FIML is a direct estimating method no replacement values are chosen for missing data. This is not ideal when wanting to conduct other analyses that require the complete dataset.

5.6.5.10 Multiple imputation.

The multiple imputation approach is an extension of both EM and FIML, however the iterative estimation process is replicated between five to ten times (Olinsky et al., 2003). Each replication produces respective datasets of imputed values. Each dataset is then used to estimate the measurement/structural model. Due to variability in the datasets, it is then possible to estimate standard errors. As this approach has the benefits of EM and the added benefit of estimated standards errors, multiple imputation could be seen as the superior option when handling missing values (Olinsky et al., 2003).

Based on the discussion of the missing values approaches, it would seem that hot-deck/imputation by matching would be the most suitable approach for this study. This approach would allow for making use of the naturally occurring distributional properties of the existing data. Other SEM methods like EM and MI would have been preferable, however, these estimation methods of the missing values are based on the assumption of multivariate normality – this condition may not be found and therefore should not be assumed.

The choice of classification/matching variables is a prerequisite to using imputation by matching. Therefore, items least plagued by missing values were identified and would serve as matching variables. A selection of items with zero reported missing values are indicated by bold font in Table 5.2. Missing values were imputed using the PRELIS programme (Jöreskog & Sörbom, 1996) by using the matching variables.

It should be noted that upon completion of the imputation that all cases with missing values were successfully imputed and no cases were eliminated - thus retaining all data.

Table 5.2.
Number of Missing Values per Interest Item with Matching Variables Indicated in Bold

	ADEVI1	ADEVI2	ADEVI3	ADEVI4	ADVI1	ADVI2	ADVI3	ADVI4
Male	1	0	0	1	4	2	0	0
Female	0	3	0	1	5	2	1	0
Combined	1	3	0	2	9	4	1	0

	ADVI5	ADVI6	ADVI7	ADVI8	ANI1	ANI2	ANI3	ANI4
Male	0	1	0	0	1	0	1	1
Female	0	0	1	0	1	1	1	2
Combined	0	1	1	0	2	1	2	3

	ARTI1	ARTI2	ARTI3	ARTI4	ARTI5	ARTI6	ATHI1	ATHI2
Male	5	2	0	1	3	1	3	0
Female	3	2	0	1	3	1	2	0
Combined	8	4	0	2	6	2	5	**0**

	ATHI3	ATHI4	ATHI5	ATHI6	ATHI7	CDEVI1	CDEVI2	CDEVI3
Male	2	1	0	1	4	5	0	2
Female	0	1	1	0	3	2	0	2
Combined	2	2	1	1	7	7	**0**	4

	CDEVI4	CDEVI5	CDEVI6	CDEVI7	COUNI1	COUNI2	COUNI3	COUNI4
Male	2	3	0	0	3	2	1	0
Female	1	2	0	0	4	0	2	1
Combined	3	5	0	0	7	2	3	1

	COUNI5	COUNI6	CULI1	CULI2	CULI3	CULI4	CULI5	CULI6
Male	2	2	1	19	1	1	1	0
Female	1	1	0	11	2	0	0	0
Combined	3	3	1	30	3	1	1	0

	FARMI1	FARMI2	FARMI3	FARMI4	FARMI5	FARMI6	FASHI1	FASHI2

Male	0	2	1	0	4	2	0	2
Female	1	1	0	0	2	0	0	2
Combined	1	3	1	0	6	2	**0**	4

	FASHI3	FASHI4	FASHI5	FASHI6	FINI1	FINI2	FINI3	FINI4
Male	2	1	0	0	0	0	1	5
Female	3	1	1	0	0	0	1	3
Combined	5	2	1	0	**0**	**0**	2	8

	FINI5	FINI6	FINI7	FINI8	INTI1	INTI2	INTI3	INTI4
Male	1	1	1	2	2	3	2	1
Female	1	1	0	2	6	1	1	0
Combined	2	2	1	4	8	4	3	1

	INTI5	LAWI1	LAWI2	LAWI3	LAWI4	LAWI5	LAWI6	LAWI7
Male	1	0	1	3	3	3	2	0
Female	1	1	0	0	2	5	0	1
Combined	2	1	1	3	5	8	2	1

	LAWI8	LAWI9	LAWI10	LEADI1	LEADI2	LEADI3	LEADI4	LEADI5
Male	0	1	0	3	3	0	1	0
Female	0	3	1	1	1	0	1	2
Combined	0	4	1	4	4	0	2	2

	LEADI6	MATHI1	MATHI2	MATHI3	MATHI4	MATHI5	MATHI6	MATHI7
Male	0	0	2	1	1	0	1	1
Female	3	0	1	0	1	1	2	2
Combined	3	**0**	3	1	2	1	3	3

	MECHI1	MECHI2	MECHI3	MECHI4	MECHI5	MECHI6	MECHI7	MECHI8
Male	1	0	0	0	1	0	1	0
Female	0	1	1	4	1	0	1	0
Combined	1	1	1	4	2	0	2	0

	MECHI9	MEDI1	MEDI2	MEDI3	MEDI4	MEDI5	MEDI6	MEDI7
Male	1	2	2	3	5	3	2	2
Female	0	0	0	0	2	1	2	0
Combined	1	2	2	3	7	4	4	2

	MEDI8	MEDI9	MEDI10	MEDI11	MEDI12	MILI1	MILI2	MILI3
Male	1	1	1	0	1	0	2	1
Female	1	1	1	0	3	0	1	0
Combined	2	2	2	0	4	**0**	3	1

	MILI4	MILI5	MILI6	MILI7	OPI1	OPI2	OPI3	OPI4
Male	1	3	0	0	1	0	0	0
Female	0	4	0	0	2	1	1	2
Combined	1	7	0	0	3	1	1	2

	OPI5	OPI6	OPI7	OPI8	PERI1	PERI2	PERI3	PERI4
Male	0	3	1	1	0	0	2	1
Female	1	3	0	0	0	0	3	0
Combined	1	6	1	1	**0**	**0**	5	1

	PERI5	PERI6	PERI7	PERI8	PERI9	PERI10	PNTI1	PNTI2
Male	3	0	3	0	3	0	1	0
Female	1	0	1	1	2	0	0	0
Combined	4	0	4	1	5	0	1	0
	PNTI3	PNTI4	PNTI5	PUBI1	PUBI2	PUBI3	PUBI4	RELI1
Male	0	0	0	1	0	1	1	4
Female	0	2	0	1	2	3	0	1
Combined	0	2	0	2	2	4	1	5

	RELI2	RELI3	RELI4	RELI5	RSKI1	RSKI2	RSKI3	RSKI4
Male	4	0	1	0	4	0	0	1
Female	3	0	2	0	1	0	3	1
Combined	7	0	3	0	5	0	3	2

	SALI1	SALI2	SALI3	SALI4	SALI5	SALI6	SALI7	SALI8
Male	2	10	0	0	1	0	0	3
Female	3	5	2	2	1	2	0	0
Combined	5	15	2	2	2	2	0	3

	SCII1	SCII2	SCII3	SCII4	SCII5	SCII6	SCII7	SUPI1
Male	0	1	0	8	0	1	2	7
Female	0	1	1	6	2	1	1	4
Combined	**0**	2	1	14	2	2	3	11

	SUPI2	SUPI3	SUPI4	SUPI5	SUPI6	SUPI7	SUPI8	WOODI1
Male	2	4	1	4	1	0	0	1
Female	7	0	1	6	0	0	1	0
Combined	9	4	2	10	1	0	1	1

	WOODI2	WOODI3	WOODI4	WOODI5	WRTI1	WRTI2	WRTI3	WRTI4
Male	1	2	0	2	29	1	1	0
Female	1	2	1	2	24	2	0	0
Combined	2	4	1	4	53	3	1	0

	WRTI5	WRTI6
Male	0	0
Female	0	1
Combined	0	1

Note. Individual CISS Basic Interest scale's items have been coded into variable names. The codes allow for ease of sub-scale identification. Coding is as follows: Adult Development = ADEVI, Advertising = ADVI, Animal Care = ANI, Art/Design = ARTI, Athletics = ATHI, Child Development = CDEVI, Counseling = COUNI, Culinary Arts = CULI, Farming = FARMI, Fashion = FASHI, Financial Services = FINI, Int'l Activities = INTI, Law/Politics = LAWI, Leadership = LEADI, Mathematics = MATHI, Mechanical Crafts = MECHI, Medical Practice = MEDI, Military = MILI, Office Practices = OPI, Performing Arts = PERI, Plants = PNTI, Public Speaking = PUBI, Religious Activities = RELI, Risk = RSKI, Sales = SALI, Science = SCII, Supervision = SUPI, Woodworking = WOODI, Writing = WRTI.

5.6.6 Item analysis

The objective of item analysis is to gain a more penetrating understanding of tests or questionnaires (Murphy & Davidshofer, 2005). The procedure is essentially an analysis of correlations between each item with a total score (Kline, 1994) as well as inter-item correlations (Murphy & Davidshofer, 2005). Test publishers are likely to construct tests that would generally aim to have items that correlate on a specific scale of investigation. Items with higher correlations are assumed to be measuring the same latent variable. When developing tests/questionnaires Nunnally (1978, as cited in Kline, 1994) indicates that item analysis is to be used to make the first item selection and then the selected items are to be subjected to factor analysis.

For this study, item analysis is conducted as a valuable precursor to fitting data to the *a priori* model. The CISS was developed to measure an interest construct carrying a specific constitutive definition. In terms of this definition specific first and second-order latent interest dimensions are identified. Items have been developed to reflect the standing of test takers on specific latent interest dimensions. If these design intentions were successful it should reflect in a number of item statistics. The item analysis helps identify whether the observed variables are consistent measures of the intended latent variable. High reliability of the provided observed latent variable manifestations would give credence to the design intentions of the test developers. While Nunnally (1978) indicates that item analysis assists in making final item selection decisions. However, the intention of this study would be to retain all items but report on those that may be possible culprits that contribute to poor latent variable representation and possible poor model fit. Further to this, the analyses will also provide initial information regarding the homogeneity of each sub-scale. For these analyses, each gender sample's data are analysed separately thereby providing some initial information regarding reliability of the observed variables across gender. While this does not directly address any of the research objectives, it does provide for valuable information regarding the effectiveness of the measurement properties across gender.

The SPSS Scale Reliability Procedure (SPSS Version 16, 2007) was used to analyse the sub-scale items.

5.6.7 Dimensionality analysis

When constructing scales, the design intention is that the items selected to represent each latent variable would be in fact measuring the intended latent variable exclusively. This is termed the uni-dimensionality assumption (Hair et al., 2006). Strict uni-dimensionality will seldom, if ever, be achieved. Essentially, uni-dimensionality would be achieved if the partial inter-item correlations would become negligibly small when controlling for a single underlying factor (Hair et al., 2006). Investigating whether the number of factors required to satisfactorily explain the observed correlation matrix corresponds to the design intention underlying the scale, and investigating whether the resultant factor loadings are high is an approach to take when wanting to test this assumption. Scales that fail the uni-dimensionality assumption (i.e., more than one factor emerges naturally for a scale that was designed to measure a single latent variable) would imply that the multiple dimensions should be specified for the instrument. Again, testing this assumption does not negate the necessity of the CFA. It rather provides further insight into the internal function of the *a priori* specified factor structure of the CISS and reasons for possible poor model fit.

The dimensionality analyses are conducted by subjecting each Basic Interest scale to an unrestricted principle axis factor analysis with varimax rotation. This analysis was performed on each of the 29 Basic scales individually for both the gender samples. Principle axis factor analysis was chosen over principle components analysis. Principle components analysis does not separate error and specific variance (Kline, 1994) whereas principle axis analysis does allow for the presence of measurement error. Human behaviour without measurement error is unlikely (Steward, 2001). Varimax rotation was chosen over oblique rotation, even though oblique rotation is considered the superior method as it can provide simple structure even when underlying factors may be related to each other (Kerlinger & Lee, 2000; Steward, 2001). However, oblique rotation can be complex to interpret (Tabachnick & Fidell, 1989) and “in many cases oblique solutions are virtually identical (to orthogonal solutions) because the correlation between the factors is so small as to be negligible” (Kline, 1994, p. 68).

The possibility exists that artefact factors, which reflect differences in item difficulty value or variance, could be extracted during the above analyses when performing analyses on a matrix of product moment correlations (Hulin, Drasgow & Parsons, 1983). Descriptive statistics were consequently calculated for the items of each Basic scale. This was to determine whether the possibility of multiple factors appearing as an artefact of differential item characteristics as apposed to a true difference in factor structure that is contrary to the design intentions of the test developers.

In cases where uni-dimensionality was not met, the possibility of meaningful factor fusion was investigated. The question therefore is whether the extracted factors constitute meaningful subthemes within the original latent Basic Interest dimension. In the case of sub-scales where the uni-dimensionality assumption was challenged, irrespective of whether meaningful factor fission occurred, the ability of a single factor to account for the observed inter-item correlation matrix was also investigated. This approach was taken to investigate the magnitude of the factor loadings when a single factor (as per the *a priori* model) is forced and to examine the magnitude of the residual correlations. The magnitude of the latter could be regarded as reflecting on the credibility of the extracted single factor solution as an explanation for the observed correlation matrix.

SPSS 16 for Windows (2007) was used for the principal factor analyses as described above. The eigenvalue-greater-than-unity rule of thumb was used to determine the number of factors to extract.

The separate gender sample results are presented in chapter 6. Differences between each gender sample are also discussed. While this does not provide information regarding the configural invariance of the CISS, it does provide valuable information that could be returned to when wanting to identify reasons for poor model fit.

5.7 STRUCTURAL EQUATION MODELLING

5.7.1 Variable type

The CISS utilises a six-point Likert-type response scale. The respondent is requested to indicate their degree of preference, or level of self-perceived skill, based on item content. The data produced by this type of response scale should strictly speaking be regarded as ordinal data. Based on the results of a Monte Carlo study by Muthén and Kaplan (1985) it is, however, standard practice to specify the data obtained from Likert scales with five or more scale points as continuous data, for the purpose of CFA (Maximum Likelihood) SEM analyses. Another strategy is to convert ordered categorical data to continuous data is to use item parcels rather than item level raw data.

In the case of this study, the use of item parcelling was a practical measure to reduce the number of measurement model parameters that had to be estimated. If individual items would have been used to represent the latent Basic Interest dimensions (200 items) model parameters would have had to be estimated. This would have precluded even fitting the Basic Interest measurement model to the combined sample. However, Sass and Smith (2006, p. 568) maintain that item parcels are "nothing more than subsets of items (or observations) from a common measure". Calculating item parcels involves taking the mean (or sum) of selected subsets of items from a scale. Item parcelling reduces the number of indicators in lengthy scales (Bandalos & Finney, 2001). It has also been found that the use of parcelling could significantly improve model fit in some circumstances (Bandalos, 2002; Bandalos & Finney, 2001). In some circumstances, it may help assure that multivariate normality is obtained when handling data using maximum likelihoods estimation methods (Sass & Smith, 2006).

Disadvantages of using item parcelling do exist. For example: Meade and Lautenschleager (2004) reported that the MI test of equality of factor loadings (i.e., metric invariance) tend to be more precise when using item level data. In a further study, Meade and Kroustalis (2006) found that the use of items versus item parcels is preferred when conducting tests of MI. From their simulation studies it was found that even though fit could be poor (when using item data), lack of invariance may be masked by using item parcels. Sass and Smith (2006) also indicate that there seems to be a lack of evidence of one single suitable approach to constructing parcels. Kim and Hagtvet (2003) indicate that using parcels may increase the likelihood of misrepresenting the latent construct.

However, due to the complexity and size of the CISS Basic scales measurement model (29 primary interest latent variables and 200 items), it was decided that item parcelling would be a suitable strategy to employ in this study. This decision was primarily due to the problem that the number of parameters to be estimated would have exceeded the number of observations in the combined sample if a model of this complexity would have been fitted at an item level of analysis.

However, Bandalos and Finney (2001) do highly recommend that uni-dimensionality is met prior to the construction of parcels. The objective of dimensionality analyses is test whether the uni-dimensionality assumption is met for each factor. Inadequate factor loadings would suggest that items should be removed and factor fission would suggest that a split should be proposed in the sub-scale factor composition. If these actions are taken then a revision of the measurement models would take place. It is, however, important to note that the researcher does not have intellectual property rights on the instrument and neither does he have any mandate from the test developer to modify the instrument and its design intention. For this study it is, therefore, not authorized to re-design the measurement model in any way. Consequently, the dimensionality investigation is, in this case, not a step in ensuring that the calculated item parcels are internally consistent observational reflections of the latent Basic Interest variables as proposed by the test authors. Rather the dimensionality analysis provides further insight into the internal function of the *a priori* specified factor structure of the CISS and reasons for possible poor model fit.

5.7.2 Item parcelling approach

As was indicated in the foregoing discussion, item parcels were constructed for the analyses in this study. However, a discussion of the different approaches follows.

A number of different approaches can be taken when generating item parcels. These approaches could include: (i) a qualitative investigation into the content of items and allocating parcels accordingly (Nasser, Takahashi & Benson, 1997), (ii) investigating the internal consistency of the scale and allocating item accordingly (Nasser et al., 1997), (iii) using factor loading information resulting from an exploratory factor analysis (Nasser et al., 1997), as well as (iv) the use of descriptive statistic information. These approaches could be considered logical/quantitative approaches to specifying item parcels (Hall, Snell & Foust, 1999). A further approach that could be considered is a random combination of items as per sub-scale (Hall et al., 1999; Kim & Hagtvet, 2003). When using a random approach, split-halves or odd-even combinations could be used (Prats, 1990; as cited by Kim & Hagtvet, 2003).

Some researchers recommend making use of a logical method as opposed to a random item selection (e.g., Bandalos, 2002; Hall et al., 1999; Sass & Smith, 2006). The construction of item parcels based on factor loadings would, in the case of this study, make sense if the uni-dimensionality assumption would be supported. This procedure would then result in item parcels that measure the single underlying latent variable approximately equally well. The construction of item parcels based on factor loadings would also make sense in the case of this study if meaningful factor fusion would occur. The construction of item parcels based on factor loadings would, however, become somewhat problematic if the un-dimensionality assumption would not be met but the identity of the factors would not be discernable or the factors would not be meaningful subthemes within the original interest factor. Such parcels do not reflect the design intentions of the test authors and the use of such parcels would therefore result in a questionable test of the extent to which the original design intentions succeeded.

The creation of item parcels could be seen as somewhat contentious if the results obtained on the item and dimensionality analyses would suggest that the design intentions of the test developers have failed. The item parcels serve as indicator variables of the latent variables. If the objective of the analysis would have been to evaluate the structural relations that exist between the latent interest dimensions, then it becomes critical to ensure that each item parcel provides a valid measure of the latent variable it was assigned to represent. Failure to do so would prevent a valid and credible test of the hypothesized structural model. Under these conditions it would be imperative that the results of the item and dimensionality analyses should be used to identify and remove inappropriate items so as to ensure that only items that validly reflect the latent variable of interest are combined in a parcel.

In the current research, however, the objective is not to test specific structural relations hypothesized to exist between specific latent variables. The objective is rather to evaluate the relationships that exist between latent variables and indicators that were designed to reflect the latent variables. Therefore, the objective is to evaluate the success with which items represent the latent (primary or secondary) interest dimension they were tasked to reflect. Again, the ideal would have been to evaluate the success with which items represent the latent interest dimension they were tasked to reflect by fitting the measurement model with the individual items as indicator variables. Since this is not feasible in this instance, all items are combined into parcels and the success with which these sets of items represent the latent interest dimension they were tasked to reflect is then evaluated. The creation of item parcels should therefore not be viewed as inappropriate if the results obtained on the item and dimensionality analyses would indicate that the items comprising the various sub-scales do not in an internally consistent manner reflect a single underlying latent variable. The measurement model will, however, invariably have to reflect this failure either in a lack of model fit or in low completely standardized factor loadings and high completely standardized measurement error variances.

The random combination of items into two parcels using split-halves or odd-even combinations should result in two approximately equally valid (or less valid) indicators of the latent variable, irrespective of the outcome of the item- and dimensionality analyses. If the latter would indicate that the items comprising a sub-scale show high internal consistency and strongly reflect a single underlying latent variable, then the randomly created item parcels should constitute valid composite indicators of each interest latent variable. If the item- and dimensionality analyses would indicate that the items comprising a sub-scale show low internal consistency and that they reflect more than one underlying latent variable, then the randomly created item parcels should constitute less valid composite indicators of each interest latent variable. Again it should be remembered that the objective with the formation of item parcels in this study is not to ensure reliable, valid composite indicators of the latent variables concerned so as to ensure a credible test of the structural relations existing between latent variables. The objective with the formation of item parcels in this study is rather to ensure composite indicators that are accurate summaries of the reliability and validity with which the individual items represent the latent variables they were earmarked to reflect so as to ensure a valid test of the measurement relations implied by the design intention.

Based on the above, it would then seem more appropriate to use a random selection of items for two item parcels per Basic Interest scale. Composite item parcels were created by sorting items in an odd-even fashion. The odd-even assignment was based on the coded item variable names. Each sub-scale was then sorted into parcels as per odd or even number and the arithmetic mean was calculated for the even numbered and for the odd numbered items. This resulted in 58 (29 sub-scales by 2 parcels) item parcels being created to represent the observed variables per latent interest variable.

5.7.3 Evaluation of multivariate normality

As item parcels will be used as indicator variables for this study, the property of the variable now becomes continuous - this is due to the composite nature of the parcel. When using continuous data in SEM, maximum likelihood estimation is preferred. Other estimation methods include generalised least squares (GLS) and full information maximum likelihood (FIML). FIML is useful when dealing with missing values. However, with all these estimation methods multivariate normality is assumed for the data (Mels, 2003).

In the event of working with non-normal data, Mels (2003) indicates that additional estimation methods could be utilized, for example: robust maximum likelihood (RML), weighted least squares (WLS) or diagonally weighted least squares (DWLS). These methods are advantageous as interpretation of the solution is not based on transformed values (Du Toit & Du Toit, 2001). Mels (2003) does make a further recommendation that RML would be the preferred approach when dealing with multivariate non-normal data.

The normality of the composite indicators (i.e., item parcels) were evaluated using PRELIS (Jöreskog & Sörbom, 1996b). The null hypothesis of univariate normality was rejected ($p<0.05$) for most indicator variables, for both samples. However, the null hypothesis of univariate normality was not rejected ($p>0.05$) for two parcels in the Interest model, in both the male and female samples (i.e., male sample: SALIP1 and SALIP2; female sample: SUPIP2 and LEADIP2). The multivariate normality results are summarised in Table 5.3. In both samples the null hypothesis of multivariate normality was rejected ($p<0.05$).

Table 5.3.
Tests Of Multivariate Normality For Continuous Variables: Interest Parcels

		Skewness			Kurtosis			Skewness and Kurtosis	
Sample	N	Value	Z-Score	P-Value	Value	Z-Score	P-Value	Chi-Sqaure	P-Value
Males	408	680.024	41.428	0.000	3815.316	21.451	0.000	2176.425	0.000
Females	402	794.840	62.336	0.000	3904.227	23.919	0.000	4457.855	0.000

Due to these findings the RML method of estimation was selected as the preferred estimation method for this research. The item parcel data was not normalized.

5.7.4 Measurement model fit

5.7.4.1 Estimation method.

In order to meet the measurement invariance research objectives of this study, LISREL 8.88 (Du Toit & Du Toit, 2001, Jöreskog & Sörbom, 1996b) was used to determine the fit of: (i) the Basic Interest models for the combined sample, (ii) the Orientation Interest model for the combined sample, (iii) the Orientation Interest model on the two gender samples and (iv) the Orientation Interest model when fitted in a multi-group analysis.

Due to the lack of normality (refer to the previous sub-section) the data was read into PRELIS (Jöreskog & Sörbom, 1996a) to compute the asymptotic covariance matrix which would serve as input for further LISREL analyses. Consequently model parameter estimation was determined by the Robust Maximum Likelihood (RML) estimation method in LISREL (Jöreskog & Sörbom, 1996b).

5.7.5 Evaluation of fit in the single group analyses

The fit of the measurement model in the single group analyses was evaluated by testing H_{01}, H_{060}, H_{0119M} and H_{0119F}. The full range of fit indices (both comparative and absolute) reported in LISREL was used to determine the adequateness of fit of the data for each sample and model. Diamantopoulos and Siguaw (2000) indicate that the use of fit indices should be interpreted holistically and integrated to form a considered decision with regards to model-data fit. The indicators of fit are described in detail in chapter 6 when evaluating fit of the respective models. Examination of the modification and resultant change indexes presented by LISREL will also be reviewed. Standardized residuals are also investigated and presented. Finally, squared multiple correlations and the completely standardized factor loadings are reviewed for each CFA.

5.7.6 Measurement invariance hypothesis tests

Only if the single group analyses would result in a rejection of H_{01} to H_{0177F} or at least indicate reasonable model fit of the models to the data, multi-group analyses will be conducted. This stage marks the beginning of the imposing measurement invariance constraints on a nested model. In the multi-group analysis, the model is simultaneously fitted to the data of both the male and female samples, initially with no model parameters constrained to be equal across groups. This is the configural invariance test. This would essentially test the hypothesis: $\xi^{g} = \xi^{g'}$.reflected in H_{0178}. The fit statistics of this unconstrained model would serve as baseline information for comparisons with increasingly constrained models. If H_{0178} is not rejected (or reasonable model fit would be obtained), full measurement invariance will be tested by testing H_{0179}. H_{0179} will be tested by fitting the Orientation Interest scale measurement model simultaneously to both gender samples in a multi-group analysis with all model parameters constrained to be equal and then calculating the difference[30] in the Satorra-Bentler χ^2 test statistic obtained under the fully constrained and fully unconstrained conditions (Meade & Lautenschlager, 2004; Vandenberg & Lance, 2000).

A significant (p<0.05) chi-square difference statistic, evaluated at the difference in the degrees of freedom of the respective models, would result in a rejection of H_{0179} and would indicate lack of full measurement invariance (Mels, 2003). If H_{0179} would be rejected, one or more model parameters would differ across the gender groups. If H_{0179} would be rejected it would make sense to search for the source of the measurement invariance. If H_{0179} would be rejected the metric invariance null hypothesis (H_{0180}) would be tested by fitting the Orientation Interest scale measurement model simultaneously to both gender samples in a multi-group analysis with all model parameters constrained to be equal but for the elements of Λ_x and then calculating the difference in the Satorra-Bentler χ^2 test statistic obtained under the partially constrained conditions and under the fully unconstrained conditions (Meade & Lautenschlager, 2004; Vandenberg & Lance, 2000). A significant (p<0.05) chi-square difference statistic, evaluated at the difference in the degrees of freedom of the respective models would result in a rejection of H_{0180} and would

[30] When the difference in statistical fit between models is calculated, based on the differences in the Sarorra-Bentler χ^2 statistic, an adjustment formula proposed by Satorra and Bentler (1999) should be employed. This should be done to reflect the fact that the difference between the two Sarorra-Bentler χ^2 values is not distributed as a chi-square distribution. The adjustment formula is provided in Tabachnick and Fidell (2001).

indicate lack of metric invariance (Mels, 2003). If H_{0179} would be rejected and irrespective of the decision on H_{0180}, a further search for differences on other model parameters would be warranted. This study will, however, not go beyond the test of metric invariance.

5.8 STATISTICAL POWER

Statistical power[31] is important when making crucial decisions regarding the rejection or not of statistical hypotheses about model fit. In the case of this study, statistical hypotheses of close fit were formulated for the various single group measurement models in terms of the parameter values attained for the Root Mean Squared Error of Approximation (RMSEA) fit index. In the context of this study statistical power refers to the conditional probability of rejecting the null hypothesis given that it is false (P[reject H_0: RMSEA $\leq$ 0.05)|H_0 false]). In the context of SEM statistical power therefore refers to the probability of rejecting an incorrect model. If the null hypothesis of close fit (H_0: RMSEA $\leq$ 0.05) would not be rejected, the question 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. If the decision not to reject the null hypothesis of close fit results under conditions of low power, it causes ambiguity because it is not clear whether the decision was due to the accuracy of the model or to the insensitivity of the test to detect specification errors in the model. The decision not to reject the null hypothesis of close fit would constitute convincing evidence on the merit of the model to the extent that it would be found that the statistical power of the evaluation of close fit had reasonably high power. Conversely, however, if the null hypothesis of close fit would be rejected under conditions of extremely high power it would create the fear that a reasonably accurate model had been rejected because of the extreme sensitivity of the test for minor specification errors in the model.

To calculate statistical power the methodology of MacCallum, Browne and Sugwara (1996) was implemented, using the model specification as indicated previously. Tabachnick and Fidell (2007) indicate that it often desired that power levels of at least 0.80[32] are found prior to continuing with analyses. In the case of evaluating the fit of a measurement model this would mean it would be desirable that the conditional probability of rejecting the close

[31] Power represents the probability that effects that actually exist have a chance of producing statistical significance on the data analysis (Tabachnick & Fidell, 2007).
[32] 0.80 would signal an 80 percent probability of achieving a significant result if an effect exists.

fit null hypothesis given that the fit of the model is actually mediocre, should be at least 0.80.

The MacCallum et al. (1996) SAS syntax was translated into SPSS for the power calculation (Spangenberg & Theron, 2005). To derive power estimates for the test of close fit, the effect sizes of 0.05 and 0.08 were captured into the syntax (as the values of RMSEA under H_0 and H_a respectively), along with a statistical significance level of 0.05. In addition, the sample size (408; 402) and degrees of freedom, based on the each respective model, was entered into the syntax. The power calculation specifications and results for each model and sample are reported below.

5.8.1 Power calculations: combined sample, Basic scales interest model

Degrees of freedom for the interest model are 1189. After running the SPSS syntax on a sample of 810, a power value of 1.00 was obtained for the test of close fit.

5.8.2 Power calculations: male sample, Orientation scales interest model

Degrees of freedom for the interest model are 1574. After running the SPSS syntax on a sample of 408, a power value of 1.00 was obtained for the test of close fit.

5.8.3 Power calculations: female sample, Orientation scales interest model

Degrees of freedom for the interest model are 1574. After running the SPSS syntax on a sample of 402, a power value of 1.00 was obtained for the test of close fit.

These power values should be taken into consideration when interpreting the results of the tests of close fit reported in chapter 6.

CHAPTER 6
RESULTS

6.1 INTRODUCTION

As described in the chapters 4 and 5, this research aims to determine whether the CISS is able to measure the latent interest variables (primary and global levels) given its constitutive definition of the interest construct as it intends to and, if so, whether the latent interest variable is measured equivalently across gender. The final operational hypotheses were described in chapter 5. This chapter aims to provide evidence that is used to decide on the validity of the operational hypotheses presented at the beginning of the previous chapter. However, prior to conducting the CFAs necessary to evaluate the measurement invariance of the given instrument, some analyses are provided that provide insight into the instrument's functioning. These analyses (i.e., item and dimensionality analyses across the two gender samples) assist in gaining understanding into the psychometric integrity of the indicator variables that represent the various latent variables (prior to the construction of item parcels). Where possible, throughout the chapter, notable gender differences are discussed.

The results are presented in the following order: (i) item analyses, (ii) dimensionality analyses (iii) CFA of the Basic Interest measurement model for the combined sample, (iv) CFA of the Orientation Interest measurement model for the combined sample, (v) CFA of the Orientation Interest measurement model for the male sample, and (vi) CFA of the Orientation Interest measurement model for the female sample. The chapter also describes the implications of conducting the measurement invariance hypothesis tests based on the CFA results.

6.2 ITEM ANALYSES

As described in the previous chapter, item analysis was conducted on each of the Basic Interest sub-scales. Item analyses were conducted to investigate: (i) the reliability of indicators of each latent variable, (ii) homogeneity of each sub-scale and (iii) screen items prior to their inclusion in composite item parcels representing the latent variables. The item analyses were conducted on each gender sample separately. The SPSS Scale Reliability Procedure (SPSS Version 16, 2007) was used to analyse the sub-scale items. A summary of the item statistics for the respective gender samples is available in Appendix 1. The detailed output of the item analyses is electronically available (on the included CD, folder: ITEM ANALYSES) in Appendix A.

Initially, problematic items identified through item statistics are flagged and discussed. Thereafter the homogeneity for each Basic Interest scale is evaluated.

6.2.1 Item analyses: interest model statistics

6.2.1.1 Sub-scale reliabilities.

Sub-scale reliabilities for each sample are reported in Tables 6.1 and 6.2. For the male sample, 5 of the 29 sub-scales obtained Cronbach Alpha values lower than a seemingly stringent cut-off value of 0.80[33]. These sub-scales were the Adult Development (0.75), Art/Design (0.78), Counseling (0.79) International Activities (0.75) and Risk (0.78) Basic scales. However, all the values were greater than the often quoted benchmark value of $\alpha > 0.70$ suggested by Nunnally (1978). Although this is frequently ignored, the conditional nature of this recommendation should be remembered when evaluating this finding.

For the female sample, 5 of 29 Basic Interest scales did not meet the 0.80 cut-off. The six sub-scales were Adult Development (0.71), Fashion (0.72), Leadership (0.78), Public Speaking (0.78) and Risk (0.68). The most concerning of these, however, is the Risk scales which even is below the rather lenient Nunnally (1978) 0.70 cut-off point. However, this may have been expected due to the shortness of the scale (only four items).

Reliability coefficients were calculated for all the Basic Interest scales before and after the imputation. The values were compared to determine if the imputation had

[33] Even though Nunnally (1978, p. 245) indicates that ”in the early stages of research on predictor tests or hypothesized measures of a construct, one saves time and energy by working with instruments that have only modest reliability, for which purpose reliabilities of 0.70 or higher will suffice” he nonetheless then continues and argues that “in many applied settings a reliability of 0.80 is not nearly high enough.. Although not frequently quoted Nunnally (1978, p. 246) continues and claims that “in those applied settings where important decisions are made with respect to specific test scores, a reliability of 0.90 is the minimum that should be tolerated, and a reliability of 0.95 should be considered the desirable standard.”

an effect on the alpha values reported. Upon review of Table 6.1 (male sample) small alpha value changes were observed for Adult Development (decrease of 0.001), Art/Design (decrease of 0.004), Counseling (decrease of 0.004) and International Activities (decrease of 0.004). No changes were observed for the Risk/Adventure sub-scale. This would suggest that imputation affected the scale reliabilities in a negligible way. For the female sample (Table 6.2), a similar trend was observed. Trivial changes in alpha values for the Athletics (increase of 0.003), Public Speaking (decrease of 0.001) and Risk/Adventure (decrease of 0.003) sub-scales emerged. No alpha changes were observed for the Adult Development, Fashion and Leadership sub-scales.

Overall, the results of the reliability analyses would suggest reasonably satisfactory levels of internal consistency with only a few sub-scales raising some concern. In addition, imputation of missing values has not affected the reliability results.

6.2.1.2 Item statistics.

Visual inspection of the means and standard deviations (i.e., for extreme means or small standard deviations), revealed no items that had to be flagged as problematic. However, inspection of the item-total statistics, particularly the corrected item total correlations, squared multiple correlations and predicted increases in alpha (when a particular item is deleted), a number of items were flagged as potentially problematic. Screening was based on cut-off values as follows: (i) corrected item-total correlations < 0.30, (ii) squared multiple correlations < 0.30 and (iii) a noticeable increase in alpha when compared to the scale's Cronbach's Alpha. Item statistic information is available in Appendix 1.

Table 6.1.
Reliability of CISS Basic Interest Scales for the Male Sample

		Pre-imputation				Post-imputation			
Sub Scale	Number of items	Valid cases	Alpha	Mean	Variance	Valid cases	Alpha	Mean	Variance
Adult Development	4	406	0.751	15.03	23.276	408	0.750	15.0147	23.223
Advertising	8	401	0.923	27.81	100.877	408	0.923	27.7426	100.098
Animal Care	4	405	0.860	13.44	28.000	408	0.859	13.4387	27.810
Art/Design	6	397	0.785	21.28	41.810	408	0.781	21.3162	41.435
Athletics	7	397	0.840	24.48	68.755	408	0.835	24.4975	67.887
Child Development	7	397	0.900	25.16	82.578	408	0.899	25.1225	81.690
Counseling	6	398	0.798	21.89	43.077	408	0.793	21.8922	42.848
Culinary Arts	6	386	0.838	19.73	47.393	408	0.835	19.7721	46.574
Farming	6	399	0.895	21.26	60.583	408	0.893	21.2770	59.749
Fashion	6	403	0.874	20.71	64.250	408	0.875	20.6863	64.324
Financial Services	8	398	0.898	28.11	90.248	408	0.896	28.1078	89.541
Int’l Activities	5	399	0.754	17.74	35.579	408	0.750	17.7402	35.249
Law/Politics	10	398	0.902	35.91	139.848	408	0.900	35.8039	137.721
Leadership	7	404	0.853	25.76	67.865	408	0.851	25.7353	67.640
Mathematics	7	402	0.891	28.91	79.856	408	0.891	28.9730	79.707
Mechanical Crafts	9	404	0.860	34.19	95.978	408	0.859	34.2255	95.630
Medical Practice	12	392	0.926	44.86	195.957	408	0.925	45.0270	195.525
Military	8	402	0.869	27.46	87.241	408	0.868	27.4387	86.763
Office Practices	9	403	0.872	31.83	97.217	408	0.871	31.9510	97.614
Performing Arts	10	397	0.890	34.68	138.410	408	0.887	34.7843	136.789
Plants	5	407	0.889	17.90	39.832	408	0.889	17.9044	39.757
Public Speaking	4	406	0.807	14.49	25.204	408	0.806	14.4755	25.189
Religious Activities	5	399	0.926	17.35	52.358	408	0.923	17.3162	51.824
Risk	4	403	0.788	13.13	28.370	408	0.788	13.1520	28.272
Sales	8	392	0.853	27.32	66.774	408	0.852	27.2353	66.922
Science	7	397	0.874	28.69	69.448	408	0.875	28.6667	69.884
Supervision	8	390	0.823	28.44	63.785	408	0.826	28.5098	64.477
Woodworking	5	402	0.856	17.53	37.626	408	0.855	17.4975	37.228
Writing	6	378	0.837	21.28	47.365	408	0.833	21.4314	46.659

Table 6.2.
Reliability of CISS Basic Interest Scales for the Female Sample

Scale	Number of items	Pre-imputation Valid cases	Alpha	Mean	Variance	Post-imputation Valid cases	Alpha	Mean	Variance
Adult Development	4	398	0.710	14.55	21.014	402	0.710	14.5821	20.982
Advertising	8	393	0.897	26.48	84.582	402	0.893	26.4229	83.247
Animal Care	4	397	0.923	14.79	39.366	402	0.920	14.8010	39.312
Art/Design	6	392	0.805	19.81	43.712	402	0.805	19.8483	43.740
Athletics	7	395	0.795	25.28	62.005	402	0.798	25.2562	62.535
Child Development	7	395	0.874	24.40	73.043	402	0.874	24.4478	72.527
Counseling	6	393	0.806	20.56	49.129	402	0.804	20.6194	48.775
Culinary Arts	6	389	0.874	19.43	55.598	402	0.873	19.4851	55.492
Farming	6	398	0.929	23.52	79.479	402	0.929	23.4975	79.488
Fashion	6	395	0.715	19.67	44.368	402	0.715	19.6343	44.218
Financial Services	8	394	0.929	29.41	125.672	402	0.928	29.3756	124.674
Int’l Activities	5	393	0.821	16.79	42.647	402	0.819	16.8557	42.538
Law/Politics	10	391	0.904	37.08	157.102	402	0.901	37.0920	153.869
Leadership	7	395	0.783	24.26	54.786	402	0.783	24.2413	54.752
Mathematics	7	395	0.895	27.84	93.001	402	0.895	27.7985	92.889
Mechanical Crafts	9	394	0.940	36.17	174.976	402	0.940	36.1020	174.511
Medical Practice	12	391	0.909	43.85	188.002	402	0.910	43.8582	187.997
Military	8	397	0.871	29.99	111.977	402	0.870	30.0224	111.518
Office Practices	9	392	0.872	31.92	102.229	402	0.870	31.8607	100.419
Performing Arts	10	394	0.878	34.98	137.295	402	0.877	34.9677	136.125
Plants	5	400	0.897	18.58	46.484	402	0.897	18.5796	46.274
Public Speaking	4	396	0.781	14.19	25.433	402	0.780	14.2164	25.287
Religious Activities	5	396	0.895	18.47	51.065	402	0.893	18.5050	50.565
Risk	4	397	0.687	13.76	23.643	402	0.684	13.7463	23.546
Sales	8	389	0.870	29.03	85.229	402	0.869	28.9801	85.371
Science	7	390	0.891	27.18	85.957	402	0.890	27.1070	85.058
Supervision	8	384	0.808	28.92	70.743	402	0.809	28.7761	70.798
Woodworking	5	396	0.840	19.23	56.507	402	0.841	19.2612	56.283
Writing	6	376	0.843	20.59	56.696	402	0.838	20.6741	55.327

For the male sample, one item (WRTI6) was flagged due to not meeting the above mentioned cut-off values. This item, from the Writing Basic scale, refers to *an interest in writing reports of a technical nature.* It is possible that it may not be viewed in the same light as the rest of the scale items, which are distinguished from item WRTI6 by following a *classical literature theme* (e.g., writing novels or a newspaper article).

In the female sample four items raised concern. These items are: ATHI6 (Athletics/Physical Fitness Basic scale), FASHI4 (Fashion Basic scale), MEDI11 (Medical item cross-loaded on the Military/Law Enforcement Basic scale) as well as WRTI6. Content analyses of the items revealed possible explanations for these findings. For example, ATHI6 refers to an interest in *activities for body improvement* versus the remaining items carrying a *competitive athletics* theme. FASHI4 carries a *personal beauty care* theme versus the remaining fashion items relating to *activities regarding clothing/hair design in fashion*. MEDI11 contains content referring to *helping in an emergency situation* which may not been seen in the same light as *traditional military activities* (e.g., an interest in commanding a military unit or enjoying a military drill).

Due to the confirmatory nature of this study, the above mentioned items were retained for the subsequent CFAs.

6.3 DIMENSIONALITY ANALYSES

As discussed in chapter 5, the uni-dimensionality assumption has been tested for each of the Basic Interest scales. The analyses assist in gaining an understanding of the item functioning per scale in the questionnaire. For these analyses both the number of factors extracted and associated factor loadings are used to determine uni-dimensionality. Should scales fail the uni-dimensionality assumption the possibility of meaningful factor fission was investigated. Therefore, the question is whether the extracted factors constitute meaningful subthemes within the original latent Basic Interest dimension. In the case of sub-scales where the uni-dimensionality assumption was challenged, irrespective of whether meaningful factor fission occurred, the ability of a single factor to account for the observed inter-item correlation matrix was also investigated. This approach was taken to investigate the magnitude of the factor loadings when a single factor (as per the *a priori* model) is forced and to examine the magnitude of the residual correlations. In addition, descriptive statistics[34] were calculated for the items of each Basic scale to determine the possibility of multiple factors appearing as an artefact of differential item characteristics.

The dimensionality analyses were conducted by subjecting each Basic Interest scale to an unrestricted principle axis factor analysis with varimax rotation. This analysis was performed on each of the 29 Basic Interest scales. The factor analysis was conducted separately for each gender sample.

SPSS 16 for Windows (2007) was used for the abovementioned analyses. The eigenvalue-greater-than-unity rule of thumb was used to determine the number of factors to extract. The detailed output of the exploratory factor analyses is electronically available (on the included CD, folder: DIMENSIONALITY) in Appendix A.

This following sub-section begins with on overview of Basic scales that failed the uni-dimensionality assumption. An executive summary is also initially provided that described items that returned low factor loadings. Thereafter, further detailed information (through the dimensionality analyses) per sub-scale is provided.

6.3.1 Dimensionality analysis results: interest model for the male sample

The results of the principle axis factor analyses for the male sample are summarised in Table 6.3. Ten of the 29 sub-scales failed the uni-dimensionality test. The affected scales are: (i) Art/Design, (ii) Counseling, (iii) International Activities, (iv) Law/Politics, (v) Leadership, (vi) Mechanical Crafts, (vii) Medical Practice, (viii) Military, (ix) Office Practices and (x) Performing Arts.

6.3.1.1 Item factor loadings: interest model for the male sample.

This sub-section aims to provide an executive summary of the item factor loadings found for the respective model and sample.

Factor loadings can be interpreted as follows: (i) 0.30 to 0.40 are considered to meet the minimal level for interpretation of structure, (ii) 0.50 or greater are considered practically significant, and (iii) loadings exceeding 0.70 are considered

[34] The full output is available in Appendices 2 & 3 for the respective sub-samples. This is also available in electronic form (on CD, folder: DESCRIPTIVES STATS) in Appendix A.

indicative of well-defined structure (Hair et al., 2006). The practical 0.50 or greater was used as a benchmark for these analyses. An item indicating a loading of 0.50 would denote that 25 percent of the variance is accounted for by the item for the factor.

Item factor loadings varied greatly for the male sample. The loadings ranged between 0.27 and 0.96. However, for the majority of item loadings the cut-off point of 0.50 was met. Results from the series of 29 PCAs that were conducted revealed that 10 of the 200 items obtained factor loadings lower than 0.50. These items were: ARTI3 (0.41), COUNI (0.44), MEDI12 (0.46), OPI1 (0.48), ATHI4 (0.36), CULI3 (0.45), PUBI (0.49), SALI1 (0.44), SCII2 (0.45) and WRTI6 (0.27). Of concern are items ATHI4 and WRTI6 as factors loading were well below the cut-off point. ATHI4 may not be able to contribute effectively to the internally consistent description of the Athletics scale. This item carries a *self-defence* theme versus the others which seem to have a *sport-orientated* theme. In the item analysis (sub-section 6.2.1.2) discussion WRTI6 had raised a flag due to problematic item statistics. Both of the affected sub-scales did, however, pass the uni-dimensionality test.

A discussion of the results of the dimensionality analyses are provided next.

Table 6.3.
Principle Factor Analyses of CISS Basic Interest Scales for the Male Sample

Basic Scale	KMO	% Variance explained	Min factor loading	Max factor loading
Adult Development	0.731	57.230	0.584	0.747
Advertising	0.914	65.437	0.587	0.893
Animal Care	0.824	70.361	0.718	0.842
Art/Design	0.700	Factor 1: 47.908	0.661	0.787
		Factor 2: 22.644	0.727	0.964
		Single forced factor: 40.430	0.509	0.712
Athletics	0.819	51.854	0.362	0.854
Child Development	0.902	62.682	0.586	0.918
Counseling	0.767	Factor 1: 49.525	0.788	0.897
		Factor 2: 16.669	0.537	0.687
		Single forced factor: 40.195	0.475	0.774

Culinary Arts	0.837	55.604	0.451	0.809
Farming	0.878	65.245	0.704	0.823
Fashion	0.816	61.696	0.641	0.844
Financial Services	0.898	58.424	0.519	0.807
Int'l Activities	0.668	Factor 1: 50.113	0.544	0.837
		Factor 2: 21.538	0.593	0.927
		Single forced factor: 37.875	0.541	0.684
Law/Politics	0.907	Factor 1: 52.882	0.582	0.747
		Factor 2: 10.443	0.686	0.795
		Single forced factor: 47.769	0.621	0.803
Leadership	0.824	Factor 1: 53.216	0.557	0.887
		Factor 2: 15.225	0.690	0.805
		Single forced factor: 45.830	0.512	0.783
Mathematics	0.862	60.847	0.616	0.895
Mechanical Crafts	0.860	Factor 1: 50.745	0.583	0.918
		Factor 2: 18.341	0.557	0.699
		Single forced factor: 45.826	0.095	0.802
Medical Practice	0.919	Factor 1: 55.144	0.549	0.746
		Factor 2: 11.169	0.546	0.887
		Single forced factor: 51.244	0.623	0.851
Military	0.865	Factor 1: 52.784	0.587	0.846
		Factor 2: 14.216	0.544	0.740
		Single forced factor: 46.657	0.452	0.805
Office Practices	0.870	Factor 1: 50.098	0.549	0.878
		Factor 2: 13.564	0.503	0.657
		Single forced factor: 45.666	0.335	0.842
Performing Arts	0.865	Factor 1: 50.108	0.503	0.718
		Factor 2: 11.407	0.521	0.778
		Single forced factor: 44.807	0.488	0.785
Plants	0.864	69.425	0.654	0.845
Public Speaking	0.765	63.776	0.499	0.857
Religious Activities	0.874	76.747	0.671	0.928
Risk	0.757	61.532	0.513	0.867
Sales	0.845	49.674	0.448	0.773
Science	0.879	58.672	0.453	0.879
Supervision	0.839	45.291	0.498	0.715
Woodworking	0.834	63.542	0.590	0.845
Writing	0.829	56.644	0.273	0.803

6.3.1.2 Art/Design scale uni-dimensionality results.

For the Art/Design scale two clear factors emerged. Factor 1 embodies interests related to *artistic activities (art, sketching)*. Factor 2 describes interests in *architectural/interior design.* The rotated factor matrix (Table 6.4) presents the items that load on the respective factors. Descriptive statistics were reviewed to determine if the two factor structure may be an artefact of the differential skewness of the items. The review found no differential skewness. Therefore, the two factors may be justified. However, due to the confirmatory nature of this study a single factor was forced on the scale as per the *a priori* model. This analysis resulted in a factor where all items loaded satisfactorily (0.50 < λ < 0.71).

Table 6.4.
Rotated Factor Matrix: Art/Design Scale

Item	Factor 1	Factor 2
ARTI1	0.089	**0.964**
ARTI2	**0.726**	0.294
ARTI4	0.168	**0.727**
ARTI5	**0.787**	0.113
ARTI3	0.410	**0.414**
ARTI6	**0.661**	0.058

6.3.1.3 Counseling scale uni-dimensionality results.

Similarly, the Counseling scale returned a two factor structure indicating a difference between *interests in psychology* (Factor 1) and *social/group development work of a non-psychological manner* (Factor 2). The results of the rotated factor matrix (Table 6.5) show the items that load on the respective factors. The descriptive statistics where reviewed and indicated no differential skewness. Upon closer inspection of the rotated factor matrix it is evident that only two items load onto the *interest in psychology* factor - thereby not creating a meaningful independent factor. The two factor solution was subjected to factor fusion and all but one item (COUNSI5, λ = 0.48) obtained loadings exceeding 0.50 (0.48 < λ < 0.77).

Table 6.5.
Rotated Factor Matrix: Counseling Scale

Item	Factor 1	Factor 2
COUNI1	0.406	**0.441**
COUNI2	**0.897**	0.245
COUNI3	0.175	**0.687**
COUNI4	**0.788**	0.280
COUNI5	0.167	**0.537**
COUNI6	0.249	**0.574**

6.3.1.4 International Activities scale uni-dimensionality results.

Dimensionality analysis for the International Activities scale revealed a two factor structure differentiating *interest in travel* (Factor 1) versus *foreign language interests* (Factor 2). The rotated factor matrix (Table 6.6) shows the items that load on the respective factors. The descriptive statistics were reviewed and indicated no differential skewness. Upon forcing a single factor satisfactory factor loadings emerged ($0.54 < \lambda < 0.68$).

Table 6.6.
Rotated Factor Matrix: International Activities Scale

Item	Factor 1	Factor 2
INTI1	0.093	**0.927**
INTI2	**0.596**	0.288
INTI3	0.291	**0.593**
INTI4	**0.837**	0.025
INTI5	**0.544**	0.354

6.3.1.5 Law/Politics scale uni-dimensionality results.

The Law/Politics scale split into two clear factors distinguishing *political* from *legal* interests. The item loadings for the respective factors, obtained from the rotated factor matrix, are presented in Table 6.7. The descriptive statistics were reviewed and indicated no differential skewness. The factor fusion resulted in a sound one factor solution with good factor loadings (0.62 < λ < 0.80).

Table 6.7.
Rotated Factor Matrix: Law/Politics Scale

Item	Factor	
	1	2
LAWI1	**0.653**	0.335
LAWI2	**0.647**	0.191
LAWI3	0.409	**0.686**
LAWI4	**0.747**	0.350
LAWI5	0.280	**0.795**
LAWI6	**0.583**	0.385
LAWI7	0.274	**0.705**
LAWI8	**0.596**	0.294
LAWI9	**0.614**	0.362
LAWI10	**0.634**	0.212

6.3.1.6 Leadership scale uni-dimensionality results.

The Leadership scale split into two factors. Factor 1 constitutes a *general leading interest* factor, whereas Factor 2 represents a *public institution/government leadership* factor. The rotated factor matrix (Table 6.8) contains the items that load on the respective factors. The descriptive statistics (Table 6.9) were reviewed and indicated that LEADI2 showed significant negative skewness (p<0.05) whilst LAWI4[35] did not show significant skewness. All the other items but LEAD13 are also negatively skewed although not significantly so (p>0.05). The emergence of two factors can therefore not be attributed to differential skewness in the items. With LEAD12 and LAWI4 as the only items loading on Factor 2 the interpretation of this factor becomes somewhat tentative. LAWI4 refers to being a *state governor,* an American political position which could be misunderstood in the South African context. LEADI2 refers to leadership at an executive level in a large corporation. When forcing a single factor, all items loaded in a satisfactory manner (0.51 < λ < 0.78).

[35] LAW14 is one of the complex items in the CISS that load on more than one of the Basic Interest scales.

Table 6.8.
Rotated Factor Matrix: Leadership Scale

Item	Factor 1	Factor 2
LEADI1	**0.682**	0.243
LEADI2	0.313	**0.690**
LEADI3	**0.887**	0.098
LEADI4	**0.731**	0.289
LEADI5	**0.570**	0.360
LEADI6	**0.557**	0.288
LAWI4	0.166	**0.805**

Table 6.9.
Descriptive Statistics For Leadership Interest Scale

	Item	LEADI1	LEADI2	LEADI3	LEADI4	LEADI5	LEADI6	LAWI4
N	Valid	408	408	408	408	408	408	408
	Mean	3.603	3.956	3.493	3.632	3.625	3.838	3.588
	Std. Deviation	1.773	1.565	1.806	1.573	1.526	1.408	1.636
	Variance	3.144	2.450	3.263	2.474	2.328	1.984	2.675
	Skewness	-0.071	-0.328*	0.026	-0.181	-0.076	-0.220	-0.088
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.372*	-0.965*	-1.393*	-1.106*	-1.031*	-0.815*	-1.163*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241

6.3.1.7 Mechanical Crafts scale uni-dimensionality results.

The Mechanical Crafts scale also split into two factors. Factor 1 represents *professional engineering occupational activities* factor. Factor 2 seems to represent a *practically orientated activities around the repair of mechanical objects* factor. The rotated factor matrix (Table 6.10) shows the items that load on the respective factors. The descriptive statistics (Table 6.11) where reviewed and indicated that MECHI3, MECHI16, MECHI8 and MECHI9 showed significant negative skewness (p<0.05). Although MECHI17 demonstrated non-significant positive skewness, differential skewness does not provide a plausible explanation for the extracted factor matrix. The factor solution derived, when forcing one factor (Table 6.12), was not as robust as previous scales. Factors loadings ranged between 0.09 and 0.80. Of concern are the items that loaded onto Factor 2 in the initial solution. The loading for MECHI1 is 0.09 and MECHI7 is 0.35. MECHI1 describes *driving earthmoving equipment* and MECHI7 describes interest in *welding*. This may suggest that these two items should be regarded as constituting a separate scale reflecting a distinct mechanical interest latent variable. However, the *a priori* model will be adhered to for the CFA analyses.

Table 6.10.
Rotated Factor Matrix: Mechanical Crafts Scale

Item	Factor 1	Factor 2
MECHI1	-0.100	**0.584**
MECHI2	**0.636**	0.318
MECHI3	**0.782**	-0.057
MECHI4	**0.675**	0.446
MECHI5	**0.826**	0.031
MECHI6	**0.918**	-0.092
MECHI7	0.160	**0.699**
MECHI8	**0.704**	0.268
MECHI9	**0.583**	0.557

Table 6.11.
Descriptive Statistics for Mechanical Crafts Scale

	Item	MECHI1	MECHI2	MECHI3	MECHI4	MECHI5	MECHI6	MECHI7	MECHI8	MECHI9
N	Valid	408	408	408	408	408	408	408	408	408
	Mean	3.569	3.855	3.975	3.745	3.882	4.042	3.422	3.944	3.792
	Std. Deviation	1.794	1.432	1.676	1.523	1.518	1.727	1.587	1.589	1.553
	Variance	3.219	2.050	2.810	2.318	2.306	2.984	2.520	2.525	2.411
	Skewness	-0.060	-0.213	-0.342*	-0.151	-0.223	-0.389*	0.027	-0.280*	-0.255*
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.381*	-0.913*	-1.163*	-0.941*	-1.042*	-1.165*	-1.038*	-1.114*	-0.983*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

Table 6.12.
Factor Fusion Item Loadings: Mechanical Crafts Scale

Item	Factor 1
MECHI1	**0.095**
MECHI2	0.714

MECHI3	0.705
MECHI4	0.787
MECHI5	0.784
MECHI6	0.802
MECHI7	**0.357**
MECHI8	0.764
MECHI9	0.722

6.3.1.8 Medical Practice scale uni-dimensionality results.

The results of the dimensionality analysis (Table 6.13) for the Medical Practice scale resulted in two factors. The first factor represents *activities associated with being a doctor or medical researcher*. The second factor embodies activities associated with *primary health care providers,* for example, nurses or nutritional advisors. Upon review of the descriptive statistics MEDI1 and MEDI8 showed statistically significant negative skewness (Table 6.14). However, differential skewness does not provide a credible explanation for the extracted factor structure. Upon forcing a single factor, satisfactory item loadings were obtained ($0.62 < \lambda < 0.85$).

Table 6.13.
Rotated Factor Matrix: Medical Practice Scale

Item	Factor 1	Factor 2
MEDI1	**0.705**	0.226
MEDI2	0.185	**0.887**
MEDI3	0.374	**0.546**
MEDI4	0.549	**0.561**
MEDI5	**0.651**	0.436
MEDI6	**0.746**	0.436
MEDI7	**0.683**	0.358
MEDI8	**0.740**	0.172
MEDI9	0.195	**0.834**
MEDI10	0.474	**0.645**
MEDI11	**0.711**	0.165
MEDI12	**0.467**	0.404

Table 6.14.
Descriptive Statistics for Medical Practice Scale

	Item	MEDI1	MEDI2	MEDI3	MEDI4	MEDI5	MEDI6
N	Valid	408	408	408	408	408	408
	Mean	4.037	3.713	3.547	3.760	3.711	3.794
	Std. Deviation	1.637	1.720	1.554	1.678	1.500	1.532
	Variance	2.679	2.957	2.416	2.817	2.250	2.346
	Skewness	-0.390*	-0.214	-0.022	-0.165	-0.139	-0.170
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.085*	-1.238*	-1.118*	-1.197*	-0.905*	-0.979*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241

6.3.1.9 Military/Law Enforcement scale uni-dimensionality results.

The Military/Law Enforcement scale split into two clear factors, consistent with the scale label, Military (Factor 1) and Law Enforcement (Factor 2). The items in the Law Enforcement factor describe policing type activities. The rotated factor matrix (Table 6.15) shows the items that load on the respective factors. The descriptive statistics indicated no differential skewness. Factor fusion resulted in a sound one-factor solution ($0.45 < \lambda < 0.80$). Items MILI1 and MEDI11 did, however, return modest factor loadings (0.45 and 0.45, respectively).

Table 6.15.
Rotated Factor Matrix: Military/Law Enforcement Scale

	Factor	
Item	1	2
MILI1	0.153	**0.740**
MILI2	**0.749**	0.319
MILI3	**0.677**	0.348
MILI4	**0.587**	0.317
MILI5	**0.709**	0.268
MILI6	**0.846**	0.107
MILI7	0.449	**0.629**
MEDI11	0.197	**0.544**

6.3.1.10 Office Practices scale uni-dimensionality results.

The Office Practices scale also returned two factors. Factor 1 carries a *clerical/secretarial* theme, whereas Factor 2 represents an *office duties with a supervisory flavour* factor. The rotated factor matrix (Table 6.16) contains the items that load on the two respective factors. No differential skewness was found. Factor fusion resulted in only a somewhat satisfactory one-factor solution with item loadings ranging from 0.33 to 0.84. OPI6 provides reason for concern as the factor loading is low (0.33). This item refers to *training others on using office equipment*. As this study is confirmatory in nature, all items in this sub-scale will however be retained in further analyses.

Table 6.16.
Rotated Factor Matrix: Office Practices Scale

Item	Factor	
	1	2
OPI1	**0.484**	0.437
OPI2	**0.589**	0.176
OPI3	0.436	**0.503**
OPI4	**0.549**	0.457
OPI5	**0.672**	0.293
OPI6	0.055	**0.657**
OPI7	**0.830**	0.249
OPI8	**0.878**	0.115
SUPI8	0.279	**0.712**

6.3.1.11 Performing Arts scale uni-dimensionality results.

Dimensionality analysis for the Performing Arts scale revealed a two factor structure differentiating *music/dance orientated activities* (Factor 1) from *acting* (Factor 2). The rotated factor matrix (Table 6.17) shows the items that load on the respective factors. No statistically differential skewness was identified upon review of the descriptive statistics. Upon forcing a single factor, mostly satisfactory factor loadings emerged ($0.44 < \lambda < 0.78$).

Table 6.17.
Rotated Factor Matrix: Performing Arts Scale

Item	Factor	
	1	2
PERI1	0.096	**0.778**
PERI2	**0.647**	0.200
PERI3	0.356	**0.692**
PERI4	**0.507**	0.464
PERI5	**0.621**	0.151
PERI6	**0.712**	0.288
PERI7	**0.718**	0.156
PERI8	**0.503**	0.375
PERI9	**0.659**	0.333
PERI10	**0.588**	0.521

6.3.2 Dimensionality analysis results: interest model for the female sample

The results of the principle axis factor analysis for the female sample are summarised in Table 6.18. Twelve of the 29 sub-scales failed the uni-dimensionality test. The affected scales are: (i) Advertising, (ii) Athletics (iii) Fashion, (iv) Law/Politics, (v) Leadership, (vi) Mathematics, (vii) Mechanical Crafts, (viii) Medical Practice, (ix) Military, (x) Performing Arts, (xi) Sales and (xii) Supervision.

Some of the above results were also found for the male sample (as described in the previous section). Factors in common (that failed the uni-dimensionality assumption) include: (i) Law/Politics, (ii) Leadership, (iii) Mechanical Crafts, (iv) Medical Practice, (v) Military and (vi) Performing Arts.

Of concern was the presence of a *Heywood case* when conducting the dimensionality analysis on the Art/Design scale. A Heywood case would signal that the communality of an item exceeded unity hence causing an inadmissible factor solution (Raykov & Marcoulides, 2008). Raykov and Marcoulides (2008) also indicate that this anomaly could be due to the following conditions: (i) not enough data have been sampled from the population to provide stable parameter estimates, (ii) too many (or few) factors being extracted in the factor solution, (iii) the initial communality estimates are not appropriate, (iv) in the studied population, a corresponding variance or correlation parameter is very small or very close to unity or (v) the factor solution is not appropriate for the observed data considered.

In this scenario, it is more likely that variance of a single item corresponds to the variance observed in the remaining items. This is due to the fact that a two factor structure was initially found, however when extraction was attempted this could not be achieved due to the correspondence seen between the single item and the remaining items. A possible solution is to identify and remove the culprit item and then reattempt the analysis. This was conducted and ART1 was identified as being the problematic item. This item contains content regarding the profession of architecture. When removing this item, a single factor solution was found with factor loadings: $0.45 < \lambda < 0.84$. However, as indicated throughout this dissertation, the *a priori* measurement model would be adhered to for the CFAs. Therefore, uni-dimensionality will be assumed for this sub-scale.

6.3.2.1 Item factor loadings: interest model for the female sample.

For the female sample item loadings ranged between 0.32 and 0.96. Problematic items (loadings < 0.50) are (18 of the 200 interest items): ADVI7 (0.45), ATHI6 (0.48), FASHI4 (0.41), FASHI6 (0.33), LAWI10 (0.46), LEADI5 (0.48), MEDI1 (0.47), MEDI8 (0.47), MILI4 (0.49), PERI2 (0.49), SUPI4 (.42), SUPI6 (.41), CDEVI4 (.36), COUNI3 (.38), PUBI4 (.49), SCII2 (0.42), WOODI4 (0.39) and WRTI6 (0.32). Items FASHI6, CDEVI4, COUNI3, WOODI4 and WRTI6 were more closely examined as they obtained item loadings of less than 0.40. It is interesting that all the scales affected by these low factor loadings passed the uni-dimensionality test except the Fashion Basic scale. The factor properties of the Fashion Basic scale will be discussed in the corresponding dimensionality analysis results section. For the remaining problematic items, the following could be hypothesized as the potential sources of the problem:

- CDEVI4 (Child Development scale) describes the job title of *playground director*[36] This type of occupation may not be known in the South African

[36] The item specifically describes that the playground director arranges games and contests for children. In South Africa, this task may be included in other job titles related to children development, for example: nursery school teacher, school principal.

context or may not be consistent with the *child nurturing* theme as with the remaining items in this sub-scale.

- COUNI3 (Counseling scale) indicates an interest in studying *group dynamics* whereas the other items carry an *adult psychological/helping theme*.
- WOODI4 (Woodworking sub-scale) is different to other Woodworking items in that the items have a *classic woodworking activity* emphasis, whereas WOODI4 describes a *large renovation project worked on at home*.
- For WRTI6, the discussion for the male sample would apply (see sub-section 6.2.1.2).

The explanations aim to explicate possible causes of low factor loadings. They are nonetheless based on a qualitative review of item content and should be tested empirically. Further hypotheses could have been formulated if more information was available on the origin and constituents of the sample.

Table 6.18.
Principle Factor Analyses of CISS Basic Interest Scales for the Female Sample

Basic scale	KMO	% Variance explained	Min factor loading	Max factor loading
Adult Development	0.737	53.585	0.551	0.692
Advertising	0.873	Factor 1: 57.664	0.516	0.819
		Factor 2: 12.593	0.585	0.739
		Single forced factor: 52.039	0.531	0.829
Animal Care	0.840	80.730	0.855	0.916
Art/Design	0.734	Factor 1: 51.024	When attempting to extract 2 factors the communality of an item exceeded 1.0[a]	
		Factor 2:18.977		
Athletics	0.783	Factor 1: 48.094	0.578	0.935
		Factor 2: 16.942	0.487	0.673
		Single forced factor: 41.319	0.244	0.800
Child Development	0.870	58.470	0.368	0.901
Counseling	0.776	51.627	0.385	0.827
Culinary Arts	0.855	61.805	0.521	0.865
Farming	0.888	73.834	0.754	0.882
Fashion	0.712	Factor 1: 42.789	0.694	0.753
		Factor 2: 19.616	0.418	0.752
		Single forced factor: 33.992	0.263	0.789
Financial Services	0.913	66.649	0.707	0.843

Int'l Activities	0.768	58.100	0.622	0.756
Law/Politics	0.899	Factor 1: 53.249	0.617	0.788
		Factor 2: 11.683	0.683	0.912
		Single forced factor: 48.222	0.594	0.789
Leadership	0.746	Factor 1: 44.036	0.563	0.820
		Factor 2: 21.899	0.746	0.830
		Single forced factor: 35.266	0.351	0.700
Mathematics	0.870	Factor 1: 62.154	0.565	0.884
		Factor 2: 15.105	0.719	0.824
		Single forced factor: 56.868	0.539	0.912
Mechanical Crafts	0.918	Factor 1: 67.821	0.571	0.817
		Factor 2: 11.685	0.508	0.908
		Single forced factor: 64.074	0.621	0.891
Medical Practice	0.897	Factor 1: 50.908	0.502	0.816
		Factor 2: 10.174	0.604	0.780
		Single forced factor: 46.694	0.413	0.796
Military	0.871	Factor 1: 55.724	0.799	0.920
		Factor 2: 16.555	0.496	0.740
		Single forced factor: 51.199	0.201	0.862
Office Practices	0.888	49.279	0.539	0.749
Performing Arts	0.856	Factor 1: 48.465	0.498	0.827
		Factor 2: 13.878	0.511	0.708
		Single forced factor: 43.293	0.450	0.810
Plants	0.857	70.859	0.690	0.837
Public Speaking	0.746	60.938	0.460	0.873
Religious Activities	0.842	70.683	0.562	0.909
Risk	0.646	51.483	0.533	0.764
Sales	0.826	Factor 1: 52.847	0.663	0.781
		Factor 2: 15.472	0.551	0.838
		Single forced factor: 46.593	0.447	0.814
Science	0.884	61.742	0.420	0.897
Supervision	0.830	Factor 1: 44.626	0.574	0.780
		Factor 2: 15.001	0.510	0.591
		Single forced factor: 37.878	0.298	0.789
Woodworking	0.823	62.462	0.393	0.876
Writing	0.823	57.004	0.323	0.837

Note. [a] When removing ART1, KMO = 785, variance explained = 53.667, min factor loading = 0.450, max factor loading= 0.836.

6.3.2.2 Advertising scale uni-dimensionality results.

For the Advertising scale two factors emerged. Factor 1 focuses on *managerial/strategic elements of advertising/marketing,* whereas Factor 2 embodies interests in "*hands on" tasks performed by individuals in advertising/marketing. The results (i.e., factor loadings) of the* rotated factor matrix are presented in Table 6.19. Descriptive statistics were reviewed (Table 6.20) to determine if the two factor structure may be an artefact of item skewness. The review found that all of the items that load on Factor 2 demonstrate statistically significant positive skewness. The items that load on Factor 1, however, do not show significant negative skewness. Differential skewness therefore does not offer a plausible explanation for the extracted factors. A single factor was forced on the scale as per the *a priori* model. The factor fusion returned a one factor solution where all items obtained satisfactory loadings ($0.53 < \lambda < 0.82$).

Table 6.19.
Rotated Factor Matrix: Advertising Scale

Item	Factor 1	Factor 2
ADVI1	**0.415**	0.326
ADVI2	**0.710**	0.326
ADVI3	0.221	**0.860**
ADVI4	0.516	**0.585**
ADVI5	**0.819**	0.348
ADVI6	0.267	**0.739**
ADVI7	0.495	**0.668**
ADVI8	**0.715**	0.169

Table 6.20.
Descriptive Statistics for Advertising Scale

	Item	ADVI1	ADVI2	ADVI3	ADVI4	ADVI5	ADVI6	ADVI7	ADVI8
N	Valid	402	402	402	402	402	402	402	402
	Mean	3.530	3.381	3.092	3.179	3.348	3.114	3.271	3.507
	Std. Deviation	1.416	1.535	1.542	1.591	1.530	1.446	1.471	1.518
	Variance	2.005	2.356	2.378	2.531	2.342	2.092	2.163	2.305
	Skewness	-0.029	0.175	0.362*	0.264*	0.117	0.241*	0.285*	-0.049
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-0.766*	-0.995*	-0.855*	-0.975*	-1.038*	-0.837*	-0.783*	-0.977*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.3 Athletic scale uni-dimensionality results.

Dimensionality analysis for the Athletic scale revealed a two factor structure, differentiating *competitive sporting activities* (Factor 1) from *personal fitness* (Factor 2). The rotated factor matrix (Table 6.21) shows the items that load on the two respective factors. The descriptive statistics revealed no differential skewness. Upon forcing a single factor solution most factor loadings ($0.24 < \lambda < 0.78$) were satisfactory. However, items ATHI4 and ATHI6 returned low factor loadings (0.32; 0.24 respectively). ATHI6 was previously flagged as a problematic item (as discussed in sub-section 6.2.1.2). The low loading for ATHI4 may be attributed to the fact that this item carries a *marital arts theme* which may not be viewed in the same light as other traditional athletic activities (e.g., running, cycling).

Table 6.21.
Rotated Factor Matrix: Athletic Scale

Item	Factor	
	1	2
ATHI1	**0.578**	0.417
ATHI2	0.336	**0.673**
ATHI3	**0.935**	0.099
ATHI4	**0.325**	0.090
ATHI5	**0.830**	0.202
ATHI6	-0.008	**0.487**
ATHI7	0.481	**0.636**

6.3.2.4 Fashion scale uni-dimensionality results.

For the Fashion scale two factors emerged. Factor represents *clothing orientated fashion activities,* whereas Factor 2 embodies *general beauty/hair care*. Although. FASHI6 tends to load on the second factor it seems somewhat out of step with the common theme shared by the other Factor 2 items (FASHI3, FASHI4). FASHI6 describes an interest in appraising jewellery/antiques. The rotated factor matrix (Table 6.22) indicates the items that load on the respective factors. Descriptive statistics were reviewed (Table 6.23) to determine if the two factor structure may be an artefact of differential skewness in the item distributions. Two items (FASHI2 and FASHI5) showed statistically significant positively skewness. All the item distributions, however, tend to be positively skewed although not all significantly.

The lack of differential skewness precludes the possibility that the extracted factor structure simply is an artefact of differences in the statistical properties of the items. Upon forcing a single factor solution on the data, the majority of items loaded satisfactorily on the single factor (0.26 < λ < 0.79), however the hair care items FASHI3 and FASHI4 showed lower loadings (0.35 and 0.26 respectively). FASHI6 also returned a factor loading lower than the 0.50 cut-off. However more attention should be paid to the hair
care items.

Table 6.22.
Rotated Factor Matrix: Fashion Scale

Item	Factor 1	Factor 2
FASHI1	**0.753**	0.105
FASHI2	**0.802**	0.176
FASHI3	0.096	**0.752**
FASHI4	0.103	**0.418**
FASHI5	**0.694**	0.246
FASHI6	0.285	**0.336**

Table 6.23.
Descriptive Statistics for Fashion Scale

	Item	FASHI1	FASHI2	FASHI3	FASHI4	FASHI5	FASHI6
N	Valid	402	402	402	402	402	402
	Mean	3.231	3.104	3.425	3.303	3.077	3.393
	Std. Deviation	1.654	1.633	1.653	1.559	1.505	1.432
	Variance	2.737	2.667	2.734	2.431	2.266	2.050
	Skewness	0.175	0.290*	0.074	0.145	0.269*	0.127
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.220*	-1.165*	-1.201*	-1.034*	-0.959*	-0.839*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.5 Law/Politics scale uni-dimensionality results.

The results of the factor analysis are similar to the results obtained for the male sample. Two factors emerged that distinguish between *political* and *legal* interests. The factor loadings for the rotated factor matrix are presented in Table 6.24. Inspection of the descriptive statistics (Table 6.25) revealed that two items; LAWI4 and LAWI6, showed significant negative skewness. However, the remaining item distributions also tended to be negatively skewed although not significantly. The extracted factor structure could therefore not be attributed to differential skewness. Factor fusion resulted in a sound one factor solution ($0.59 < \lambda < 0.78$). Of interest is that LAWI8 loaded on the legal factor as apposed to the political factor as seen with the male sample. The content of LAWI8's concerns negotiating conflicts between irate people. This might suggest a difference between the genders in how conflict should be handled (politically or legally).

Table 6.24.
Rotated Factor Matrix: Law/Politics Scale

Item	Factor	
	1	2
LAWI1	**0.617**	0.376
LAWI2	**0.658**	0.223
LAWI3	0.380	**0.683**
LAWI4	**0.763**	0.315
LAWI5	0.143	**0.912**
LAWI6	**0.788**	0.249
LAWI7	0.345	**0.639**
LAWI8	0.378	**0.469**
LAWI9	**0.785**	0.262
LAWI10	**0.468**	0.389

Table 6.25.
Descriptive Statistics for Law/Politics Scale

	Item	LAWI1	LAWI2	LAWI3	LAWI4	LAWI5	LAWI6	LAWI7	LAWI8	LAWI9	LAWI10
N	Valid	402	402	402	402	402	402	402	402	402	402
	Mean	3.624	3.794	3.624	3.883	3.597	3.915	3.592	3.550	3.759	3.754
	Std. Deviation	1.643	1.539	1.799	1.712	1.779	1.825	1.700	1.464	1.879	1.640
	Variance	2.699	2.369	3.238	2.931	3.164	3.329	2.891	2.143	3.530	2.690
	Skewness	-0.045	-0.162	-0.028	-0.282*	-0.053	-0.297*	-0.038	-0.031	-0.210	-0.193
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.151*	-0.986*	-1.413*	-1.229*	-1.389*	-1.342	-1.231*	-0.973*	-1.443*	-1.137*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.6 Leadership scale uni-dimensionality results.

As with the male sample, the Leadership scale split into two factors. Factor 1 represents a *general leading interest,* whereas Factor 2 embodies activities related to *leading a large public institution*. Table 6.26 contains the item factor loadings as listed in the rotated factor matrix. The item descriptive statistics (Table 6.27) were reviewed and indicated that LEADI2 and LAWI4 showed significant negative skewness. The consistency in the distributional form across all items, however, makes it unlikely that the two factor structure emerged simply as an artefact of differential skewness. When forcing a single factor solution most loadings were satisfactory ($0.35 < \lambda < 0.70$) except for LAWI4 (refers to being a *state governor,* an American political position which could be misunderstood in the South African context). A similar trend was observed in the male sample for this item (i.e., low loading of 0.35).

Table 6.26.
Rotated Factor Matrix: Leadership Scale

Item	Factor	
	1	2
LEADI1	**0.527**	0.430
LEADI2	0.170	**0.830**
LEADI3	**0.820**	0.001
LEADI4	**0.812**	-0.010
LEADI5	**0.483**	0.351
LEADI6	**0.563**	0.221
LAWI4	0.013	**0.746**

Table 6.27.
Descriptive Statistics for Leadership Scale

	Item	LEADI1	LEADI2	LEADI3	LEADI4	LEADI5	LEADI6	LAWI4
N	Valid	408	408	408	408	408	408	402
	Mean	3.603	3.956	3.493	3.632	3.625	3.838	3.883
	Std. Deviation	1.773	1.565	1.806	1.573	1.526	1.408	1.712
	Variance	3.144	2.450	3.263	2.474	2.328	1.984	2.931
	Skewness	-0.071	-0.328*	0.026	-0.181	-0.076	-0.220	-0.282*
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.122
	Kurtosis	-1.372*	-0.965*	-1.393*	-1.106*	-1.031*	-0.815*	-1.229*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.2428

6.3.2.7 Mathematics scale uni-dimensionality results.

The results of the dimensionality analysis for the Mathematics scale revealed a two factor structure, differentiating *classical mathematical interests (e.g., algebra, mathematical puzzles)* (Factor 1) from *computer science interests* (Factor 2). The factor loadings for the rotated factor matrix are presented in Table 6.28. The item descriptive statistics (Table 6.29) revealed that all the Mathematics items were significantly negatively skew. Upon forcing a single factor solution all factor loadings ($0.54< \lambda < 0.91$) were above 0.50.

Table 6.28.
Rotated Factor Matrix: Mathematics Scale

Item	Factor	
	1	2
MATHI1	0.249	**0.719**
MATHI2	**0.760**	0.211
MATHI3	**0.884**	0.223
MATHI4	0.250	**0.824**
MATHI5	**0.872**	0.303
MATHI6	**0.565**	0.348
MATHI7	**0.802**	0.312

Table 6.29.
Descriptive statistics for Mathematics scale

	Item	MATHI1	MATHI2	MATHI3	MATHI4	MATHI5	MATHI6	MATHI7
N	Valid	402	402	402	402	402	402	402
	Mean	4.037	4.025	3.978	4.047	3.886	3.955	3.871
	Std. Deviation	1.635	1.738	1.885	1.615	1.870	1.693	1.799
	Variance	2.674	3.022	3.553	2.609	3.498	2.866	3.235
	Skewness	-0.404*	-0.336*	-0.392*	-0.395*	-0.311*	-0.277*	-0.255*
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-0.986*	-1.225*	-1.334*	-1.007*	-1.361*	-1.200*	-1.379*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.8 Mechanical Crafts scale uni-dimensionality results.

The Mechanical Crafts scale split into two factors. Factor 1 embodies *traditional trades (e.g., welder, electrician).* Factor 2 comprises of items that specific indicate the word “engineering” in the majority of cases. Table 6.30 contains the rotated factor matrix results. The item descriptive statistics (Table 6.31) where reviewed and indicated that all items, except item MECHI3, showed significant negative skewness. The two factor solution is therefore likely to be due to the fact that mechanical Interest latent variable comprises two meaningful sub-themes and not merely an artefact of differential skewness. Upon forcing a single factor solution, all item loadings were satisfactory ($0.62 < \lambda < 0.89$).

When comparing the male and female samples, the factor structures were fairly different. For males, Factor 1 indicates *diagnostic type engineering occupational activities* and Factor 2: *practically orientated activities around the repair of mechanical objects*. Of further interest is that when forcing a single factor, the female sample's item loadings showed satisfactory results however, for the males two items (MECHI1, MECHI7) showed low factor loadings. As indicated previously, MECHI1 indicates *driving earthmoving equipment* and MECHI7 describes interest in *welding*. Therefore, it seems like females may not distinguish these interests as much as men.

Table 6.30.
Rotated Factor Matrix: Mechanical Crafts Scale

Item	Factor	
	1	2
MECHI1	**0.790**	0.248
MECHI2	**0.727**	0.508
MECHI3	0.229	**0.716**
MECHI4	**0.801**	0.396
MECHI5	0.571	**0.677**
MECHI6	0.284	**0.908**
MECHI7	**0.817**	0.243
MECHI8	0.466	**0.625**
MECHI9	**0.780**	0.407

Table 6.31.
Descriptive Statistics for Mechanical Crafts Scale

	Item	MECHI1	MECHI2	MECHI3	MECHI4	MECHI5	MECHI6	MECHI7	MECHI8	MECHI9
N	Valid	402	402	402	402	402	402	402	402	402
	Mean	4.127	4.192	3.729	4.080	4.047	3.953	3.990	3.905	4.080
	Std. Deviation	1.908	1.774	1.639	1.807	1.673	1.794	1.925	1.760	1.800
	Variance	3.642	3.148	2.687	3.265	2.798	3.217	3.706	3.098	3.241
	Skewness	-0.496*	-0.526*	-0.153	-0.420*	-0.454*	-0.309*	-0.397*	-0.263*	-0.455*

Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
Kurtosis	-1.274*	-1.155*	-1.179*	-1.258*	-1.038*	-1.353*	-1.375*	-1.281*	-1.187*
Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.9 Medical Practice scale uni-dimensionality results.

The dimensionality analysis for the Medical Practice scale revealed a two factor structure differentiating *primary health care activities (e.g., nursing, emergency care)* (Factor 1) and *providing medical advice through consultation (e.g., nutritional advice, physical therapy, doctor),* (Factor 2). The rotated factor matrix (Table 6.32) shows the items that load on the respective factors. The item descriptive statistics indicated no differential skewness. Upon forcing a single factor solution, satisfactory factor loadings emerged ($0.62 < \lambda < 0.85$).

Results for the male sample contrast with the above mentioned results. For males, Factor 1 indicated *activities associated with being a doctor or medical researcher*. The second factor embodies activities associated with *primary health care providers;* for example, nurses or nutritional advisors. It seems that the females place doctors with other healthcare practitioners, whereas the males seem to distinguish between doctors, as researchers (scientists), and healthcare providers (e.g., nurses). However, with both samples, when forcing a single factor, satisfactory loading are observed.

Table 6.32.
Rotated Factor Matrix: Medical Practice Scale

Item	Factor 1	Factor 2
MEDI1	**0.477**	0.340
MEDI2	**0.816**	0.160
MEDI3	0.074	**0.604**
MEDI4	**0.702**	0.284
MEDI5	0.315	**0.780**
MEDI6	0.458	**0.732**
MEDI7	**0.502**	0.435
MEDI8	**0.475**	0.430
MEDI9	**0.786**	0.146

MEDI10	**0.699**	0.360
MEDI11	**0.572**	0.394
MEDI12	**0.515**	0.425

6.3.2.10 Military/Law Enforcement scale uni-dimensionality results.

As with the male sample, the Military/Law Enforcement scale split into two clear factors as indicated by the sub-scale label, Military (Factor 1) and Law Enforcement (Factor 2). The items on the Law Enforcement factor describe police type activities. The rotated factor matrix (Table 6.33) indicates how the items on the respective factors. The descriptive statistics (Table 6.34) revealed that all the sub-scale items were negatively skewed but for the first item. Therefore differential skewness does not offer a plausible explanation for the extracted solution. The factor fusion resulted in a one factor solution ($0.20 < \lambda < 0.87$) with items MILI1 and MEDI11 returning lower factor loadings (0.40; 0.20 respectively). Item MEDI11 was flagged as problematic in the item reliability analysis (sub-section 5.4.1.1). MEDI11 contains content referring to *helping in an emergency situation* which may not been seen in the same light as *traditional military activities* (e.g., an interest in commanding a military unit or enjoying a military drill). However, the item is once again retained in further analysis.

An interesting difference between the two samples is that the loading for MEDI11, on the single forced factor for the male sample, was 0.45. A possible reason could be that males might view *military/law enforcement* and *helping in an emergency situation* as being heroic acts.

Table 6.33.
Rotated Factor Matrix: Military/Law Enforcement Scale

Item	Factor	
	1	2
MILI1	0.111	**0.740**
MILI2	**0.893**	0.167
MILI3	**0.799**	0.326
MILI4	0.492	**0.496**
MILI5	**0.841**	0.194

MILI6	**0.920**	0.131
MILI7	0.470	**0.703**
MEDI11	0.043	**0.361**

Table 6.34.
Descriptive Statistics for Military/Law Enforcement scale

	Item	MILI1	MILI2	MILI3	MILI4	MILI5	MILI6	MILI7	MEDI11
N	Valid	402	402	402	402	402	402	402	402
	Mean	3.291	4.002	3.883	3.689	4.005	3.973	3.585	3.595
	Std. Deviation	1.697	1.933	1.900	1.653	1.870	1.960	1.796	1.581
	Variance	2.880	3.738	3.610	2.733	3.496	3.842	3.226	2.501
	Skewness	0.212	-0.384*	-0.283*	-0.055	-0.401*	-0.369*	-0.009	-0.020
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.240*	-1.412*	-1.440*	-1.181*	-1.304*	-1.443*	-1.385*	-1.141*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.2.11 Performing Arts scale uni-dimensionality results.

For the Performing Arts scale two factors emerged. Factor 1 focuses on *performance with a team/group,* whereas Factor 2 indicates interests in *solo performance* (see Table 6.36 for the respective factor loadings). No differential item skewness was evident upon investigation of the individual item descriptive statistics. Upon forcing a single factor solution, satisfactory factor loadings emerged ($0.45 < \lambda < 0.81$). Three items did, however, return loadings lower than 0.50: PERI2 (0.48), PERI5 (0.47) and PERI7 (0.45). These loading could be viewed as borderline acceptable.

Table 6.35.
Rotated Factor Matrix: Performing Arts Scale

Item	Factor	
	1	2
PERI1	**0.827**	0.032
PERI2	**0.494**	0.133
PERI3	**0.688**	0.193
PERI4	**0.532**	0.511
PERI5	0.131	**0.708**
PERI6	0.498	**0.637**
PERI7	0.114	**0.693**
PERI8	**0.580**	0.285

PERI9	**0.641**	0.371
PERI10	**0.795**	0.269

6.3.2.12 Sales scale uni-dimensionality results.

The results of the dimensionality analysis for the Sales scale revealed a clear two factor structure differentiating *external selling (sourcing clients)* (Factor 1) from *internal sales (retail)* (Factor 2). The rotated factor matrix with item loadings is presented in Table 6.36. No differential item skewness was observed. Upon forcing a single factor solution on the data, satisfactory factor loadings emerged ($0.44 < \lambda < 0.85$). SALI7 obtained a factor loading of 0.44. This may be due to the item content describing the management of a speciality retail outlet.

Table 6.36.
Rotated Factor Matrix: Sales Scale

Item	Factor	
	1	2
SALI1	**0.708**	0.165
SALI2	0.420	**0.551**
SALI3	**0.701**	0.391
SALI4	**0.731**	0.309
SALI5	**0.663**	0.195
SALI6	**0.781**	0.156
SALI7	0.111	**0.682**
SALI8	0.275	**0.838**

6.3.2.13 Supervision scale uni-dimensionality results.

The Supervision scale also split into two factors. Factor 1 represents *supervision of staff that do <u>not</u> deal with clients face-to-face* whereas Factor 2 entails *supervision in client facing industries, (e.g., retail or hospitality).* One item, SUPI7, describes the supervisory task of interviewing job applicants. Table 6.37 contains the rotated factor matrix with respective item loadings. Differential skewness does offer a plausible alternative hypothesis for the extracted factor solution. SUPI1, SUP2, SUPI5 and SUPI8 are negatively skewed (although not significantly) and load on Factor 1 whereas SUPI, SUP4, SUPI6 and SUPI7 are all positively skewed (although not significantly so) and three of these items load on Factor 2. However, the meaning suggested by the shared item content seems to provide a more convincing explanation for the factor structure than the skewness artefact hypothesis (Table 6.38). Factor fusion resulted in a one factor solution (0.49 < λ < .071) with SUPI6 returning a negligible low loading of 0.49.

Table 6.37.
Rotated Factor Matrix: Supervision Scale

Item	Factor	
	1	2
SUPI1	**0.712**	0.131
SUPI2	**0.780**	0.009
SUPI3	**0.574**	0.510
SUPI4	0.421	**0.424**
SUPI5	**0.603**	0.292
SUPI6	0.144	**0.413**
SUPI7	0.019	**0.591**
SUPI8	**0.649**	0.431

Table 6.38.
Descriptive Statistics for Supervision Scale

	Item	SUPI1	SUPI2	SUPI3	SUPI4	SUPI5	SUPI6	SUPI7	SUPI8
N	Valid	402	402	402	402	402	402	402	402
	Mean	3.751	3.998	3.438	3.463	3.853	3.318	3.316	3.639
	Std. Deviation	1.654	1.670	1.566	1.566	1.740	1.532	1.514	1.564
	Variance	2.736	2.791	2.451	2.454	3.028	2.347	2.291	2.446
	Skewness	-0.190	-0.348*	0.018	0.050	-0.224	0.190	0.112	-0.130
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.176*	-1.136*	-1.087*	-1.072*	-1.288*	-0.922*	-1.025*	-1.068*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

6.3.3 Residual correlations

When the reproduced correlation matrix is subtracted from the observed correlation matrix the result is referred to as the residual correlation matrix. While it would be desirable for this difference to be zero it is more likely to be observed only in the perfect dataset with perfect analysis (Gorsuch, 2003). In real data the best outcome would be to obtain a limited number of large residual correlations.

The discrepancy between observed and reproduced inter-item correlations (provided by SPSS) were examined for each sub-scale to evaluate the credibility of the extracted factor solutions. The degree to which the residuals are nonzero would be suggestive that an alternative factor structure could explain the data better or there could be change variations in the correlations due to sampling error (Gorsuch, 2003). A small percentage of nonredundant residuals with absolute values greater than 0.05 would suggest that the extracted factor solution provides a convincing explanation for the observed inter-item correlation matrix. In the case where factor fission occurred, the residual correlations were also examined for the forced single factor solution.

A summary indicating percentage of large residual correlations per scale are provided in Table 6.39. Overall, many of the scales returned residual percentages lower that 25%. However, there are a number of scales that returned higher percentages and these provide some reason for concern. These scales include:

- For male sample: Adult Development (66 percent), Athletics (52 percent), Culinary Arts (40 percent), Financial Services (35 percent), Mathematics (33 percent), Military (28 percent), Performing Arts (40 percent), Plants (30 percent), Religious Activities (30 percent), Sales (32 percent), Science (42 percent), Supervision (35 percent), and Writing (53 percent).
- For female sample: Athletics (52 percent), Counseling (60 percent), Culinary Arts (60 percent), Fashion (26 percent), International Activities (40 percent), Medical Practices (31 percent), Offices Practices (47 percent), Performing Arts (42 percent), Plants (30 percent), Religious Activities (40 percent), Risk (100 percent), Sales (32 percent), and Writing (46 percent). Again the Art/Design scale is difficult to interpret as the communality of a variable exceeded 1 and extraction was terminated. However, this would substantiate a model review of the Art/Design scale for the female sample.

For the above, alternative item-factor configurations may elicit a more suitable explanation for inter-item correlations.

In the instances where a single factor was forced, interesting observations were made. For the male sample: initial solutions that suggested a factor fissure returned lower than 25 percent of residuals. In some instances zero large residuals were observed (Art/Design, Counseling and International Activities). However, when forcing a single factor all applicable scales returned large residuals greater than 25 percent. This would suggest that the initial factor fissure solutions would provide a more plausible explanation for the observed inter-item correlation matrix. A similar observation was made for the female sample.

While a thorough analysis of the factor structures and hypothesized factor content was conducted, the foregoing discussion would suggest that the initial solutions would seem to summarise the data best. However, in the instances mentioned, certain scales could be reviewed to improve the overall factor solution.

Table 6.39.
Percentage of Nonredundant Residuals per Basic Interest Scale

	Male Sample	Female Sample
Basic Interest Scale	% Nonredundant Residuals	% Nonredundant Residuals
Adult Development	Initial Solution: 66	Initial Solution: 0
Advertising	Initial Solution: 21	Initial Solution: 25 Single forced factor: 71
Animal Care	Initial Solution: 0	Initial Solution: 0
Art/Design	Initial Solution: 0 Single forced factor: 80	Cannot be determined due to Heywood case observed.
Athletics	Initial Solution: 52	Initial Solution: 52 Single forced factor: 52
Child Development	Initial Solution: 19	Initial Solution: 23
Counseling	Initial Solution: 0 Single forced factor: 66	Initial Solution: 60
Culinary Arts	Initial Solution: 40	Initial Solution: 33
Farming	Initial Solution: 26	Initial Solution: 13
Fashion	Initial Solution: 46	Initial Solution: 26 Single forced factor: 60
Financial Services	Initial Solution: 35	Initial Solution: 32
Int’l Activities	Initial Solution: 0 Single forced factor: 60	Initial Solution: 40
Law/Politics	Initial Solution: 11 Single forced factor: 53	Initial Solution: 22 Single forced factor: 60
Leadership	Initial Solution: 14	Initial Solution: 4

	Single forced factor: 57	Single forced factor: 57
Mathematics	Initial Solution: 33	Initial Solution: 0
		Single forced factor: 28
Mechanical Crafts	Initial Solution: 16	Initial Solution: 5
	Single forced factor: 66	Single forced factor: 61
Medical Practice	Initial Solution: 22	Initial Solution: 31
	Single forced factor: 54	Single forced factor: 48
Military	Initial Solution: 28	Initial Solution: 0
	Single forced factor: 67	Single forced factor: 75
Office Practices	Initial Solution: 16	Initial Solution: 47
	Single forced factor: 44	
Performing Arts	Initial Solution: 40	Initial Solution: 42
	Single forced factor: 62	Single forced factor: 64
Plants	Initial Solution: 30	Initial Solution: 30
Public Speaking	Initial Solution: 0	Initial Solution: 0
Religious Activities	Initial Solution: 30	Initial Solution: 40
Risk	Initial Solution: 0	Initial Solution: 100
Sales	Initial Solution: 32	Initial Solution: 32
		Single forced factor: 64
Science	Initial Solution: 42	Initial Solution: 23
Supervision	Initial Solution: 35	Initial Solution: 25
		Single forced factor: 50
Woodworking	Initial Solution: 20	Initial Solution: 10
Writing	Initial Solution: 53	Initial Solution: 46

Note. Figures reported in this table should be interpreted as percentages.

6.3.4 Conclusions derived from the item and dimensionality analyses

The purpose of the foregoing analyses was to provide insight into the CISS Basic Interest scales functioning. Further to this, the analyses assist in gaining an understanding about the psychometric integrity of the indicator variables that were tasked to represent each of the latent Basic interest variables. As item parcels were chosen as the variable type for the CFAs, the item and dimensionality analyses provide valuable information with regards to the original items as per the indicated measurement model.

The item analyses revealed that sufficient internal consistency has been established for the CISS Basic Interest scales. In many cases, the scales achieved alpha values exceeding 0.80 in both samples. This does provide some evidence in support of the homogeneity of each scale as proposed by the test publishers. At a more detailed level, the item statistics revealed some items that were flagged as being potentially problematic. Upon investigating item content, hypothetical explanations could justify why these items could be problematic. However, considering that the Basic Interest scales are measured through 200 items, it could be expected that some items may not perform as well as others.

As far as the dimensionality analyses are concerned, a number of observations were made. While many factors passed the uni-dimensionality assumption, a fair number did not. In the event of factors splitting, a forced single factor solution was attempted. In many of these cases the factor loadings were satisfactory (> 0.50). Differential skewness was examined to determine whether the factor fissure was due to significant skewness. Artefact factors as a result of skewness were not found as a plausible explanation for factor fissure.

Of major interest 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, provide a more convincing explanation for the observed inter-item correlation matrix. This is suggestive that these factors could be better explained by further sub-facets of the respective interest.

Based on this information, it may be expected that model fit would be jeopardized. The extent to which the CISS can successfully measure the constitutively defined interest construct may not be completely as per the design intention in some cases. However, this is a South African specific study and the researcher aims to provide this information as observations of how local data presents itself. Therefore, the constructs could be explained differently in the South African context.

However, conclusions on how the data fits the measurement model can only be provided in the CFAs as reported in the following section. These initial analyses do, however, provide mitigating circumstances for poor model fit.

6.4 RESULTS OF BASIC INTEREST SCALES CFA: COMBINED SAMPLE

As indicated in the operational hypotheses, the Basic Interest measurement model was subjected to the combined sample data. This analysis specifically aims to evaluate whether the first-order Basic scales interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the combined sample (operational hypothesis 1).

Operational hypothesis 1 is tested by testing H_{01}: RMSEA = 0. A holistic descriptive evaluation of model fit, based on the array of model fit indices as reported by LISREL, is subsequently provided. Thereafter, an examination of the standardized residuals, factor loadings, squared multiple correlations as well as the latent variable inter-correlations, is reported.

The detailed output of the confirmatory factor analyses is electronically available (on the included CD, folder: CFA, Basic Interest Scales) in Appendix A.

6.4.1 Overall fit assessment

Upon fitting the data of the combined sample (*N*=810) to the Basic Interest model the fit indices indicated in Table 6.40 were obtained. An admissible final solution of parameter estimates was obtained after 25 iterations.

The chi-square test statistic provides information regarding the difference between the observed and estimated covariance matrices as a function of sample size (Pousette & Hanse, 2002). However, the Satorra-Bentler (Satorra & Bentler, 1999) chi-square result is interpreted as it is better suited to multivariate non-normal data. This statistic is the result of the use of RML estimation. The Satorra-Bentler χ^2 test statistic (6276.66) is significant ($p<0.05$) thus resulting in a rejection of the null hypothesis of exact model fit [$H_0:\Sigma= \Sigma(\theta)$]. This indicates that the Basic Interest model is not able to reproduce the observed covariance matrix to a degree of accuracy that could be explained in terms of sampling error only. However, the result is not surprising considering that the χ^2 statistic is distributed asymptotically as a χ^2 distribution. This causes the frustrating dilemma that just at the point where the distributional assumption of the test statistic becomes tenable the statistical power of the test also becomes extremely high. It therefore becomes extremely unlikely to obtain the desired insignificant χ^2 statistic in a large sample even when the model fits the empirical data quite well. Given the sample size involved in this study it therefore seems somewhat premature to conclude poor model fit based on the large and significant χ^2 alone.

Table 6.40.
Goodness-Of-Fit Indicators for Combined Sample: Basic Interest Model

Degrees of Freedom = 1189
Minimum Fit Function Chi-Square = 6063.34 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 6895.32 (P = 0.0)
Satorra-Bentler Scaled Chi-Square = 6276.66 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 5087.66
90 Percent Confidence Interval for NCP = (4843.23 ; 5339.40)
Minimum Fit Function Value = 7.49
Population Discrepancy Function Value (F0) = 6.29
90 Percent Confidence Interval for F0 = (5.99 ; 6.60)
Root Mean Square Error of Approximation (RMSEA) = 0.073
90 Percent Confidence Interval for RMSEA = (0.071 ; 0.075)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00
Expected Cross-Validation Index (ECVI) = 9.05
90 Percent Confidence Interval for ECVI = (8.75 ; 9.36)
ECVI for Saturated Model = 4.23
ECVI for Independence Model = 161.77
Chi-Square for Independence Model with 1653 Degrees of Freedom = 130759.95
Independence AIC = 130875.95
Model AIC = 7320.66
Saturated AIC = 3422.00
Independence CAIC = 131206.38
Model CAIC = 10294.52
Saturated CAIC = 13169.63
Normed Fit Index (NFI) = 0.95
Non-Normed Fit Index (NNFI) = 0.95

Parsimony Normed Fit Index (PNFI) = 0.68
Comparative Fit Index (CFI) = 0.96
Incremental Fit Index (IFI) = 0.96
Relative Fit Index (RFI) = 0.93
Critical N (CN) = 169.25
Root Mean Square Residual (RMR) = 1.33
Standardized RMR = 0.063
Goodness of Fit Index (GFI) = 0.77
Adjusted Goodness of Fit Index (AGFI) = 0.67
Parsimony Goodness of Fit Index (PGFI) = 0.54

Consequently it may be considered unrealistic to expect the null hypothesis of exact fit to be tenable (Browne & Cudeck, 1993).

Pousette and Hanse (2002) suggest that the chi-square statistic should be treated as a descriptive badness-of-fit measure. This can be achieved through using the normed χ^2 measure to identify inappropriate models. Normed χ^2 values less than 1.0 indicate an "overfitted" model (Schumacker & Lomax, 1996) whilst ratio values more than 2.0 (or the more liberal limit of 5.0) indicate that the model does not fit the observed data and needs improvement (Pousette & Hanse, 2002). For the combined sample, the normed χ^2 expressed as the Satorra-Bentler χ^2 estimate expressed in terms of the degrees of freedom (χ^2/df = 5.28) suggests that the model does not fit the observed data (even at the liberal level). However, Kelloway (1998) indicates that these guidelines have little empirical justification and recommends against a strong reliance on this indicator.

When assuming that this measurement model only approximates the dynamics underlying the CISS that created the observed covariance matrix, the χ^2 test statistic will follow a non-central χ^2 distribution with noncentrality parameter, λ. The estimated λ value (5087.66) reflects the estimated discrepancy between the observed (Σ) and estimated population covariance [$\Sigma(\theta)$] matrices (Diamantopoulos & Siguaw, 2000). Based on a 90 percent confidence interval (4843.23 ; 5339.40) the λ value falls closer to the upper limit of the interval. The high value attained would indicate a higher level of discrepancy between Σ and $\Sigma(\theta)$ at a significance level of p<0.10.

RMSEA determines the error due to approximation, per degree of freedom of the model (i.e., the discrepancy between Σ and $\Sigma(\theta)$ per degree of freedom) . According to Diamantopoulos and Siguaw (2000), it is regarded as one of the most informative fit indices as it takes model complexity into consideration. The RMSEA value for the combined sample is 0.073. This value fails to meet the criterion for close fit (RMSEA ≤ 0.05). The confidence interval for this index is (0.071; 0.075). Confidence intervals assist in assessing the precision of the fit statistics. For example, a small RMSEA value with a large confidence interval indicates the estimated discrepancy value is quiet imprecise, thereby negating any possibility to determine accurately the degree of fit in the population. On the other hand, small intervals indicate a higher level of precision in reflecting the model fit in the population (Byrne, 2001). A test of close fit is also performed by LISREL to determine the probability of obtaining a RMSEA value of 0.05 in the sample given the assumption that the model fits closely in the population. The exceedence probability for the test of close fit is smaller than 0.05. Hence, H_{01} is rejected. Strictly speaking, therefore the study fails to find support for operational hypothesis 1. The array of fit statistics available in Table 6.40 will, however, was also interpreted to obtain a descriptive evaluation of the degree to which the model fits the data in the combined sample.

The expected cross-validation index (ECVI) expresses the difference between the reproduced sample covariance matrix $\hat{\Sigma}$ derived from fitting the model on the sample at hand, and the expected covariance matrix that would be obtained in an independent sample of the same size, from the same population (Byrne, 1998; Diamantapolous & Siguaw, 2000). It, therefore, focuses on the difference between $\hat{\Sigma}$ and Σ. Since the model ECVI (9.05) is smaller than the value obtained for the independence model (161.77) but larger than the ECVI value associated with the saturated model (4.23), a model more closely resembling the saturated model seems to have a better chance of being replicated in a cross-validation sample than the fitted model.

The assessment of parsimonious fit acknowledges that model fit can always be improved by adding more 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). In defining and fitting models it would seem essential to find the most parsimonious model that achieves satisfactory fit with as few model parameters as possible (Jöreskog & Sörbom, 1993). The parsimonious normed fit index (PNFI = 0.68) and the parsimonious goodness-of-fit index (PGFI = 0.54) approach model fit from this perspective. PNFI and PGFI range from 0 to 1, with higher values indicating a more parsimonious fit. However, neither index is likely to reach the 0.90 cut-off used for other fit indices, and there is also no standard for how high either index should be to indicate parsimonious fit (Kelloway, 1998). According to Kelloway (1998) these indices are more meaningfully used when comparing two competing theoretical models and are not very useful indicators in this CFA analysis.

The values for this model's Aiken information criterion (AIC= 7320.66) suggest that the fitted measurement model provides a more parsimonious fit than the independent/null model (130875.95) but not the saturated model (3422.00) since smaller values on these indices indicate a more parsimonious model (Kelloway, 1998). Values for the consistent Aiken information criterion (10294.52) imply that the fitted measurement model provides a more parsimonious fit than both the independent/null model (131206.38) but once again not with the saturated model (13169.63). This, in conjunction with the ECVI results, indicates that the measurement model does lack influential paths.

Indices of comparative fit that use as a baseline an independence or null model, contrast the ability of the model to reproduce the observed covariance matrix with that of a model known *a priori* to fit the data poorly. In other words a model that postulates no paths between the variables in the model (Diamantopoulos & Siguaw, 2000). The fit indices presented include the normed fit index (NFI= 0.95), the non-normed fit index (NNFI= 0.95), the comparative fit index (CFI= 0.96), the incremental fit index (IFI=0.96) and the relative fit index (RFI =0.93).The closer the values are to unity, the better the fit. However, 0.90 could be considered indicative of a well fitting model (Diamantopoulos & Siguaw, 2000; Kelloway, 1998). In the current results, all of these indices exceed the 0.90 level, which would be indicative of satisfactory comparative fit relative to the independence model.

The critical sample size statistic (CN) refers to the size that the sample would have to reach in order to accept the χ^2 statistic as significant at the 0.05 significance level (Diamantopoulos & Siguaw, 2000). The estimated CN value (169.25) falls below the recommended threshold value of 200. This threshold is regarded as indicative of the model providing an adequate representation of the data (Diamantopoulos & Siguaw, 2000) although this proposed threshold should be used with caution (Hu & Bentler, 1995).

The standardized RMR may be considered a summary measure of standardized residuals which represents the average difference between the elements of the sample covariance matrix and the fitted covariance matrix. If the model fit is good, the fitted residuals ($S - \hat{\Sigma}$) should be small in comparison to the magnitude of the elements in S (Diamantopoulos & Siguaw, 2000). The RMR (1.33) and standardized RMR (0.063) indicate reasonable fit as values less than 0.05 on the latter index suggest the model fits the data well (Kelloway, 1998).

The goodness-of-fit index (GFI) and the adjusted goodness-of-fit index (AGFI) reflect how closely the model comes to perfectly reproducing the sample covariance matrix (Diamantopoulos & Siguaw, 2000). The AGFI (0.67) adjusts the GFI (0.77) for the degrees of freedom in the model (Diamantopoulos & Siguaw, 2000; Jöreskog & Sörbom, 1993) and should be between zero and 1.0 with values exceeding 0.9 indicating that the model fits the data well (Jöreskog & Sörbom, 1993; Kelloway, 1998). For the fit of this model, both the GFI and AGFI are below the acceptable cut-off level. Kelloway (1998), however, states that GFI and AGFI should be used with some circumspection as guidelines for the interpretation are grounded in experience and therefore somewhat arbitrary.

In conclusion, when the abovementioned model fit statistics are considered holistically they seem to suggest reasonable, but not close fit. In addition, the model does outperform the independence model but not the saturated model and therefore model fit may benefit from the inclusion of a number of additional paths. The latter finding seems to suggest that the design intention of the CISS that each sub-scale should only reflect a single, specific latent Basic Interest did not fully succeed.

6.4.2 Examination of residuals

Residuals represent the differences between corresponding cells in the observed and fitted covariance matrices (Diamantopoulos & Siguaw, 2000; Jöreskog & Sörbom, 1993). As such, residuals - and especially standardized residuals - provide valuable diagnostic information on sources of lack of fit in models (Jöreskog & Sörbom, 1993; Kelloway, 1998). Standardized residuals can be interpreted as standard normal deviates (i.e., *z*-scores). Large positive and negative standardized residuals with absolute values greater than 2.58 would be indicative of relationships (or the lack thereof) between indicator variables that the model fails to explain (Diamantopoulos & Siguaw, 2000). Large positive residuals would indicate that the model underestimates the covariance between two observed variables. Adding paths to the model that could account for the covariance may remedy the problem. Conversely, large negative residuals would indicate that the model overestimates the covariance between specific observed variables. Rectifying this situation would therefore lie in removing some of the paths that are associated with the indicator variables in question (Diamantopoulos & Siguaw, 2000; Kelloway, 1998).

```
- 2|754421
 - 1|9887665442211110000000
 - 0|99999999999998888888888888888888888877777777777777777777777777777777776+97
   0|11111111111111111111111111111111111111111111111111111111111111111111111+99
   1|00000000011112333569
   2|3
   3|
   4|5
   5|
   6|
   7|5
```

Figure 6.1. Stem-and-leaf plot of standardized residuals for the combined sample: Basic Interest model.

The stem-and-leaf plot (Figure 6.1) provides graphical information regarding the sample standardized residual distribution. In this case, the results do confirm the foregoing discussion regarding the reasonable to mediocre fit of this model and the need for additional paths. Although the median of the distribution is zero, the distribution is positively skewed. Only a small number of large negative standardized residuals occur (the smallest negative residual is-2.75) but a far greater number of large positive standardized residuals occur (largest positive residual 7.50). Overall, the prevalence of large positive and the small number of large negative residuals would suggest that the observed covariance terms in the observed covariance matrix are typically underestimated by the derived model parameter estimates. Adding paths to the model may rectify the problem. This possibility was further reviewed in the modification indices section.

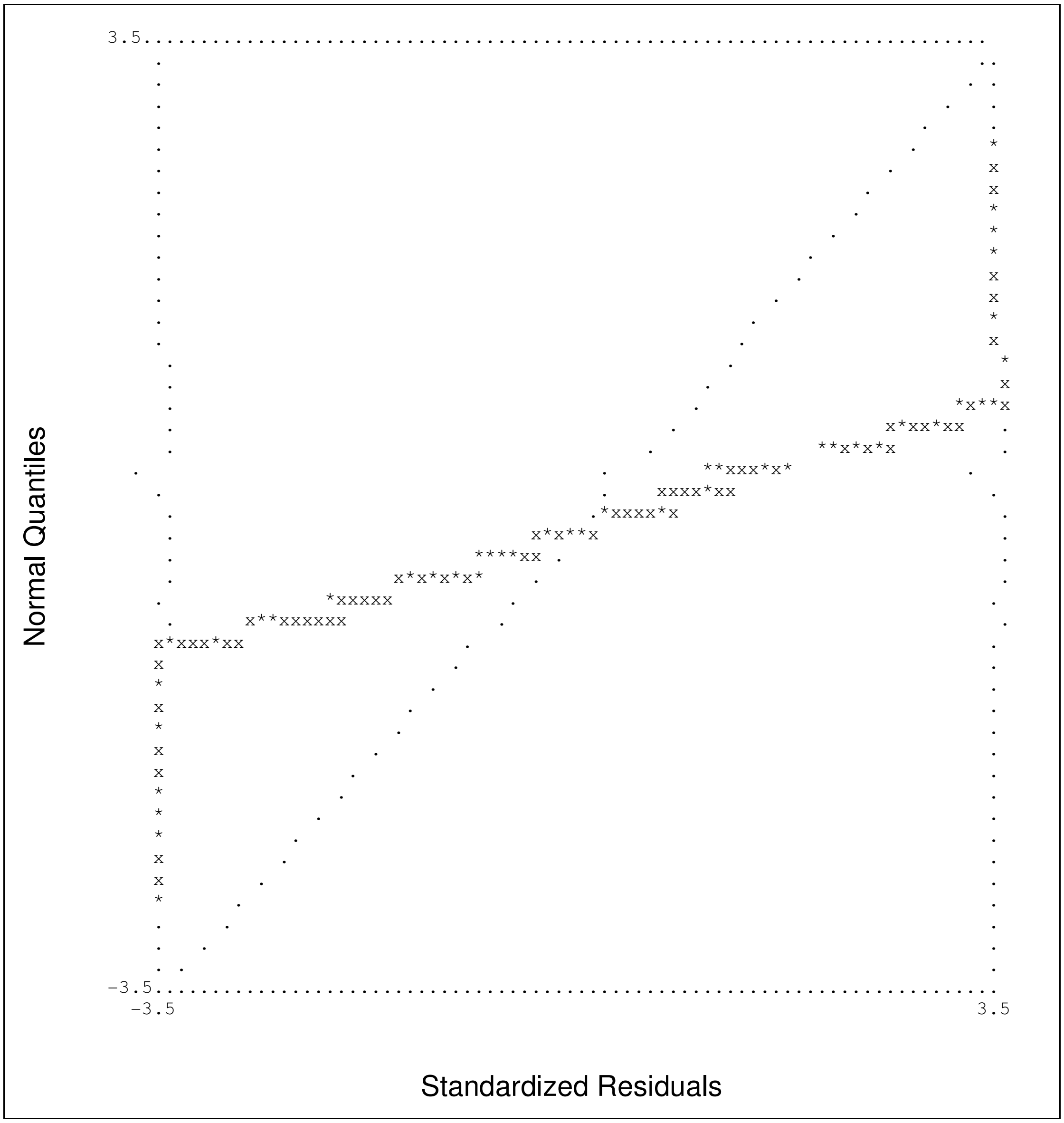

Figure 6.2. Q-Plot of standardized residuals for combined sample: Basic Interest model.

The Q-plot displayed in Figure 6.2 provides an additional graphical display of residuals by plotting the standardized residuals (horizontal axis) against the quantiles of the normal distribution (Diamantopoulos & Siguaw, 2000). The Q-plot also indicates reasonable to mediocre model fit as there are progressively large angular deviations of the standardized residuals for some pairs of observed variables from the 45° reference line in the Q-plot, especially in the upper and lower regions of the X-axis.

6.4.3 Model modification indices

Examining the modification indices returned by LISREL for the currently fixed parameters of the model may also provide an additional way of determining if adding one or more paths would significantly improve the fit of the model. The indices aim to estimate the decrease that would occur in the χ^2 statistic if parameters that are currently fixed are set free and the model is re-estimated. Modification indices with large values (larger than 6.64; Theron, 2006) identify currently fixed parameters that would improve the fit of the model significantly (p<0.01) if set free (Diamantopoulos & Siguaw, 2000). Diamantopoulos and Siguaw (2000) and Kelloway (1998) do, however, suggest that modifications to the model based on these statistics should be theoretically/substantially justified. Paths would not be freed in this study as the purpose is to evaluate the fit of the *a priori* model indicated by the test authors. Modification indices calculated for the Λ_X and Θ_δ matrices will, however, be examined as comments on the adequacy of the fitted measurement model. If only a limited number of additional factor loadings would be proposed and only a limited number of (or no) correlated measurement error terms would be proposed, this would comment positively on the merits of the model derived from the design intentions of the test developers.

Upon inspection of the modification indices for the Λ_x matrix, a number of paths could have been freed that would significantly improve model fit. Approximately 50% of the currently free elements in the Λ_x matrix were identified as being significant (> 6.64). This corroborates the conclusion derived from the Aiken information criterion, the ECVI statistic and the inspection of the stem-and-leaf plot of standardized residuals that the current model's claim that each sub-scale of items only reflect a specific single latent Basic Interest dimension is should be questioned. This would then suggest that greater provision should be made for cross loading items in the *a priori* model. The use of the original items as indicator variables might have resulted in better fit of the *a priori* model but this seems unlikely as the item parcels are unweighted linear composites of the sub-scale items. The converse could, however, also be found when fitting the model with the individual items as indicators.

The magnitude of the predicted factor loadings that should be found if currently fixed elements in Λ_X would be freed is reflected in the LISREL output in the completely standardized expected change values. An investigation of the completely standardized expected change matrix showed only one satisfactory (loadings greater than a stringent value of 0.70) loading. A pathway could be freed between the Writing scale and the second item parcel representing International Activities. This item parcel contains elements regarding language interest. The potential loading would be above the stringent cut-off level of 0.70.

The numerous significant modification index values in Λ_X along with the rather modest completely standardized expected change values suggest that the response to the items allocated to a specific sub-scale does not solely depend on that specific latent Basic Interest dimension but also on numerous other latent Basic Interest dimensions in the *interest* space. The numerous significant modification index values in Λ_X along with the rather modest completely standardized expected change values, therefore, seem to suggest that sub-scale items do not cluster closely around the upper end of the interest dimension they were earmarked to reflect in the (29 dimensional) interest space but rather that they are more scattered around the interest space.

As far as the theta-delta (Θ_δ) modification indices are concerned, approximately 35 percent of the matrix is significant (> 6.64). However, upon review of standardized expected changes as a result of freeing off-diagonal error terms in theta-delta, no correlated measurement error terms can be proposed. This finding comments positively on the fit of the measurement model.

However, as previously indicated, no changes will be made to the model.

6.4.4 Assessment of the estimated model parameters of the Basic Interest scales model

The completely standardized factor loading matrix (Λ_x), reflected in Table 6.41, indicates the regression of X_i on ξ_i and is used to evaluate the significance of the Basic Interest scale item parcel loadings as specified by the *a priori* model. The completely standardized

λ parameter estimates reflect the average change in standard deviation units in the manifest variable X (in this case item parcels) 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. The results provided in Table 6.41 describe factor loadings (significance through *t*-values > I1.96I; Theron, 2006) and associated standard error estimates per item parcel. All specified factor loadings are significant ($p<0.05$). H_{02} to H_{059} can therefore all be rejected in favour of H_{a2} to H_{a59}. By pruning any of these pathways, model fit will deteriorate significantly (Kelloway, 1998). Table 6.41 further indicates the magnitude of the completely standardized factor loading freed in accordance with the model specification. In the case of item parcels reflecting a single latent variable, the completely standardized factor loadings can be interpreted as correlation coefficients. By using a stringent cut-off point of 0.80, most factor loadings appear satisfactory. Six factor loadings were below this cut-off point; ADEVIP2 (0.78), COUNIP2 (0.74), INTIP1 (0.63), PUBIP2 (0.78), RSKIP2 (0.79) and WOODIP2 (0.71). Of most concern is INTIP1 which returned the lowest factor loading. The items comprising this parcel include the foreign language interest content.

The total variance in the *i*th item parcel (X_i) can be decomposed into variance due to variance in the latent variable the item parcel was designed to reflect (ξ_j), variance due to variance in other systematic latent effects the item parcel was not designed to reflect, as well as random measurement error. The latter two sources of variance in the item parcel are acknowledged in the model specification through the measurement error term (δ_i). The measurement error terms (δ) thus does not differentiate between systematic and random sources of error or non-relevant variance. The square of the completely standardized factor loadings λ_{ij} given in Table 6.41 could be interpreted as the proportion systematic-relevant item parcel variance given that each item parcels loads on one latent Basic Interest variable only. Since reliability could be defined as the extent to which variance in item parcels can be attributed to systematic sources, irrespective of whether the source of variance is relevant to the measurement intention or not, the square of the completely standardized factor loading values shown in Table 6.41 could be simultaneously interpreted as lower bound estimates of the item reliabilities (Diamantopoulos & Siguaw, 2000; Jöreskog & Sörbom, 1996a). Thus, given the values calculated for Λ_x shown in Table 6.40, the extent to which the true item reliabilities would be under-estimated is not likely to be considerable and in most cases the item parcels seem to provide relatively uncontaminated reflections of their designated latent dimensions.

The proportion of item parcel variance as per the latent variable it has been designed to reflect as per the measurement model is indicated by the squared multiple correlations for the observed indicator variables as shown in Table 6.42. The squared multiple correlations for the following item parcels were not as satisfactory (>0.70) as for the remainder of the item parcels ADEVIP1 (0.66), ADEVIP2 (0.61), ATHIP1 (0.54), ATHIP2 (0.63), COUNIP2 (0.54), INTIP1 (0.39), PERIP2 (0.67), PUBIP2 (0.61), RSKIP1 (0.65), RSKIP2 (0.63) and WOODIP2 (0.50). These item parcels provide relatively contaminated reflections of their designed dimension.

Table 6.41.
Completely Standardized Factor Loading Matrix for Combined Sample: Basic Interest Scales

Adult Development		Advertising		Animal Care		Art/Design		Athletics	
ADEVIP1	0.81 (0.07) 30.74*	ADVIP1	0.93 (0.11) 42.01*	ANIP1	0.88 (0.07) 39.48*	ARTIP1	0.89 (0.09) 33.97*	ATHIP1	0.74 (0.17) 24.39*
ADEVIP2	0.78 (0.07) 28.75*	ADVIP2	0.95 (0.11) 43.92*	ANIP2	0.92 (0.07) 40.02*	ARTIP2	0.84 (0.09) 31.01*	ATHIP2	0.79 (0.11) 25.07*
Child Development		**Counseling**		**Culinary Arts**		**Farming**		**Fashion**	
CDEVIP1	0.92 (0.11) 43.59*	COUNIP1	0.82 (0.10) 29.40*	CULIP1	0.95 (0.10) 34.55*	FARMIP1	0.92 (0.09) 44.72*	FASHIP1	0.90 (0.10) 37.11*
CDEVIP2	0.90 (0.09) 39.66*	COUNIP2	0.74 (0.12) 25.30*	CULIP2	0.83 (0.11) 30.98*	FARMIP2	0.95 (0.09) 47.93*	FASHIP2	0.83 (0.09) 35.54*
Financial Services		**Int'l Activities**		**Law/Politics**		**Leadership**		**Mathematics**	
FINIP1	0.93 (0.11) 43.70*	INTIP1	0.63 (0.13) 19.82*	LAWIP1	0.86 (0.15) 36.92*	LEADIP1	0.88 (0.12) 36.70*	MATHIP1	0.89 (0.13) 40.83*
FINIP2	0.89 (0.12) 39.60*	INTIP2	0.86 (0.08) 31.18*	LAWIP2	0.91 (0.14) 40.10*	LEADIP2	0.83 (0.09) 31.69*	MATHIP2	0.94 (0.09) 41.06*
Mechanical Crafts		**Medical Practice**		**Military**		**Office Practices**		**Performing Arts**	
MECHIP1	0.95 (0.13) 45.15*	MEDIP1	0.99 (0.14) 47.36*	MILIP1	0.88 (0.15) 36.26*	OPIP1	0.95 (0.13) 40.91*	PERIP1	0.92 (0.12) 36.62*
MECHIP2	0.91 (0.12) 44.14*	MEDIP2	0.90 (0.17) 40.42*	MILIP2	0.88 (0.09) 43.74*	OPIP2	0.84 (0.12) 33.22*	PERIP2	0.82 (0.17) 31.72*

Plants		Public Speaking		Religious Activities		Risk		Sales	
	0.95		0.86		0.87		0.81		0.90
PNTIP1	(0.09)	PUBIP1	(0.07)	RELIP1	(0.10)	RSKIP1	(0.08)	SALIP1	(0.11)
	44.30*		34.11*		34.85*		29.27*		36.20*
	0.88		0.78		1.00		0.79		0.93
PNTIP2	(0.06)	PUBIP2	(0.07)	RELIP2	(0.06)	RSKIP2	(0.08)	SALIP2	(0.11)
	36.54*		29.47*		50.91*		29.00*		39.09*
Science		Supervision		Woodworking		Writing			
	0.85		0.88		0.91		0.84		
SCIIP1	(0.14)	SUPIP1	(0.11)	WOODIP1	(0.10)	WRTIP1	(0.10)		
	36.68*		35.43*		40.56*		33.62*		
	0.86		0.86		0.71		0.89		
SCIIP2	(0.09)	SUPIP2	(0.11)	WOODIP2	(0.09)	WRTIP2	(0.09)		
	32.43*		33.61*		24.72*		34.27*		

* *t*-values > I 1.96 I indicate significant path coefficients; values in brackets represent standard error estimates

Table 6.42.
Squared Multiple Correlations for Item Parcels: Basic Interest Scales

ADEVIP1	0.66	ADVIP1	0.87	ANIP1	0.78	ARTIP1	0.79	ATHIP1	0.54
ADEVIP2	0.61	ADVIP2	0.90	ANIP2	0.84	ARTIP2	0.70	ATHIP2	0.63
CDEVIP1	0.85	COUNIP1	0.67	CULIP1	0.91	FARMIP1	0.84	FASHIP1	0.81
CDEVIP2	0.82	COUNIP2	0.54	CULIP2	0.68	FARMIP2	0.90	FASHIP2	0.69
FINIP1	0.86	INTIP1	0.39	LAWIP1	0.74	LEADIP1	0.77	MATHIP1	0.80
FINIP2	0.79	INTIP2	0.74	LAWIP2	0.83	LEADIP2	0.70	MATHIP2	0.89
MECHIP1	0.91	MEDIP1	0.98	MILIP1	0.77	OPIP1	0.90	PERIP1	0.84
MECHIP2	0.82	MEDIP2	0.81	MILIP2	0.77	OPIP2	0.71	PERIP2	0.67
PNTIP1	0.89	PUBIP1	0.74	RELIP1	0.75	RSKIP1	0.65	SALIP1	0.81
PNTIP2	0.77	PUBIP2	0.61	RELIP2	1.00	RSKIP2	0.63	SALIP2	0.87
SCIIP1	0.72	SUPIP1	0.78	WOODIP1	0.82	WRTIP1	0.70		
SCIIP2	0.73	SUPIP2	0.74	WOODIP2	0.50	WRTIP2	0.79		

The phi-matrix of correlations between the 29 latent Basic Interests sub-scales is provided in Table 6.43. The off-diagonal elements of the Φ matrix are the Basic Interest scale correlations disattenuated for measurement error. As the Φ matrix is positive definite and off-diagonal entries tend to contain relatively moderate correlations, the results tend to provide some support for discriminant validity of the Basic Interest scales.

The correlations observed seem to agree with the intended clustering as indicated in Table 3.1. When investigating how the Basic Interest scales that cluster under a specific Orientation Interest scale tend to correlate amongst themselves, a correlation above 0.30

was returned for the majority of cases. An exception to this includes International Activities and Fashion (0.25), and Writing and Culinary Arts (0.29) which are intended to represent the Creating Orientation scales.

6.4.5 Summary of model fit assessment for the combined sample

Overall, the model statistics do indicate that reasonable to mediocre fit has been found for the combined sample. That said the model fails because it only modestly captures the complexity of the dynamics underlying the CISS. The results show that the model would benefit from adding additional pathways. This was reflected in the fit statistics, echoed by the distributional form of the standardized residuals. Modification indices calculated for the factor loading matrix also indicate a sizable number of paths that could be added to effectively decrease the chi-square statistic although these cross loadings are generally not expected to be very high.

The completely standardized factor loading matrix and squared multiple correlations indicate that the item parcels generally reflect the latent Basic Interest dimensions they were tasked to represent reasonably well, although a limited number of parcels only loaded rather modestly on the latent variables they were meant to represent. These findings explored holistically seems to suggest that the behavioural responses to the items allocated to a specific interest sub-scale, although primarily determined by the latent Basic Interest dimension they were tasked to reflect, nonetheless depend on the whole of the interest space. As the measurement model does not reflect this, the model fit tends to be adversely affected.

It could be said that the CISS Basic Interest scales demonstrate fair levels of construct validity.

Table 6.43.
Completely Standardized Phi Matrix for the Combined Sample: Basic Interest Scales

	Adult Dev	Advertising	Animal Care	Art/Design	Athletics	Child Dev	Counseling	Culinary Arts	Farming	Fashion	Financial Services	Int'lActivities	Law/Politics	Leadership	Mathematics	Mechanical Crafts	Medical Prac	Military	Office Prac	Perf Arts	Plants	Public Speaking	Religious Activities	Risk	Sales	Science	Supervision	Woodworking	Writing
Adult Dev	1.00																												
Advertising	0.38	1.00																											
Animal Care	0.30	-0.07	1.00																										
Art/Design	0.26	0.51	0.07	1.00																									
Athletics	0.43	0.43	0.19	0.36	1.00																								
Child Dev	0.71	0.12	0.39	0.19	0.30	1.00																							
Counseling	0.84	0.51	0.06	0.35	0.41	0.60	1.00																						
Culinary Arts	0.31	0.33	0.24	0.41	0.37	0.39	0.34	1.00																					
Farming	0.33	-0.02	0.70	0.08	0.16	0.31	0.00	0.22	1.00																				
Fashion	0.39	0.31	0.24	0.45	0.17	0.58	0.40	0.43	0.17	1.00																			
Fin Services	0.41	0.46	0.25	0.09	0.26	0.20	0.26	0.17	0.46	0.19	1.00																		
Int'l Activities	0.37	0.52	-0.02	0.45	0.49	0.19	0.55	0.40	-0.07	0.25	0.04	1.00																	
Law/Politics	0.65	0.41	0.31	0.13	0.24	0.34	0.51	0.18	0.42	0.30	0.74	0.22	1.00																
Leadership	0.65	0.65	0.01	0.30	0.50	0.22	0.68	0.31	0.11	0.16	0.64	0.52	0.75	1.00															
Mathematics	0.40	0.13	0.04	0.16	0.31	0.08	0.18	0.02	0.16	-0.07	0.44	0.07	0.32	0.39	1.00														
Mech Crafts	0.19	0.00	0.33	0.13	0.26	0.05	-0.12	0.05	0.58	-0.10	0.44	-0.18	0.33	0.18	0.59	1.00													
Medical Prac	0.59	0.11	0.33	0.13	0.48	0.56	0.59	0.22	0.25	0.32	0.19	0.28	0.33	0.30	0.38	0.22	1.00												
Military	0.44	0.13	0.57	0.05	0.39	0.37	0.24	0.21	0.63	0.19	0.55	0.02	0.72	0.37	0.27	0.58	0.43	1.00											
Office Prac	0.48	0.19	0.39	0.06	0.17	0.46	0.28	0.25	0.46	0.41	0.71	-0.01	0.54	0.43	0.31	0.37	0.30	0.52	1.00										
Perf Arts	0.52	0.35	0.34	0.35	0.34	0.43	0.37	0.31	0.32	0.54	0.35	0.31	0.56	0.36	0.13	0.16	0.31	0.44	0.33	1.00									
Plants	0.36	0.13	0.58	0.41	0.26	0.31	0.15	0.42	0.73	0.29	0.27	0.15	0.28	0.15	0.16	0.39	0.28	0.35	0.30	0.33	1.00								
Public Speaking	0.69	0.57	0.16	0.28	0.36	0.35	0.63	0.29	0.21	0.34	0.48	0.43	0.74	0.73	0.22	0.11	0.30	0.39	0.38	0.67	0.26	1.00							
Rel Activities	0.49	0.08	0.34	-0.01	0.19	0.40	0.32	0.14	0.44	0.28	0.38	-0.04	0.42	0.23	0.19	0.26	0.33	0.42	0.43	0.37	0.32	0.39	1.00						
Risk	0.11	0.28	0.29	0.33	0.54	0.04	0.12	0.31	0.26	-0.03	0.16	0.51	0.25	0.34	0.14	0.37	0.18	0.45	-0.03	0.25	0.25	0.23	-0.05	1.00					
Sales	0.38	0.51	0.39	0.22	0.30	0.32	0.28	0.35	0.45	0.47	0.71	0.17	0.57	0.48	0.16	0.33	0.22	0.52	0.63	0.46	0.36	0.50	0.38	0.21	1.00				
Science	0.45	0.14	0.11	0.17	0.31	0.08	0.21	0.04	0.24	-0.06	0.37	0.12	0.35	0.40	0.87	0.61	0.52	0.33	0.23	0.18	0.27	0.21	0.20	0.21	0.13	1.00			
Supervision	0.64	0.48	0.32	0.18	0.40	0.41	0.48	0.40	0.45	0.33	0.76	0.23	0.70	0.77	0.34	0.44	0.33	0.65	0.81	0.39	0.32	0.57	0.43	0.23	0.76	0.34	1.00		
Woodworking	0.25	0.07	0.46	0.36	0.24	0.23	0.00	0.32	0.73	0.18	0.38	-0.04	0.30	0.13	0.26	0.76	0.16	0.53	0.39	0.25	0.66	0.17	0.33	0.36	0.41	0.26	0.41	1.00	
Writing	0.67	0.42	0.25	0.41	0.22	0.38	0.52	0.29	0.33	0.35	0.41	0.33	0.64	0.49	0.28	0.18	0.33	0.39	0.37	0.59	0.40	0.68	0.37	0.18	0.39	0.37	0.42	0.28	1.00

6.5 RESULTS OF ORIENTATION INTEREST SCALES CFA: COMBINED SAMPLE

Due to the ratio of sample size to number of parameters to be estimated, independent sample CFAs could not be conducted on the Basic Interest model. However, it was possible to conduct independent sample CFA analyses at the global factor level, namely: the Orientation Interest model. Hence, the Orientation Interest model was subjected to combined sample data prior to evaluating model fit for the separate gender samples (a necessary prerequisite to the measurement invariance tests).

The analysis reported on in this section specifically aims to evaluate (operational hypothesis 2) whether the Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the combined sample. Operational hypothesis 2 is tested by testing H_{060}: RMSEA =0.

The detailed output of the confirmatory factor analyses is electronically available (on the included CD, folder: CFA, ORIENTATION INTEREST SCALES) in Appendix A.

6.5.1 Overall fit assessment

The Orientation Interest measurement model was confronted with the item parcel data of the combined sample and a final solution was found after 61 iterations. Results are presented in Table 6.44.

The Satorra-Bentler χ^2 test statistic (26916.05) is significant (p<0.01), thus resulting in a rejection of the null hypothesis of exact model fit [H_{060}: $\Sigma = \Sigma(\theta)$]. The normed χ^2 (17.1) indicates that the measurement model is severely ‘under-fitted’ and may benefit from major improvement. The estimated λ value for the combined sample (25342.05) is very high. The 90 percent confidence interval for NCP is (24812.89; 25877.47). This strongly suggest that the estimated discrepancy between the observed (Σ) and estimated population covariance [$\Sigma(\theta)$] matrices is unacceptably high (Diamantopoulos & Siguaw, 2000). The RMSEA value

(0.14) for the combined sample by far exceed the criteria of close model fit (i.e., <0.05) (Browne & Cudeck, 1993). The 90% confidence interval for the RMSEA sample estimate returns a rather unusual "interval" (0.14; 0.14) which seems to be indicating that a very precise, albeit large, estimate had been obtained. The test of close fit performed by LISREL shows that the conditional probability of the obtained RMSEA value of 0.14 under H_{060}: RMSEA ≤ 0.05 is sufficiently low to ensure a rejection of the close fit null hypothesis. These results suggest that poor fit has been achieved for the Orientation Interest scales model.

The model ECVI (33.61) is far smaller than the value obtained for the independence model (161.77) but larger than the ECVI value associated with the saturated model (4.23), and thus a model more closely resembling the saturated model seems to have a better chance of being replicated in a cross-validation sample than the fitted model. This once again strengthens the argument for poor fit (Diamantopoulos & Siguaw, 2000). The parsimonious normed fit index (PNFI = 0.76) and the parsimonious goodness-of-fit index (PGFI = 0.41) values fall well short of the 0.90 cut-off value.

The values for this model's AIC (27190.05) suggests that the fitted measurement model provides a more parsimonious fit than the independent/null model (130875.95) but not than the saturated model (3422.00) since smaller values on these indices indicate a more parsimonious model (Kelloway, 1998). Values for the CAIC (27970.55) imply that the fitted measurement model provides a more parsimonious fit than the independent/null model (131206.38) but once again not than the saturated model (13169.63). This, in conjunction with the ECVI results, indicates that the measurement model does lack influential paths.

The indices of relative fit given in Table 6.44 all fall short of the critical value of 0.90 and therefore indicate inadequate comparative fit when compared to the independence model (Diamantopoulos & Siguaw, 2000; Kelloway, 1998). These indices include the NFI (0.79), NNFI (0.79), CFI (0.80), IFI (0.80), and the RFI (0.78). Additionally the estimated CN value (52.32) is well below the recommended threshold value of 200. In a similar vein, the RMR (0.14) and standardized RMR (0.077) suggest marginal fit. The AGFI (0.39) and the GFI (0.44) both fall well below the accepted cut-off of 0.90 which indicates the model does not come even close to perfectly reproducing the sample covariance matrix and, therefore, further suggests poor fit (Jöreskog & Sörbom, 1993; Kelloway, 1998).

Overall, the model fit statistics would point towards poor fit. The model fails to provide a credible explanation for the observed covariances between the item parcels. The model does outperform the independence model but underperforms relative to the saturated model which would imply that the model complexity could be captured better with many additional paths.

Table 6.44.
Goodness-Of-Fit Indicators for Combined Sample: Orientation Interest Model

Degrees of Freedom = 1574
Minimum Fit Function Chi-Square = 21420.64 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 29772.43 (P = 0.0)
Satorra-Bentler scaled Chi-Square = 26916.05 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 25342.05
90 Percent Confidence Interval for NCP = (24812.89 ; 25877.47)

Minimum Fit Function Value = 26.48
Population Discrepancy Function Value (F0) = 31.33
90 Percent Confidence Interval for F0 = (30.67 ; 31.99)
Root Mean Square Error of Approximation (RMSEA) = 0.14
90 Percent Confidence Interval for RMSEA = (0.14 ; 0.14)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00

Expected Cross-Validation Index (ECVI) = 33.61
90 Percent Confidence Interval for ECVI = (32.96 ; 34.27)
ECVI for Saturated Model = 4.23
ECVI for Independence Model = 161.77

Chi-Square for Independence Model with 1653 Degrees of Freedom = 130759.95
Independence AIC = 130875.95
Model AIC = 27190.05
Saturated AIC = 3422.00
Independence CAIC = 131206.38
Model CAIC = 27970.55

Saturated CAIC = 13169.63

Normed Fit Index (NFI) = 0.79

Non-Normed Fit Index (NNFI) = 0.79

Parsimony Normed Fit Index (PNFI) = 0.76

Comparative Fit Index (CFI) = 0.80

Incremental Fit Index (IFI) = 0.80

Relative Fit Index (RFI) = 0.78

Critical N (CN) = 52.32

Root Mean Square Residual (RMR) = 2.45

Standardized RMR = 0.12

Goodness of Fit Index (GFI) = 0.44

Adjusted Goodness of Fit Index (AGFI) = 0.39

Parsimony Goodness of Fit Index (PGFI) = 0.41

6.5.2 Examination of residuals

The stem-and-leaf plot (Figure 6.3) does show a distribution centred around the median of zero.

```
- 5|6
 - 4|
 - 3|4
 - 2|211
 - 1|666544433333322222222222111111111110000000
 - 0|99999999999999999999888888888888888888888877777777777777777777777777777777+94
   0|111111111111111111111111111111111111111111111111111111111111111111111111+99
   1|0000001111111122222223444444566778
   2|012267
   3|333
   4|
   5|09
   6|
   7|5
   8|
   9|
  10|6
```

Figure 6.3. Stem-and-leaf plot of standardized residuals for the combined sample: Orientation Interest model.

The distribution is, however, positively skewed with numerous large positive standardized residuals (largest 10.60) and a smaller number of large negative (largest -5.60) standardized residuals. The large positive residuals would suggest that many paths would have to be added to improve model fit. This corroborates the finding derived from the fit statistics that this model fits poorly.

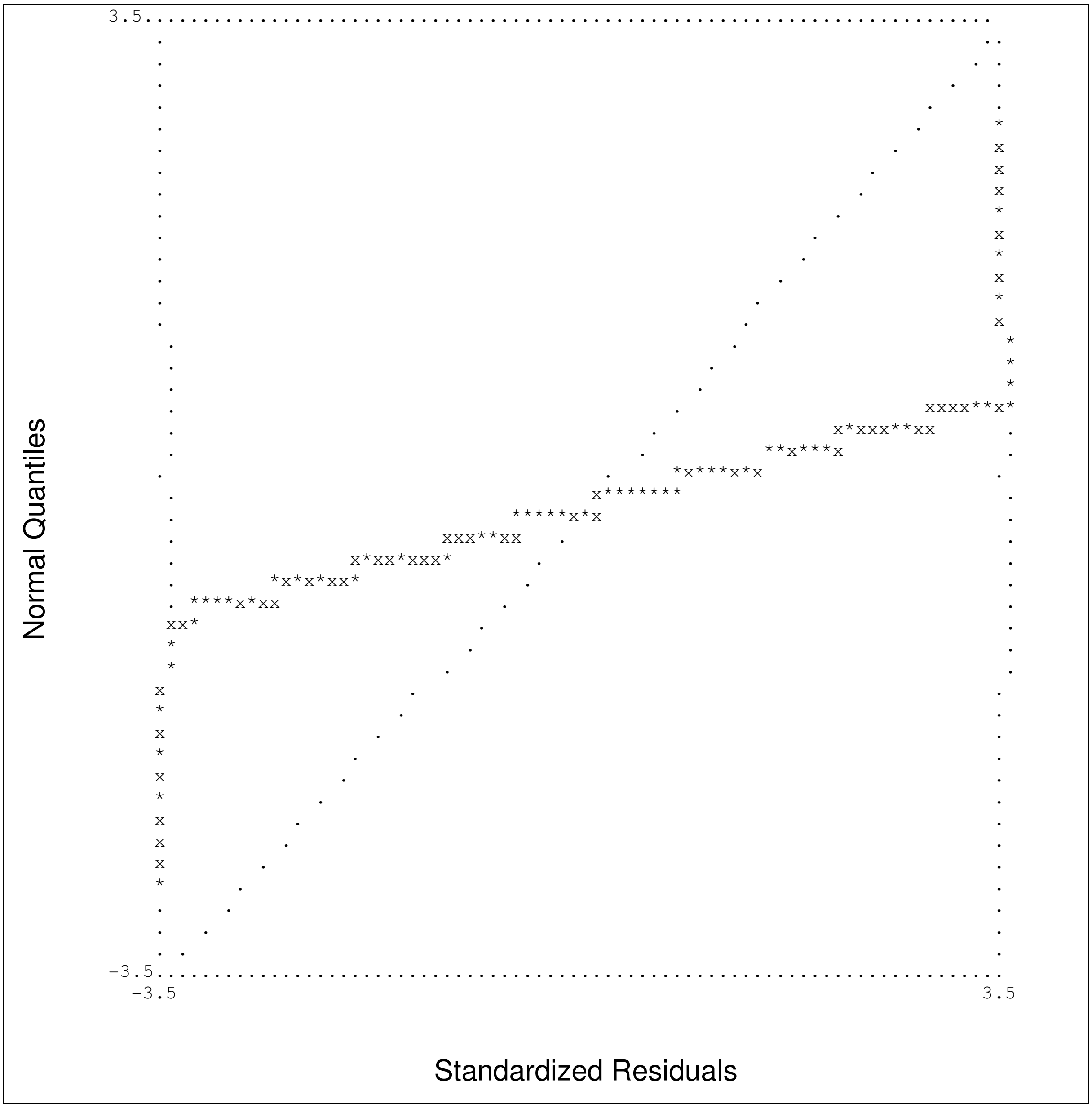

Figure 6.4. Q-Plot of standardized residuals for combined sample: Orientation Interest model.

In the Q-plot (Figure 6.4) for this model, few standardized residuals are placed on the reference line. The plotted residuals show a large angular deviation from the reference line, therefore, suggesting poor fit.

6.5.3 Model modification indices

As with the Basic scales model, inspection of the Λ_x modification index matrix revealed a number of paths that, when freed, would significantly improve model fit. Approximately 52% of the currently fixed elements in the Λ_x matrix, if freed, would be significant, which suggests that the item parcels calculated for a specific Basic scale do not only reflect the higher-order Orientation scale on which that Basic Interest loads, but also other Orientations. This large number of significant modification index value in Λ_X again points to problems with the current *a priori* model and corroborates the earlier indicators of poor model fit.

When examining the completely standardized expected change matrix no pathways seemed to show a satisfactory loading (> 0.70). A similar conclusion therefore seems to be indicated as in the case of the Basic Interest model. As the same item parcels are used to represent the latent Orientation variables that were used to represent the latent Basic Interest dimensions - this is not altogether unexpected. The numerous significant modification index values in Λ_X along with the rather modest completely standardized expected change values suggest that the response to the items allocated to a specific sub-scale does not solely depend on that specific latent Orientation Interest dimension but also on a number of other latent Orientation Interest dimensions in the interest space. The numerous significant modification index values in Λ_X along with the rather modest completely standardized expected change values therefore seem to suggest that sub-scale items do not cluster closely around the upper end of the Orientation dimension they were earmarked to reflect in the (seven dimensional) interest space. They rather are more scattered around the interest space.

Theta-delta modification indices for the combined sample also suggest that error term parameters could be freed to improve model fit. However, the standardized expected change values do not support making this decision. The absence of strong correlations between the measurement error terms comments positively on the fitted model.

6.5.4 Assessment of the parameter estimates of the Orientation Interest scales model

Due to the poor model fit it does not seem suitable to interpret the Orientation Interest measurement model parameter estimates. The poor model fit means that the model with its current parameter estimates cannot, to any acceptable degree of accuracy, reproduce the observed covariance matrix. Therefore, the parameter estimates are not credible. The same can be said for discussing the factor loading estimates, measurement error variance estimates or inter-latent variables correlations.

6.5.5 Summary of model fit assessment for the combined sample

The model fit results, obtained here, were poor, in relation to the reasonable-mediocre fit that was observed for the Basic Interest model (see section 6.4). Overall, for this model, the fit statistics indicate inadequate fit. This would suggest that the model fails to adequately account for the covariance observed between the item parcels.

Even though these results seem to raise some questions regarding the global level factor structure of the Interest model, gender sample CFAs should still be conducted to determine if the poor fit in the combined sample could be due to poor fit in either of the independent samples. However, should poor fit be found for these samples, the tests for measurement invariance would not be justified.

6.6 RESULTS OF ORIENTATION INTEREST SCALES CFA: MALE SAMPLE

The analysis reported on in this section and the subsequent section (paragraph 6.7) specifically aims to evaluate (operational hypothesis 3) whether the Orientation scale interest measurement model implied by the scoring key of the CISS can closely reproduce the covariances observed between the item parcels formed from the items comprising each of the Basic scales in the separate gender samples. Operational hypothesis 3 is tested by testing H_{0119M}: RMSEA = 0.

The detailed output of the confirmatory factor analyses is electronically available (on the included CD, folder: CFA, Basic Interest Scales) in Appendix A.

6.6.1 Overall fit assessment

Upon fitting the data of the male sample (*n*=408) to the Orientation Interest model the following fit statistics were obtained after 88 iterations.

Table 6.45.
Goodness-Of-Fit Indicators for Male Sample: Orientation Interest Model

Degrees of Freedom = 1574

Minimum Fit Function Chi-Square = 22973.38 (P = 0.0)

Normal Theory Weighted Least Squares Chi-Square = 36700.92 (P = 0.0)

Satorra-Bentler scaled Chi-Square = 33005.69 (P = 0.0)

Estimated Non-centrality Parameter (NCP) = 31431.69

90 Percent Confidence Interval for NCP = (30843.51 ; 32025.82)

Minimum Fit Function Value = 28.40

Population Discrepancy Function Value (F0) = 38.85

90 Percent Confidence Interval for F0 = (38.13 ; 39.59)

Root Mean Square Error of Approximation (RMSEA) = 0.16

90 Percent Confidence Interval for RMSEA = (0.16 ; 0.16)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00

Expected Cross-Validation Index (ECVI) = 41.14

90 Percent Confidence Interval for ECVI = (40.41 ; 41.87)

ECVI for Saturated Model = 4.23

ECVI for Independence Model = 203.20

Chi-Square for Independence Model with 1653 Degrees of Freedom = 164275.13

Independence AIC = 164391.13

Model AIC = 33279.69

Saturated AIC = 3422.00

Independence CAIC = 164721.55

Model CAIC = 34060.19
Saturated CAIC = 13169.63

Normed Fit Index (NFI) = 0.80
Non-Normed Fit Index (NNFI) = 0.80
Parsimony Normed Fit Index (PNFI) = 0.76
Comparative Fit Index (CFI) = 0.81
Incremental Fit Index (IFI) = 0.81
Relative Fit Index (RFI) = 0.79

Critical N (CN) = 42.85

Root Mean Square Residual (RMR) = 2.48
Standardized RMR = 0.13
Goodness of Fit Index (GFI) = 0.39
Adjusted Goodness of Fit Index (AGFI) = 0.34
Parsimony Goodness of Fit Index (PGFI) = 0.36

The null hypothesis of exact model fit [H_{0119M}: $\Sigma = \Sigma(\theta)$] is rejected due to the Satorra-Bentler $\chi 2$ test statistic (33005.69) being significant (p<0.01). The normed $\chi 2$ (20.96) indicates that the measurement model is severely 'under-fitted'. The estimated λ value for the combined sample (31431.69) is very high. The 90 percent confidence interval for NCP is (30843.51; 32025.82). This strongly suggests that the estimated discrepancy between the observed (Σ) and estimated population covariance [$\Sigma(\theta)$] matrices is very high (Diamantopoulos & Siguaw, 2000). The RMSEA value (0.16) for the male sample does not meet the criteria of close model fit (Browne & Cudeck, 1993). Once again the 90% confidence interval for the RMSEA sample estimate returns a rather unusual "interval" (0.16; 0.16) which seems to be indicating that a very precise, albeit large, estimate had been obtained. The null hypothesis of close fit (H_{0119M}: RMSEA $\leq$ 0.05) was also rejected. These results indicate that poor fit has been achieved for the Orientation Interest scales model for the male sample.

Interpretation of the model ECVI [4.23 (saturated model) < 41.14 < 203.20 (independence model)] once again suggests that a model, more closely resembling the saturated model, seems to have a better chance of being replicated in a cross-validation sample than the current fitted model. This further substantiates the finding of lack of model fit (Diamantopoulos & Siguaw, 2000).

The values for this model's AIC (33279.69) suggests that the fitted measurement model provides a more parsimonious fit than the independent/null model (164391.13) but not the saturated model (3422.00) since smaller values on these indices indicate a more parsimonious model (Kelloway, 1998). Values for the CAIC (34060.19) imply that the fitted measurement model provides a more parsimonious fit than the independent/null model (164721.55) but once again not with the saturated model (13169.63).

As with the combined sample, the comparative fit indices did not meet the 0.90 cut-off value. These comparative fit indices are as follows: NFI (0.80), NNFI (0.80), CFI (0.81), IFI (0.81), and the RFI (0.79). The model also provide a limited representation of the data, this is indicated by the CN value [200 > 42.85 (CN)] (Diamantopoulos & Siguaw, 2000). Further to this, the RMR (0.14) and standardized RMR (0.077) suggest marginal fit. Finally, both the AGFI (0.34) and the GFI (0.39) are well below the cut-off value (0.90). In conclusion, the model is not able to reproduce the sample covariance matrix accurately (Jöreskog & Sörbom, 1993; Kelloway, 1998).

6.6.2 Examination of residuals

The stem-and-leaf plot (Figure 6.5) shows a positively skewed distribution centred around a median of zero. Somewhat surprisingly, only a limited number of large positive (largest 7.90) and large negative (largest -4.40) standardized residuals were observed for this model. This once again supports the poor model fit findings observed through the fit statistics discussion.
The Q-plot (Figure 6.6) for this model, also suggest poor fit. The plotted residuals show a large angular deviation from the reference line thereby indicating poor fit.

```
- 4|4
 - 3|
 - 2|210000
 - 1|99777666555554433333322222222222211111111111111110000000 0
 - 0|99999999999999999999998888888888888888888888888888777777777777777777777 7+99
   0|111111111111111111111111111111111111111111111111111111111111111111111111+91
   1|000000000000111111111111111112222222222223333334444455555566678 99
   2|0335
   3|56
   4|4
   5|
   6|
   7|9
```

Figure 6.5. Stem-and-leaf plot of standardized residuals for the male sample: Orientation Interest model.

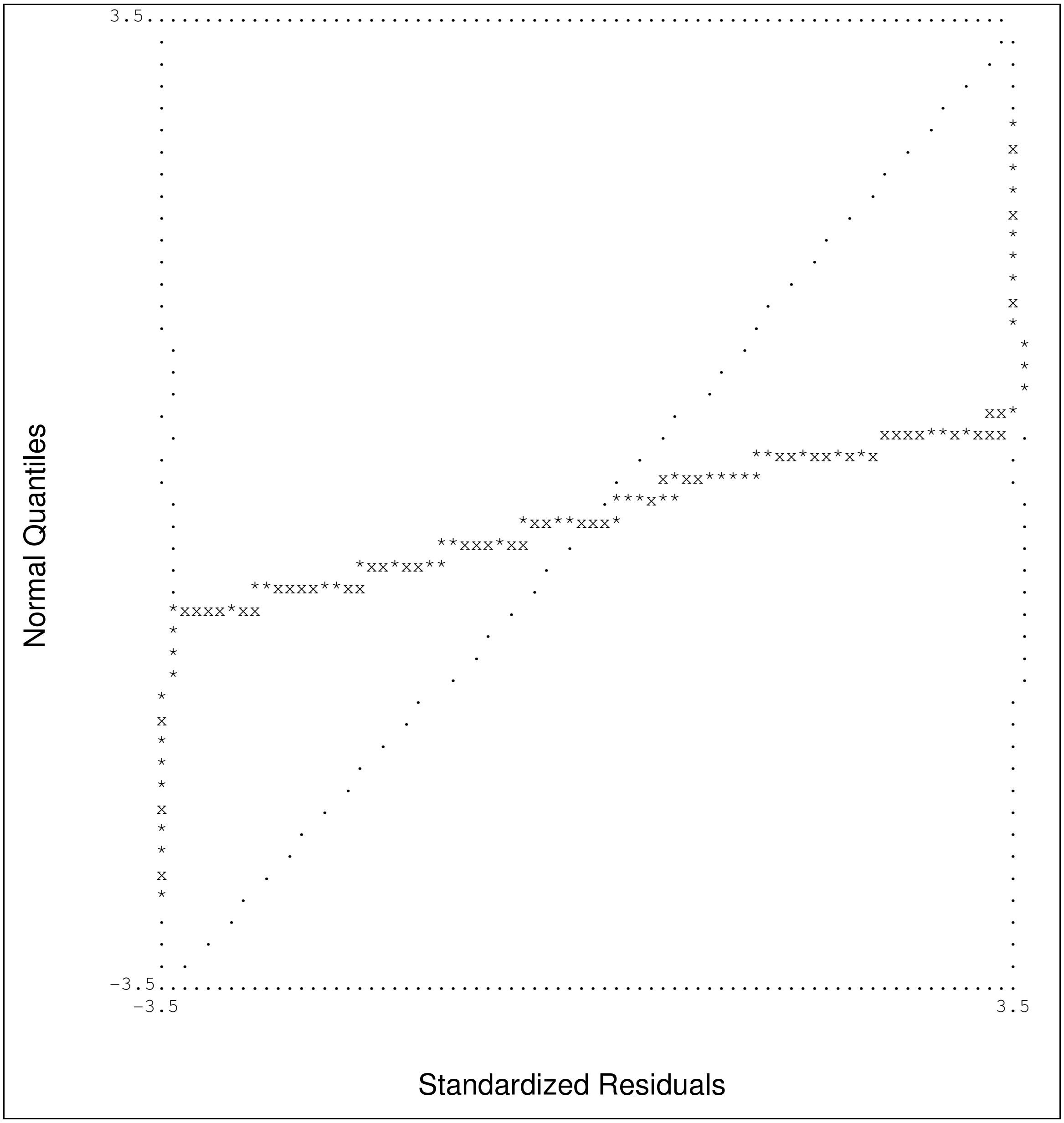

Figure 6.6. Q-Plot of standardized residuals for male sample: Orientation Interest model.

6.6.3 Model modification indices

Approximately 60% of the modification index values calculated for the Λ_x matrix are significant thus indicating that the current model needs to be aggressively revised to obtain better fit.

The completely standardized expected change matrix did not, however, provide suitable loadings should a change be affected. A similar conclusion, therefore, seems to be indicated as in the case of the Orientation model fitted to the combined sample.

As was seen for the combined, a fair number of error term parameters (theta-delta modification indices) could be freed to improve model fit, but the predicted error term correlations do not support this decision.

6.6.4 Assessment of the model parameter estimates for the Orientation Interest scales model

Due to the poor model fit it would not seem plausible to interpret the Orientation Interest Measurement model parameter estimates. The poor model fit indicates that the model with its current parameter estimates cannot, to any acceptable degree of accuracy, reproduce the observed covariance matrix. The parameter estimates are, therefore, not credible. It would not seem valuable to discuss factor loading estimates, measurement error variance estimates or inter-latent variable correlations.

6.6.5 Summary of model fit assessment for the male sample

The purpose of this section was to evaluate the Orientation Interest model fit for the male sample. As seen with the combined sample, the model fits poorly. The Orientation scale measurement model fails to adequately explain the covariance observed between the item parcels.

For the male sample, the model is not able (to with any degree of accuracy) describe the nature of the complexity underlying the CISS on the second-order factor level. The model needs major modification. How this could be applied is not clear. A second-order measurement model would provide a more accurate description of the measurement intention of the test developers. The purpose of the Orientation scales is to summarize the data obtained at the Basic scale level. The Basic Interest scale items were not designed to measure the latent Orientation dimensions directly. A model in which the Basic scale item parcels serve as indicators of the latent Basic Interest dimensions and the latter in turn are mapped onto the latent Orientation Interest dimensions they reflect would provide a more accurate description of the measurement intention of the test developers and model the dynamics underlying the CISS more accurately. Unfortunately sample sizes were not conducive to conduct this more appropriate approach to the model specification.

As a result of the poor model fit, interpreting any of the measurement model parameters would not be meaningful.

The CFA results for the female sample are reported next. However, as model fit is reported as problematic for both the combined and male samples, it is highly probable that the female sample's model fit would be congruent with the existing findings.

6.7 RESULTS OF ORIENTATION INTEREST SCALES CFA: FEMALE SAMPLE

6.7.1 Overall fit assessment

The female sample (n=402) was also subject to confirmatory factor analysis by fitting the data to the Orientation Interest model, as done with the combined and male sample. The model converged after 72 iterations and the results are presented below and summarised in Table 6.46.

The overall fit statistics found for the female sample echo those found for the combined and male sample. The null hypothesis [H_{0119F}: $\Sigma = \Sigma(\theta)$] of exact fit was rejected when examining the Satorra-Bentler $\chi2$ test statistic [(14203.38,(p<0.01)] The model is indicated as being "under-fitted' and would require extensive adjustment, this is due to the value obtained for the normed $\chi2$ (9.02). This indicates that the measurement model is and may benefit from major improvement. The estimated λ value for the female sample is 12629.38 [NCP confidence interval (12251.52 ; 13014.14)]. Once again, this strongly suggests that the estimated discrepancy between the observed (Σθ) and estimated population covariance (Σ(θ)) matrices is high (Diamantopoulos & Siguaw, 2000). Close fit hypothesis is not found for the female sample [RMSEA= 0.14, “confidence interval” (0.14; 0.14)] (Browne & Cudeck, 1993). Similarly, the test of close fit shows that the conditional probability of the obtained RMSEA value of 0.14 under H_{0119F}: RMSEA ≤0.05 is much lower ensuring a rejection of the close fit null hypothesis.

As indicated above, poor fit is once again, found for the female sample (Orientation Interest scales).

Table 6.46.
Goodness-Of-Fit Indicators for Female Sample: Orientation Interest Model

Degrees of Freedom = 1574
Minimum Fit Function Chi-Square = 12004.52 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 15397.60 (P = 0.0)
Satorra-Bentler scaled Chi-Square = 14203.38 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 12629.38
90 Percent Confidence Interval for NCP = (12251.52 ; 13014.14)

Minimum Fit Function Value = 29.94
Population Discrepancy Function Value (F0) = 31.49
90 Percent Confidence Interval for F0 = (30.55 ; 32.45)
Root Mean Square Error of Approximation (RMSEA) = 0.14
90 Percent Confidence Interval for RMSEA = (0.14 ; 0.14)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00

Expected Cross-Validation Index (ECVI) = 36.10
90 Percent Confidence Interval for ECVI = (35.16 ; 37.06)
ECVI for Saturated Model = 8.53
ECVI for Independence Model = 147.89

Chi-Square for Independence Model with 1653 Degrees of Freedom = 59188.56
Independence AIC = 59304.56
Model AIC = 14477.38
Saturated AIC = 3422.00
Independence CAIC = 59594.35

Model CAIC = 15161.90

Saturated CAIC = 11970.93

Normed Fit Index (NFI) = 0.76

Non-Normed Fit Index (NNFI) = 0.77

Parsimony Normed Fit Index (PNFI) = 0.72

Comparative Fit Index (CFI) = 0.78

Incremental Fit Index (IFI) = 0.78

Relative Fit Index (RFI) = 0.75

Critical N (CN) = 49.21

Root Mean Square Residual (RMR) = 3.01

Standardized RMR = 0.14

Goodness of Fit Index (GFI) = 0.43

Adjusted Goodness of Fit Index (AGFI) = 0.38

Parsimony Goodness of Fit Index (PGFI) = 0.40

The ECVI (8.53<36.10 < 147.89) statistics further substantiate the poor model fit finding. A model more closely resembling the saturated model is likely to replicate effectively in a cross-validation sample (Diamantopoulos & Siguaw, 2000).

The values for this model's AIC (14477.38) suggest that the fitted measurement model provides a more parsimonious fit than the independent/null model (59304.56) but not the saturated model (3422.00) (Kelloway, 1998). Similar results are seen for the CAIC statistics [11970.93 (saturated model) < 15161.90 < 59594.35 (independence model)]. In both of these fit indices the overall conclusion would be that the model could benefit from several impacting paths.

The relative fit indices support the conclusions seen in the foregoing discussions. All indices do not meet the 0.90 cut-off value. In fact, they are well below this benchmark (Diamantopoulos & Siguaw, 2000; Kelloway, 1998). Results are: NFI (0.76), NNFI (0.77), CFI (0.78), IFI (0.78), and the RFI (0.75). Additionally the estimated CN value (49.21) is well below the recommended threshold value of 200. In a similar vein, the RMR (3.01) and standardized RMR (0.14) indicate marginal fit. The AGFI (0.38) and the GFI (0.40) also fall below 0.90 and further substantiate a poor model fit observation (Jöreskog & Sörbom, 1993; Kelloway, 1998).

As seen with the combined and male sample, the female sample results indicate poor fit.

6.7.2 Examination of residuals

For the stem-and-leaf plot (Figure 6.7) shows a positively skewed distribution centred around the a median of zero. However, limited large positive (3.70) and negative (-2.70) standardized residuals. As seen with the combined and female analyses, these results suggest echo poor model fit finding observed throughout the fit statistics information.

The plotted residuals once again indicate a large angular deviation from the 45° reference line in the Q-plot (Figure 6.8). Again substantiating the low levels of fit described in the foregoing discussions.

```
- 2|7
 - 2|
 - 1|75
 - 1|4322111111000
 - 0|9999988888888888888888888777777777777777777777777777777777776666666666666+75
 - 0|4444444444444444444444444444444444444444444444444444444444444444444444444+91
   0|1111111111111111111111111111111111111111111111111111111111111111111111111+95
   0|5555555555555555555555555555555555555555555555555555555555555555555555555+05
   1|00000011111223444
   1|5556789
   2|023
   2|588
   3|
   3|7
```

Figure 6.7. Stem-and-leaf plot of standardized residuals for the combined sample: Orientation Interest model.

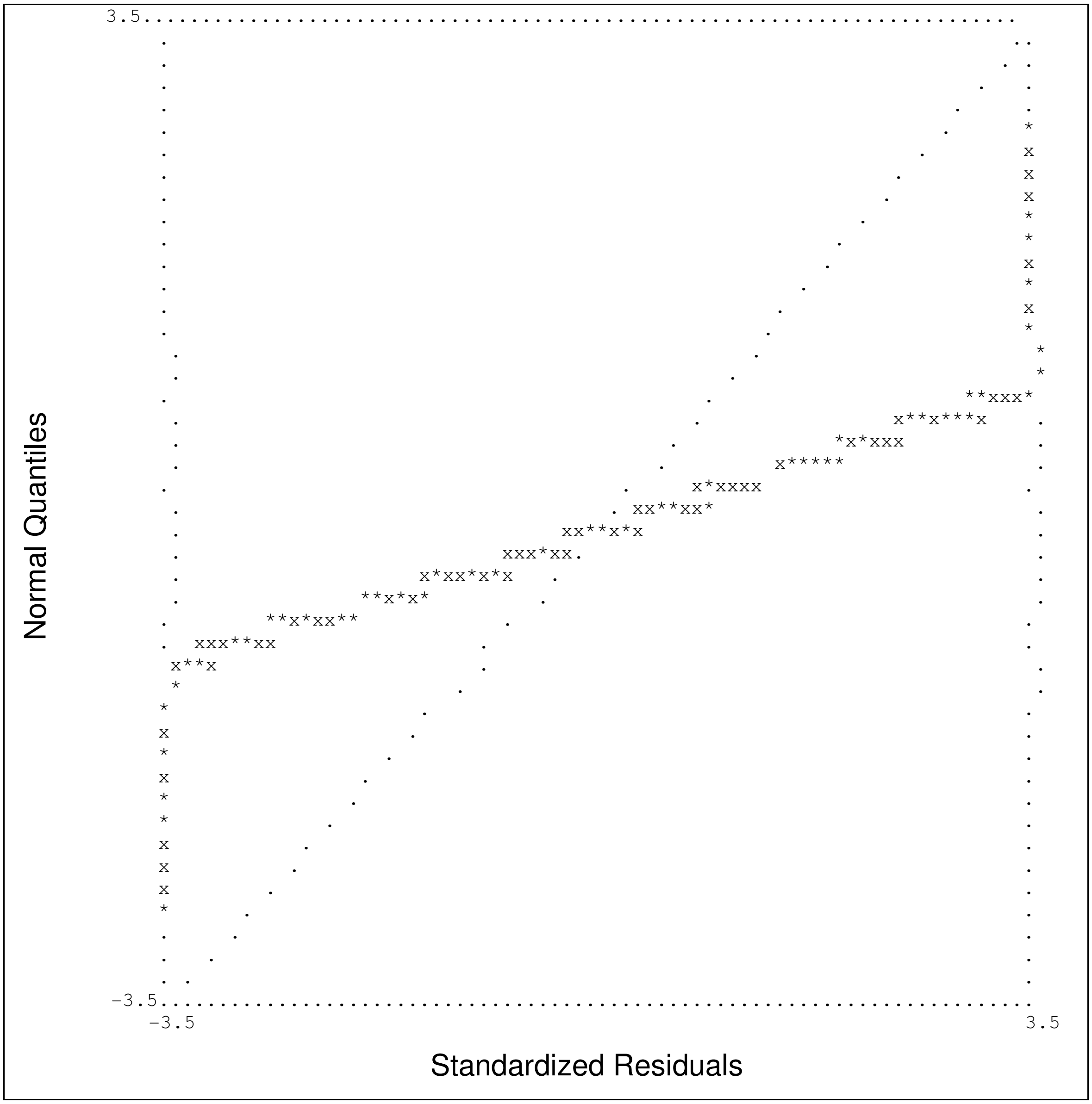

Figure 6.8. Q-Plot of standardized residuals for combined sample: Basic scales model.

6.7.3 Model modification indices

When exploring the Λ_x matrix, results indicate that approximately 60% of the elements are significant. This suggests, as seen with the combined and male sample, that the model would need to be extensively amended to obtain better fit.

When examining the completely standardized expected change matrix no pathways seemed to show a satisfactory loading. Therefore, even though modifying the model could improve fit, the predicted factor loadings, as a result of modification, do not justify freeing these parameters. A similar conclusion can, therefore, be made as seen with the combined and male sample. The same also applies for the proposed freeing of error term correlations.

6.7.4 Assessment of the Orientation Interest scales model

Due to the poor model fit there is no point in interpreting the Orientation Interest measurement model parameter estimates. The poor model fit means that the model with its current parameter estimates cannot, to any acceptable degree of accuracy, reproduce the observed covariance matrix. The parameter estimates are therefore not credible. There is therefore no point in discussing the factor loading estimates, measurement error variance estimates or inter-latent variables correlations.

6.7.5 Summary of model fit assessment for the female sample

Results for this model have, unfortunately, yielded similar results as seen with the combined and male samples. Overall, inadequate fit was found. The model fails to capture the complexity of the process underlying the CISS.

The evaluation of the female sample did not provide suitable fit (at least close fit) results. Consequently, the decision was made to abandon the measurement invariance tests. This decision was in line with the argument underlying the research objectives indicated in chapter 4. For measurement invariance tests to be meaningfully conducted on the CISS on the second-order factor level, credible evidence should have existed that the instrument provides a construct valid measure of the seven Orientation Interest constructs within each gender group. It makes little sense to examine whether measurement model parameters differ across groups if the measurement model does not fit the data of the groups separately. It is, therefore, inappropriate to evaluate the equivalence of a measure that does not fit well with the existing data.

6.8 SUMMARY OF CONFIRMATORY FACTOR ANALYSES

To summarise, the data fitted the CISS measurement model at the Basic scale level reasonably well. However, the exact and close fit hypothesis had to be rejected for the Basic Interest scales. When evaluating model fit of the CISS measurement model at the global Orientation Interests level, the results were inadequate to justify any further MI investigation. This is due to a rejection of the close fit null hypotheses associated with operational hypotheses two, three and four. All three CFAs could not provide evidence in support of close fit. The Orientation measurement model did not fit the data of the combined sample. This was not brought about by differential fit of the model in the two samples. The Orientation measurement model failed to accurately reproduce the observed covariance matrix in both gender samples.

A second-order measurement model would provide a more accurate description of the measurement intention of the test developers. The purpose of the Orientation scales is to summarize the data obtained at the Basic scale level. The Basic Interest scale items were not designed to measure the latent Orientation dimensions directly. A model in which the Basic scale item parcels serve as indicators of the latent Basic Interest dimensions, and the latter in turn are mapped onto the latent Orientation Interest dimensions they reflect would provide a more accurate description of the measurement intention of the test developers and model the dynamics underlying the CISS more accurately. The findings on the Basic Interest measurement model, moreover, seem to suggest that the behavioural responses to the items allocated to a specific interest sub-scale (although primarily determined by the latent Basic Interest dimension they were tasked to reflect) nonetheless depend on the whole of the interest space. If the second-order measurement model is not going to acknowledge this the model fit will remain under pressure.

Fitting a second-order measurement model in which all the elements of Λ_Y are freed could be one option. The approach favoured by Cattell (Cattell, Eber, Tatsuoka, 1970; Gerbing & Tuley, 1991) could be another way of thinking about the CISS. Cattell proposed an approach to sub-scale construction that differs from the conventional approach that aims to construct homogenous sub-scales characterized by high inter-item correlations and high internal consistency.

The approach preferred by Cattell still aims to develop items that are intended to primarily represent a specific latent dimension. At the same time, however, each item to a lesser degree also reflects all of the remaining latent dimensions comprising the construct domain with a pattern of small positive and negative loadings (Gerbing & Tuley, 1991). According to Cattell it is not possible to isolate behavioural indicators that are pure reflections of only a single latent dimension. Although the behavioural indicators placed in a specific sub-scale would primarily reflect the latent dimension measured by that sub-scale, the behavioural indicators would also be (positively and negatively) influenced by all the remaining factors comprising the construct of interest, albeit to a lesser degree. When computing a sub-scale total score the positive and negative loading patterns on the remaining factors are expected to cancel each other out in what Cattell referred to as a *suppressor action* (Cattell, Eber, Tatsuoka, 1970; Gerbing & Tuley, 1991). If such suppressor action did operate on the CISS, it should also have operated (albeit less effectively than on the full length sub-scale) on the level of the item parcels.

A measurement model in which all elements of Λ_Y are not freed should then still reflect the fact that each item of a sub-scale to a lesser degree also reflects all of the remaining latent dimensions comprising the construct domain. If such suppressor action did operate on the CISS, the Basic Interest measurement models have fitted at least closely. This line of reasoning would suggest that the items loading high on one basic Interest dimension are not sufficiently scattered on a hyper plane in the interest space to ensure the effective operation of the suppressor action. A third solution to the problem would be to take the opposite tack and attempt to improve the homogeneity of the sub-scales.

CHAPTER 7
DISCUSSION OF RESULTS

7.1 INTRODUCTION

Making a career choice is certainly a daunting task. Interest assessment is proposed as one of many sources of information that could assist the individual in making and managing their career decisions (Lowman, 1991). If an individual is able to find high levels of fit between what interests them and their ultimate career choice they could achieve career satisfaction which would impact overall psychological wellbeing.

Interest assessment could be used by the I/O psychologist in assisting organisations in making selection decisions as well as, decisions on career planning and career development. These decisions, as backed by person-job/environment fit research (Hesketh 2000; Spokane, Meir & Catalano, 2000), would be of benefit to overall succession planning initiatives in organisations. However, Schmidt and Hunter (1998) found low correlation coefficients between interest results and job performance. It should, however, be noted that there was a slightly higher relationship between interests and performance in training. The limited evidence presented by Schmidt and Hunter (1998) assumes a rather simplistic linear relationship between interests and performance. This leaves the question; whether their findings cannot be attributed to an oversimplified stance regarding the structure of job performance and its associated structural relationship with the person attributes that drive performance which should include interests and the degree of interest-occupation match.

While interest assessment can be helpful both to the individual and organisation, it has met with scepticism from the gender equality movement. Interest assessment instruments have been condemned as being gender biased and typically forcing people into gendered type occupations (Campbell & Hansen, 1981; Kaplan & Saccuzzo, 2001; Murphy & Davidshofer, 2005).

With the backdrop of possible gender bias and perceived unfairness in the assessment of interests, organisations and individuals should take note that their decisions could be adversely affected. Many important individual and organisational decisions may be made with interest assessment results that may not be reflected in the same light across both genders. The instrument used to assess interest attaches a specific connotative definition (Kerlinger & Lee, 2000) to the interest latent variable. Specific latent interest dimensions are distinguished in terms of this conceptualization. Specific items are designed to serve as effect indicators (Hair et al., 2006) of these latent interest dimensions. This design intention is reflected in the scoring key of the instrument. A very specific measurement model is therefore implied by the design intentions (and the scoring key) of the developers of the instrument. A critical question is whether the measurement model reflecting the design intentions of the developers fits data obtained from the instrument at least reasonably well. A further critical question is whether the same measurement model (in terms of the number of latent variables, factor loading pattern and relationships between latent variables) fits the data of the two gender groups separately (i.e., the question whether configural invariance exists). A final critical question is whether the measurement model parameters remain the same across gender. The measurement models underlying interest instruments should be equivalent across gender which would mean that the instrument measures the proposed areas of assessment (interests) in the same manner across the two gender groups. Should these measurement models not display transference across the different groups (in this case gender), then, depending on the nature of the measurement model discrepancies, the test is ultimately testing different latent variables across the respective groups or the test is measuring the same latent variable differently across the respective groups.

While equivalence in terms of the number of factors and the associated pattern of factor loadings (i.e., configural invariance) (Vandenberg & Lance, 2000) might satisfy one level of equivalence, it certainly is not a sufficient condition in ensuring that the latent interest dimension in one group has been equivalently measured in the other group(s). The magnitude of the measurement model parameters could still differ across the different groups and this would still imply non-equivalence in measurement. To be able to confidently interpret observed score differences between genders as indicative of latent score differences, full measurement invariance needs to be indicated.

For equivalence to be observed, the identification and control of bias would be a necessary requirement. Van de Vijver and Leung (1997) describe *bias* as a generic term used for all the nuisance factors threatening the validity of cross-group (cultural) comparisons. Construct, method, and item bias indicates where bias may originate. Van de Vijver and Leung (1997) have also provided a hierarchy that describe different levels of equivalence. Only once equivalence has been met at the highest level can observed scores be compared with confidence. Such confidence rests in the fact that the differences observed in scores between different groups are reflective of a true difference on the underlying latent variable and not due to systematic group effects in measurement.

This then raises the question: how should measurement equivalence be evaluated? Van de Vijver and Leung (1997) have suggested making use of an exploratory factor analytic approach to study measurement equivalence. This approach is a data-driven procedure. However, this study aimed to make use of a procedure that allows for specific hypotheses to be tested regarding measurement equivalence/invariance. Vandenberg and Lance (2000) raised the issue that measurement invariance research in organisational settings should be conducted routinely. Through confirmatory factor analytic procedures, the researcher is able to fit the measurement model implied by the constitutive definition of the construct and the design intentions of the test publishers to data. If reasonable measurement model fit along with significant ($p<.05$) and reasonably high completely standardized factor loadings would be found, it would permit the within gender group use of the instrument to measure the interest construct as constitutively defined. Cross-gender group comparisons would, however, thereby not be sanctioned. An additional vital question is therefore whether the measurement model parameters are the same across the gender groups. In examining these differences the confirmatory technique allows for placing increasing constraints on the model to determine at which level of measurement equivalence is being threatened (Vandenberg & Lance, 2000). Vandenberg and Lance (2000) indicate that invariance should be addressed at three important levels: configural, metric and scalar. These coincide with the taxonomy of

Van de Vijver and Leung (1998), but the confirmatory technique allows for testing the measurement models as per the design intentions of the questionnaire authors/publishers.

For this study, a well-known interest assessment instrument (Gregory, 2004), the CISS (Campbell et al., 1992), was chosen for the purposes of investigating the measurement invariance issue. This questionnaire was chosen as the authors have indicated that they took actions to ensure modernity and relevance to current expectations. This was emphasised as the authors propose that the questionnaire transcends gender archetypes (Campbell et al., 1992).

Evidence on the psychometric integrity of the instrument is reported in the test manual (Campbell et al., 1992). The validity and reliability analysis results reported in the manual, however, all originate from studies performed outside of South Africa. No South African studies that evaluated the reliability and construct validity of the CISS could be traced in the literature. Moreover, none of the studies on the psychometric integrity of the CISS evaluated the fit, through confirmatory factory analytic procedures, of the measurement model implied by the design intentions of the developers.

Therefore, it was decided that research should be conducted on this instrument in the South African setting and the gender equivalence issue should be investigated. The local questionnaire distributors welcomed and authorized the research on this instrument. However, the researcher was not in a position to amend the instrument and the underlying measurement model in any way because the intellectual property rights for the instrument does not reside with the researcher and neither was the researcher mandated by the test publisher to modify the design of the instrument in any way. This further substantiated the need for confirmatory analyses versus an exploratory approach.

7.2 DISCUSSION

Based on the arguments presented in this study, the following central research question was proposed:

"Do the measurement models, implied by the design intentions of the developers of the CISS, fit data obtained for a South African sample on the instrument and are the measurement model parameters equivalent across gender subsamples?"

While this central research question provided the overall objective of the study, a number of specific research questions were suggested. These were as follows:

1. When fitting the interest and skill measurement models to the combined sample, do the models fit the data adequately?
2. When independently fitting the interest and skill measurement models to separate gender samples, do the models fit the data adequately?
3. When fitting the measurement models to the separate gender samples simultaneously, do the models fit adequately when holding only the pattern of factor loadings invariant, while all measurement model parameter estimates are allowed to vary between groups (configural invariance)?
4. When fitting the measurement models to the separate gender samples simultaneously, with all measurement model parameters constrained to be equal across groups, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (omnibus test of measurement invariance)?
5. When fitting the measurement model to the separate gender samples simultaneously with only the factor loadings constrained to be equal across groups but all other model parameters estimated freely, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (lack of metric invariance)?
6. When fitting the measurement model to the separate gender samples simultaneously, with the items (or item parcels) intercepts constrained to be equal across groups but all other model parameters estimated freely, do the fit of the models deteriorate significantly ($p<0.05$) in comparison to the fit obtained when all model parameters are estimated freely (lack of scalar invariance)?

Being able to answer the central research question was dependent on answers gathered for questions 3 to 5. However, prior to answering questions 3 to 5, the measurement models for the CISS should fit the combined sample and the separate gender subsamples.

Upon investigating the measurement model fit of the Basic Interest scales, reasonable model fit was found for the combined sample although the close fit null hypothesis had to be rejected. Unfortunately, independent gender sample confirmatory factor analyses (CFAs) could not be conducted on the Basic Interest scales. Fitting the Basic Interest measurement to the separate gender samples would have meant that the number of model parameters that had to be estimated would have exceeded the number of observations in the gender samples.

However, the possibility of using the gender samples to examine measurement model fit at the global level of measurement (Orientation scales) was investigated. Fortunately, it was found that the number of model parameters that had to be estimated in the Orientation scale measurement model would allow the Orientation scales measurement model to be tested on the separate gender samples. The fit of the Orientation Interest measurement model was evaluated on the combined sample first. Poor model fit was found. However, independent sample CFAs were conducted nonetheless to determine whether the poor fit was due to poor fit within one of the subsamples. Again poor fit was found for both the male and female samples. Therefore, the global interest measurement model did not fit the data from the sample (and gender subsamples) closely. As a result, questions 3 to 5 could not be investigated in this study.

Even though poor fit was found with the Orientation Interest scales some interesting findings should be emphasised. Reasonable model fit was found for the primary level of interest measurement in the CISS, namely the Basic Interest scales. Due to this finding it could be said that the instrument to some degree satisfied the rudimentary levels of construct validity (Hair et al., 2006). Construct validity for this instrument has not been previously evaluated for the South African market. To provide more conclusive evidence on the construct validity of the CISS in South Africa additional analyses would be required. These would have to include, amongst others, the fitting of a structural model that maps the Basic Interest latent variables on to latent outcome or criterion variables that are theoretically assumed to be dependent on the specific latent Basic Interest dimensions. While the global level of measurement forms as a summary component in the measurement of interests, it is not the core level of measurement in this instrument. The Basic scales provide a wealth of information for the career seeker/employee/employer. The positive finding for the combined sample provides some reassurance that the model is able to replicate in the South African context. Of concern, though, is that the gender measurement invariance is still not known for this questionnaire.

During the preliminary phases of the analyses, item and dimensionality analyses were conducted on the respective subsamples. Results for the item analyses revealed some problematic items. Some items were problematic for one sample but not the other and vice versa. However, internal consistency figures for the Basic Interest scales were above the often reported (but somewhat controversial) benchmark of 0.70 (Nunnally, 1978) for both samples. Nonetheless, reliability of the data (on the Basic Interest scales) for South Africa has been established.

Dimensionality analysis was performed via exploratory factor analysis on each Basic Interest scale. The purpose of these analyses was to investigate uni-dimensionality as a possible indicator of poor model fit. An initial finding was that a number of scales particularly for the Basic Interest scales did not meet the uni-dimensionality assumption. In all of these cases where uni-dimensionality was not found a maximum of two factors were extracted. In all cases factor fission presented itself as a plausible explanation for the extraction of more than one factor. This would suggest that some of the scales contain meaningful sub-facets of career interest. This could have been anticipated due to the breadth of item content comprising the Basic Interest scales. However, when forcing single factors the vast majority of “split” factors returned acceptable factor loadings (>0 .50). Gender differences were noted.

Even though gender differences were observed, the utility of the gender differences seen in the dimensionality analyses is limited at this point of the research on the CISS. At this stage of the discussion the poor fit finding for the Orientation Interest model (global level of measurement) should be explored. Specifically the question should be considered why the Basic Interest measurement model should provide at least reasonable fit, but at the global level of measurement, the model fit was poor. This may be due to the diversity of the seven factor structure specified at this level. It is diverse in that a number of Basic Interest factors are said to load onto each of the respective Orientation Interest scales. For example: the Producing Orientation scale summarises information that spans working with cars, machinery (Mechanical Basic scale) to caring for pets, raising and training animals (Animal Care Basic scale). Therefore, complex factors are assumed to load onto a major global factor. It would also seem that each of the Basic scales load with an equal weighting on the respective Orientation scale. This may not be a realistic assumption.

In addition, the seven Orientation scales are indicated by the questionnaire authors as a summary of all 29 factors. The development of the global level of measurement was based on an exploratory empirical study with a specific US sample. There is a possibility that the functioning of the data in the South African context could be a reason for problems experienced with model fit. This could be investigated by conducting item analyses and exploratory factor analysis at the global level. Uni-dimensionality cannot be expected for the seven factors due to the variety of specific themes that the primary factors represent that are combined in each Orientation factor. It should be noted that the authors of the CISS, did indicate that upon conducting the principle components analyses in the development/refinement of the instrument (Campbell et al., 1992) three options (the appropriateness of a five, six and seven-component solution was considered) for permissible factor structures had emerged for the Interest Orientations. It was decided that the seven factor structure represented the second order factor structure in the best manner. It could be argued that this approach to the development of the Orientation scales may have limited the replication ability of the model in the South African context.

A second-order measurement model would provide a more accurate description of the measurement intention of the test developers. The purpose of the Orientation scales is to summarize the data obtained at the Basic scale level. The Basic Interest scale items were not designed to measure the latent Orientation dimensions directly. A model in which the Basic scale item parcels serve as indicators of the latent Basic Interest dimensions and the latter in turn are mapped onto the latent Orientation Interest dimensions they reflect, would provide a more accurate description of the measurement intention of the test developers and model the dynamics underlying the CISS more accurately. Moreover, the findings on the Basic Interest measurement model seem to suggest that the behavioural responses to the items allocated to a specific interest sub-scale (although primarily determined by the latent Basic Interest dimension they were tasked to reflect) nonetheless depend on the whole of the interest space. If the second-order measurement model is not going to acknowledge this the model fit will remain under pressure.

7.3 LIMITATIONS OF THE STUDY

A particular limitation of the study is sample size [male (*n*=408), female (*n*=402), combined (*N*=810)]. The measurement invariance tests could not be conducted at the primary level of measurement (i.e., Basic Interest scales) due to sample size restrictions. This is regrettable as the reasonable fit attained for the Basic Interest model, would have allowed for the execution of the multi-group measurement invariance tests. This would have been ideal considering this model is the core of the CISS. A further limitation of the sample is the lack of descriptive demographic information regarding the composition of the sample. Some of the observations made during the analyses could have been a function of the composition of the sample, thereby supporting the creation of further hypotheses to be tested. Also, in the event of obtaining further information regarding the composition of the sample, for example educational background, race or stage of employment, further invariance tests could be conducted.

In the present study it would have been ideal to use individual items as indicator variables that represent the latent interest dimensions in the model, due to recommendations made regarding the appropriateness of the item as apposed to item parcels for measurement invariance tests (Bandalos, 2002; Bandalos & Finney, 2001; Meade & Lautenschleager, 2004). In addition, by making use of individual items differential item functioning also could have been evaluated for the models which may have assisted in determining item level sources of model fit problems. However, once again sample size would have caused estimation problems, due to the sample size restrictions. Even though item parcels could be considered composite manifestation of the latent variables, when wanting to evaluate model fit and subsequent measurement invariance tests for a questionnaire, the questionnaire in its purest form should be utilized.

This research did not investigate the self-perceived Basic and Orientation Skill measurement models. This should be addressed in a further study, but with sufficiently large samples that would allow the CFAs to be conducted at the primary and global levels of measurement.

This study was unable to provide information regarding differences in interest scores between the gender groups. This information would have been valuable in understanding the composition of interest in the South African context across the genders. *Mean* differences may provide valuable information about the perception of careers in South Africa across the genders. This could not be achieved as full measurement equivalence could not be established.

The nature of the findings of the study, in conjunction with the limitations thereof, precludes any definite verdict on the construct validity of the CISS as a measure of interest in South Africa, as well as on the measurement invariance of the instrument across gender groups.

7.4 RECOMMENDATIONS FOR RESEARCHERS AND PRACTITIONERS

If possible, data on a larger sample should be gathered in order to conduct the measurement invariance tests at the primary factor level; Basic Interest and Skill scales. It would also be suggested that the sample size should be adequate enough to conduct the analyses using the original items of the questionnaire. Specific attention should be given to investigating the item statistics, uni-dimensionality and confirmatory factor analyses on the Basic and Orientation Skill models.

The construct validity of the CISS should be further evaluated by mapping the Basic Interest dimensions onto outcome latent variables that are theoretically thought to be affected by the interest dimensions and the fit of the resultant structural model should be evaluated.

The second-order measurement model should be fitted on the combined sample (preferably with individual items as indicator variables) and on the separate gender samples. The possibility of conducting measurement invariance analysis on the second-order measurement model should be considered.

For South African practitioners, it would be recommended that the interpretation and communication of results to questionnaire respondents should be focussed on the measurement attained at the primary level as model fit was reasonable here. Not only is this level of measurement more detailed and perhaps more helpful to the client, its measurement model CFA results suggest that a fair amount of confidence can be placed in how the model could be replicated in the South African population.

Researchers and practitioners are cautioned against calculating latent *means* for both the Basic and Orientation scales and making comparisons between the genders. This type of comparison can not be completed before measurement invariance has been established.

REFERENCES

Association for Measurement and Evaluation in Guidance (1973). AMEG Commission report on sex bias in interest measurement. *Measurement and Evaluation in Guidance, 6,* 171-177.

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

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 parceling issues in structural equation modeling. In G.A Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Berry, J.W., Poortinga, Y.H., Segall, M.H., & Dasan, P.R. (2002). *Cross-cultural psychology: Research and applications* (2nd ed.). Cambridge: Cambridge University Press.

Boggs, K.R. (1999). Campbell Interest and Skill Survey: review and critique. *Measurement and Evaluation in Counseling and Development, 32* (3), 168 – 182.

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

Byrne, B.M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models.* New York: Springer Verlag.

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

Byrne, B.M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105* (3), 456-466.

Byrne, B.M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34* (2), 155 – 175.

Campbell, D.P. (1974). *Manual for the SVIB/SCII.* Stanford, CA: Stanford University Press.

Campbell, D.P. (1993). A new integrated battery of psychological surveys. *Journal of Counseling and Development, 71* (5), 575 – 587.

Campbell, D.P. (1995). The Campbell Interest and Skill Survey (CISS): A product of ninety years of psychometric evolution. *Journal of Career Assessment, 3* (4), 391 – 410.

Campbell, D.P., Hyne, S.A., & Nilsen, D.L. (1992). *Manual for the Campbell Interest and Skill Survey.* Minneapolis, MN: NCS Pearson, Inc.

Campbell, D., & Hansen, J. (1981). *Manual for the SVIB-SCII.* Palo Alto, CA: Stanford University Press.

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

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling, 9* (2), 233-255.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace.

De Bruin, G.P. (2001). Career counselling assessment. In C. Foxcroft & G. Roodt (Eds.), *An introduction to psychological assessment in the South African context.* Cape Town: Oxford University Press.

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

Donnay, D.A.C. (1997). E.K. Strong's legacy and beyond: 70 years of the Strong Interest Inventory. *The Career Development Quarterly, 46* (1), 2 – 22.

Dunbar-Isaacson, H. (2006). *An investigation into the measurement invariance of the performance index.* Stellenbosch: Unpublished dissertation, University of Stellenbosch.

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

Edwards, A.L. (1957). *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts..

Edwards, A.L. (1970). *The measurement of personality traits by scales and inventories.* New York: Holt, Rinehart & Winston.

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

Fritzsche, B.A., Powell, A.B., & Hoffman, R. (1999). Person-environment congruence as a predictor of customer service performance. *Journal of Vocational Behavior, 54,* 59-70.

Gerbing, D.W., & Tuley, M.R. (1991). The 16PF related to the five-factor model of personality: Multiple-indicator measurement versus a priori scales. *Multivariate Behavioural Research, 26* (2), 271-289.

Gorsuch,R.L. (2003). Factor analysis. In J.A. Schinka & W.F. Velicer, (Eds.), *Handbook of psychology: Volume 2 research methods in psychology* (pp. 143-164). Hoboken, NJ:John Wiley & Sons.

Greenhaus, J.H., Callanan, G.A., & Godshalk, V.M. (2000). Career management (3rd ed.). Orlando, FL: Harcourt Inc.

Gregory, R (2004). *Psychological testing: History, principles, and applications* (4th ed.). Boston, MA: Pearson.

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

Hall, R.J., Snell, A.F., & Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2* (3), 233-251.

Hansen, J.C., & Campbell, D.P. (1985). *Manual for the SVIB-SCII* (4th ed.). Standford, CA: Stanford University Press.

Hansen, J. C., & Leuty, M. (2007). Evidence of validity for the Skill scale scores of the Campbell Interest and Skill Survey. *Journal of Vocational Behavior, 71*, 23-44.

Hansen, J.C., & Neuman, J. (1999). Evidence of concurrent prediction of the Campbell Interest and Skills Survey for college major selection. *Journal of Career Assessment, 7*, 239-247.

Hesketh, B. (2000). The next millennium of 'fit' research: comments on 'The congruence myth: an analysis of the efficacy of the person-environment fit model' by H.E.A. Tinsley. *Journal of Vocational Behavior, 56*, 190-196.

Holland, J.L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6,* 35 - 45.

Holland, J.L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117 – 144.

Hu, L.T., & Bentler P.M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and applications.* Thousand Oaks, California: Sage Publications.

Hulin, C.L., Drasgow F., & Parsons C.K. (1983). *Item response theory: Application to psychological measurement.* Homewood, Ill.: Jones-Irwin Publishers.

Janda, L.H. (1998). *Psychological testing: Theory and applications.* Boston, MA: Allyn and Bacon.

Johnson, T.J., Kulesa, P., Cho, Y.I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36* (2), 264-277.

Jöreskog, K.G., & Sörbom, D (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* United States of America: Scientific Software International, Inc.

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

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

Kaplan, R.M., & Saccuzzo, D.P. (2001). *Psychological testing: Principles, applications, and issues* (5^{th} ed.). Belmont, CA: Wadsworth/Thomson Learning.

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

Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Harcourt College Publishers.

Kim, S., & Hagtvet, K.A. (2003). The impact of misspecified item parcelling on representing latent variables in covariance structure modelling: a simulation study. *Structural Equation Modeling, 10* (1), 101-127.

Kline, P. (1994). *An easy guide to factor analysis.* London: Routledge

Kline, R.B. (2005). *Principles and practice of structural equation modelling* (2^{nd} ed.). New York: The Guilford Press.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lowman, R.L. (1991). *The clinical practice of career assessment: Interests, abilities, and personality.* Washington, DC: APA.

MacCallum, R.C. (1995). Model specification: procedures, strategies and related issues. In Hoyle, R.H. (Ed.), *Structural equation modelling: Concepts, issues and applications.* Thousand Oaks, CA: SAGE Publications, Inc.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods, 1* (2), 130-149.

Meade, A.W., & Lautenschlager, G.J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7* (4), 361-388.

Meade, A.W., & Kroustalis (2006). Problems with item parcelling for confirmatory factor analytic tests of measurement invariance. *Organizational Research Methods, 9* (3), 369-403.

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

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.

Murphy, K.R., & Davidshofer, C.O. (2005). *Psychological testing: principles and applications.* (6^{th} ed.). Upper Saddle River, NJ: Pearson Education Inc.

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.

Nakamura, J., & Csikszentmihalyi, M. (2005). The concept of flow. In C.R. Snyder & S.J. Lopez (Eds.), *Handbook of Positive* Psychology (pp. 89-105). New York: Oxford University Press.

Nasser, F., Takahashi, T., & Benson,J. (1997). The structure of test anxiety in Israeli-Arab high school students: An application of confirmatory factor analysis with mini-scales. *Anxiety, Stress, and Coping, 10*, 129 – 151.

Neuman, W.L. (2000). *Social research methods: qualitative and quantitative approach* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

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

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modelling. *European Journal of Operational Research, 151*, 53-79.

Owen, K., & Taljaard, J.J. (1988). *Handleiding vir die gebruik van Sielkundige en Skolastiese toetse van IPEN en die NIPN.* Pretoria: Raad vir Geesteswetenskaplike Navorsing.

Pousette, A., & Hanse, J.J. (2002). Job characteristics as predictors of ill-health and sickness absenteeism in different occupational types: A multi-group structural equation modelling approach. *Work and Stress, 16* (3), 229-250.

Raykov, T., & Marcoulides, G.A. (2008). *An introduction to applied multivariate analysis.* New York: Routledge.

Republic of South Africa (RSA) (1977). Health Professions Act (Act no 66 of 1995). *Government Gazette.*

Riordan, C.M., & Vandenberg, R.J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20* (3), 643-671.

Rasch, G. (1980). *Probabilistic models for intelligence and attainment testing*. Chicago: University of Chicago Press.

Robitschek, C. (2004). Vocational psychology assessment: Positive human characteristics leading to positive work outcomes. In S.J. Lopez, & C.R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures.* Washington, DC: APA.

Sass, D.A., & Smith, P.L. (2006). The effects of parcelling unidimensional scales on structural parameter estimates in structural equation modelling. *Structural Equation Modeling, 13* (4), 566-586.

Satorra, A., & Bentler, P. M. (1999). *A scale difference chi-square test statistic for moment structure analysis* (UCLA Statistics Series 260). Los Angeles: University of California, Department of Psychology.

Sauermann, H. (2005). Vocational choice – a decision making perspective. *Journal of Vocational. Behavior, 66* (2), 273-303.

Savickas, M.L., Taber, B.J., & Spokane, A.R. (2002). Convergent and discriminant validity of five interest inventories. *Journal of Vocational Behavior, 61*, 139–184.

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* (2), 262-274.

Schumacker, R.E., & Lomax, R.G. (1996). *A beginner's guide to structural equation modeling.* Mahaw, NJ: Lawrence Erlbaum Associates, Publishers.

Sekaran, U. (1983). Methodological and theoretical issues and advancements in cross-cultural research. *Journal of International Business Studies, 14* (2), 61-73.

Singh, R., & Greenhaus, J.H. (2004). The relation between career decision-making strategies and person-job fit: a study of job changers. *Journal of Vocational Behavior, 64,* 198-221.

Spangenberg, H.H., & Theron C.C. (2005). Promoting ethical follower behaviour through leadership of ethics: The development and psychometric evaluation of the ethical leadership inventory (ELI). *SA Journal of Business Management, 36* (2), 1-18.

Spokane, A.R., Meir, E.I., & Catalano, M. (2000). Person-Environment congruence and Holland's theory: A review and reconsideration. *Journal of Vocational Behavior, 57*, 137-187.

Steenkamp, J.-B.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, *103* (2684), 677-680.

Stewart, D. (2001). Factor analysis. *Journal of Consumer Psychology, 10* (1&2), 75-82.

Strong, E. K., Jr. (1927). *Vocational Interest Blank.* Stanford, CA: Stanford University Press.

Sullivan B.A., & Hansen J.C. (2004). Evidence of construct validity of the interest scales on the Campbell Interest and Skill Survey. *Journal of Vocational Behavior, 66*, 179-202.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics.* New York: Harper Collins Publishers.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

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

Theron, C.C. (1999). *Psychometric implications of corrections for attenuation and restriction of range for selection validation research.* Unpublished doctoral dissertation, University of Stellenbosch.

Theron, C.C. (2006). *Intermediate Statistics and Computer Usage*. Unpublished class notes (Industrial Psychology 815), University of Stellenbosch.

Theron, C.C., & Spangenberg, H.H. (2004). *Towards a comprehensive leadership unit performance structural model: The development of second-order factors for the leadership behaviour inventory (LBI).* Manuscript accepted for publication in Management Dynamics.

Tucker, L.R. 1951. *A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984).* Washington, DC: Department of the Army.

Van de Vijver, F.J.R. (2003). Bias and equivalence: Cross-cultural perspectives. In J.A. Harkness, F.J.R. Van de Vijver & P.P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 143-155). New York: John Wiley & Sons.

Van de Vijver, F.J.R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage Publications Inc.

Van de Vijver, F.J.R., & Poortinga Y.H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment, 13*, 21-29.

Van Herk, H., Poortinga, Y.H., & Verhallen, T.M.M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology, 35* (3), 346-360.

Van de Vijver, F.J.R., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 47*, 263-279.

Vandenberg, R.J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods, 5* (2), 139-158.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Watkins, C.E., & Campbell (Eds.) (2000). *Testing and assessment in counseling practice* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

APPENDIX 1: ITEM STATISTICS: INTEREST MODEL

	Male Sample					Female Sample				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
ADEVI1	11.0588	13.8245	0.549	0.324	0.691	10.8507	12.6310	0.548	0.302	0.614
ADEVI2	11.2426	12.8968	0.607	0.391	0.657	10.6816	12.9807	0.450	0.219	0.676
ADEVI3	11.4387	14.4238	0.532	0.315	0.701	11.0697	12.7034	0.504	0.263	0.641
ADEVI4	11.3039	15.4504	0.499	0.260	0.718	11.1443	13.4704	0.482	0.249	0.654
ADVI1	24.1225	82.3830	0.567	0.358	0.926	22.8930	69.2828	0.507	0.324	0.894
ADVI2	24.2083	75.8214	0.778	0.673	0.910	23.0423	64.1005	0.683	0.594	0.879
ADVI3	24.4387	76.6400	0.757	0.611	0.912	23.3308	64.2519	0.672	0.623	0.880
ADVI4	24.4069	73.5884	0.799	0.698	0.908	23.2438	62.2696	0.735	0.624	0.874
ADVI5	24.2426	74.4840	0.851	0.752	0.904	23.0746	62.3685	0.767	0.663	0.871
ADVI6	24.3505	81.5206	0.664	0.568	0.919	23.3085	65.8697	0.651	0.566	0.882
ADVI7	24.3137	76.5008	0.818	0.728	0.907	23.1517	62.9669	0.776	0.640	0.870
ADVI8	24.1152	78.8884	0.699	0.513	0.916	22.9154	66.5464	0.581	0.483	0.888
ANI1	10.0784	16.8194	0.658	0.436	0.839	11.1468	22.6043	0.818	0.693	0.896
ANI2	10.1029	16.2302	0.711	0.518	0.818	11.1045	22.8070	0.811	0.662	0.898
ANI3	10.0588	16.0260	0.753	0.571	0.800	11.1194	21.9907	0.861	0.747	0.881
ANI4	10.0760	16.4340	0.695	0.486	0.824	11.0323	23.3980	0.775	0.621	0.910
ARTI1	17.6985	30.1423	0.502	0.584	0.755	16.5149	32.3502	0.559	0.523	0.776
ARTI2	17.6863	28.2895	0.642	0.513	0.719	16.5423	29.5606	0.685	0.558	0.746
ARTI4	17.7010	30.4804	0.509	0.533	0.753	16.3905	32.3833	0.512	0.427	0.786
ARTI5	17.8554	29.4557	0.535	0.465	0.747	16.5398	30.1393	0.640	0.553	0.757
ARTI3	17.7770	30.0951	0.540	0.300	0.745	16.8582	32.0771	0.498	0.308	0.790
ARTI6	17.8627	31.7502	0.444	0.339	0.768	16.3955	32.8332	0.491	0.373	0.791
ATHI1	21.0294	49.1687	0.694	0.567	0.792	21.6169	45.5038	0.650	0.458	0.741
ATHI2	20.9583	53.4847	0.529	0.417	0.818	21.6418	47.9512	0.570	0.425	0.757
ATHI3	21.1716	46.6290	0.758	0.737	0.779	21.5721	43.0634	0.662	0.695	0.736
ATHI4	20.9559	56.6467	0.332	0.145	0.848	21.5796	52.3989	0.277	0.125	0.811
ATHI5	21.1054	47.9127	0.733	0.706	0.785	21.5050	42.8441	0.684	0.691	0.731
ATHI6	20.8358	51.6413	0.469	0.254	0.831	22.0796	54.2430	0.204	0.175	0.823
ATHI7	20.9289	53.3881	0.603	0.447	0.808	21.5423	46.7476	0.689	0.511	0.739
CDEVI1	21.6618	58.1900	0.781	0.662	0.876	20.8955	51.9541	0.758	0.658	0.842
CDEVI2	21.4069	61.8635	0.688	0.531	0.887	20.9353	52.9883	0.757	0.654	0.843
CDEVI3	21.4755	56.2254	0.856	0.762	0.867	20.9254	50.4433	0.822	0.747	0.833
CDEVI4	21.5686	65.6071	0.559	0.318	0.901	20.8856	61.3385	0.347	0.135	0.894
CDEVI5	21.4559	61.5066	0.717	0.528	0.884	21.0448	55.0404	0.621	0.421	0.861
CDEVI6	21.6299	58.5089	0.730	0.616	0.883	20.9328	52.7810	0.700	0.579	0.850
CDEVI7	21.5368	65.2222	0.611	0.388	0.895	21.0672	56.2723	0.593	0.395	0.864
COUNI1	18.0515	30.5059	0.547	0.306	0.763	17.2189	34.3609	0.624	0.406	0.766
COUNI2	18.1250	29.2546	0.643	0.629	0.739	17.2736	31.8451	0.702	0.695	0.746
COUNI3	18.2525	32.2531	0.515	0.303	0.770	17.0149	39.4761	0.361	0.145	0.820
COUNI4	18.1691	29.0401	0.629	0.616	0.742	17.3358	31.5652	0.705	0.695	0.745
COUNI5	18.4975	32.7568	0.436	0.219	0.788	17.1294	36.6017	0.506	0.267	0.792

	Male Sample					Female Sample				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
COUNI6	18.3652	32.0751	0.514	0.287	0.770	17.1244	37.1665	0.512	0.282	0.790
CULI1	16.4706	33.6011	0.565	0.467	0.817	16.0448	40.1626	0.643	0.482	0.859
CULI2	16.5417	31.6788	0.729	0.539	0.782	16.2264	37.2030	0.794	0.646	0.832
CULI3	16.3873	36.4049	0.412	0.225	0.846	16.2861	43.3968	0.489	0.276	0.883
CULI4	16.6471	31.6098	0.733	0.553	0.782	16.3582	37.9761	0.753	0.633	0.839
CULI5	16.3725	34.9174	0.563	0.395	0.816	16.3905	39.5803	0.675	0.525	0.853
CULI6	16.4412	32.2766	0.664	0.496	0.796	16.1194	38.7139	0.713	0.533	0.847
FARMI1	17.7549	41.5123	0.745	0.578	0.868	19.6368	54.3915	0.845	0.729	0.907
FARMI2	17.6642	43.6241	0.715	0.556	0.873	19.4975	55.1384	0.826	0.740	0.910
FARMI3	17.7255	43.4920	0.663	0.469	0.881	19.6070	57.1020	0.756	0.627	0.919
FARMI4	17.6716	42.2506	0.731	0.584	0.870	19.4900	58.1757	0.778	0.636	0.916
FARMI5	17.6765	41.6248	0.764	0.629	0.865	19.5149	55.7167	0.822	0.734	0.910
FARMI6	17.8922	41.8262	0.662	0.446	0.882	19.7413	55.4591	0.725	0.558	0.924
FASHI1	17.0735	47.9012	0.588	0.490	0.865	16.4030	32.4357	0.480	0.471	0.615
FASHI2	17.1863	47.2379	0.659	0.557	0.854	16.5299	31.3669	0.557	0.507	0.591
FASHI3	17.1985	42.7197	0.717	0.664	0.844	16.2090	34.0061	0.388	0.194	0.645
FASHI4	17.2623	43.0588	0.685	0.669	0.850	16.2313	29.7992	0.270	0.134	0.728
FASHI5	17.4706	45.1146	0.770	0.611	0.835	16.5572	32.2972	0.564	0.447	0.594
FASHI6	17.2402	48.8169	0.651	0.448	0.856	16.2413	36.1686	0.348	0.212	0.657
FINI1	24.5319	70.6968	0.654	0.454	0.884	25.6269	97.5013	0.722	0.544	0.921
FINI2	24.5147	68.9678	0.674	0.519	0.882	25.7313	95.4189	0.772	0.618	0.917
FINI3	24.7059	73.6184	0.492	0.346	0.899	25.4776	97.2925	0.681	0.529	0.924
FINI4	24.6225	68.2896	0.747	0.583	0.875	25.7164	94.2436	0.809	0.687	0.914
FINI5	24.5539	68.1789	0.693	0.514	0.880	25.5697	94.4303	0.781	0.625	0.916
FINI6	24.6103	69.1623	0.750	0.639	0.876	25.9552	99.5541	0.718	0.672	0.921
FINI7	24.4632	68.4016	0.756	0.625	0.875	25.7811	96.1764	0.804	0.736	0.915
FINI8	24.7525	69.3268	0.656	0.489	0.884	25.7711	96.8901	0.740	0.595	0.919
INTI1	14.2451	24.6327	0.460	0.420	0.722	13.2562	29.9616	0.555	0.395	0.800
INTI2	14.2941	23.7315	0.558	0.367	0.687	13.4851	28.5497	0.615	0.414	0.783
INTI3	14.2402	23.9864	0.533	0.398	0.696	13.5448	29.0217	0.597	0.388	0.788
INTI4	14.0760	24.1981	0.445	0.373	0.730	13.7065	27.2104	0.624	0.495	0.780
INTI5	14.1054	23.3623	0.574	0.338	0.681	13.4303	27.5276	0.665	0.472	0.767
LAWI1	32.1569	110.9680	0.680	0.515	0.889	33.4677	126.0351	0.681	0.493	0.890
LAWI2	32.0319	117.2791	0.586	0.415	0.895	33.2985	130.0703	0.610	0.447	0.895
LAWI3	32.2843	109.4374	0.697	0.572	0.888	33.4677	123.2072	0.687	0.587	0.890
LAWI4	32.2157	109.2457	0.755	0.602	0.884	33.2090	122.8789	0.739	0.637	0.886
LAWI5	32.3407	111.9451	0.647	0.578	0.891	33.4950	126.4601	0.606	0.606	0.895
LAWI6	32.2328	111.9039	0.662	0.543	0.890	33.1766	121.9263	0.710	0.628	0.888
LAWI7	32.3382	115.8067	0.604	0.489	0.894	33.5000	126.6147	0.637	0.499	0.893
LAWI8	32.2328	115.0046	0.616	0.413	0.893	33.5423	132.5830	0.568	0.379	0.897
LAWI9	32.3088	111.2754	0.670	0.551	0.890	33.3333	120.7814	0.716	0.660	0.888
LAWI10	32.0931	114.9888	0.591	0.391	0.895	33.3383	129.3815	0.584	0.380	0.896
LEADI1	22.1324	48.5377	0.646	0.504	0.825	20.9552	38.8509	0.621	0.425	0.728
LEADI2	21.7794	52.2215	0.573	0.452	0.835	20.2960	41.0618	0.502	0.492	0.753
LEADI3	22.2426	47.5700	0.675	0.617	0.820	21.1617	41.0636	0.513	0.537	0.751

	Male Sample					Female Sample				
Item	scale Mean if Item Deleted	scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	scale Mean if Item Deleted	scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
LEADI4	22.1029	49.5963	0.703	0.576	0.817	21.0473	42.0402	0.504	0.521	0.753
LEADI5	22.1103	51.4350	0.634	0.425	0.827	20.8483	41.7051	0.544	0.310	0.746
LEADI6	21.8971	53.6847	0.580	0.369	0.835	20.7811	43.1240	0.527	0.330	0.750
LAWI4	22.1471	53.4722	0.480	0.414	0.849	20.3582	44.0360	0.343	0.426	0.786
MATHI1	24.8995	61.6828	0.602	0.543	0.885	23.7612	75.1398	0.532	0.459	0.900
MATHI2	24.8578	61.0608	0.639	0.479	0.880	23.7736	69.2628	0.712	0.592	0.881
MATHI3	24.8260	58.6306	0.709	0.624	0.872	23.8209	64.9304	0.803	0.759	0.869
MATHI4	24.9706	61.3259	0.593	0.555	0.886	23.7512	74.4866	0.567	0.494	0.896
MATHI5	24.6961	56.0057	0.825	0.732	0.857	23.9129	64.1545	0.842	0.795	0.864
MATHI6	24.7647	60.4113	0.671	0.473	0.877	23.8433	71.6786	0.640	0.420	0.889
MATHI7	24.8235	58.2636	0.775	0.642	0.864	23.9279	66.2516	0.799	0.703	0.870
MECHI1	30.6569	88.2800	0.123	0.313	0.885	31.9751	138.1739	0.729	0.682	0.935
MECHI2	30.3701	76.2779	0.692	0.520	0.827	31.9104	135.7276	0.862	0.770	0.927
MECHI3	30.2500	75.3771	0.599	0.628	0.835	32.3731	147.8554	0.601	0.560	0.941
MECHI4	30.4804	73.7687	0.747	0.609	0.820	32.0224	136.0020	0.836	0.764	0.928
MECHI5	30.3431	75.7640	0.664	0.647	0.829	32.0547	138.3760	0.847	0.773	0.928
MECHI6	30.1838	72.9907	0.666	0.742	0.827	32.1493	139.9228	0.739	0.758	0.934
MECHI7	30.8039	81.6814	0.398	0.366	0.854	32.1119	137.3266	0.742	0.670	0.934
MECHI8	30.2819	74.1931	0.691	0.554	0.825	32.1965	140.6820	0.736	0.604	0.934
MECHI9	30.4338	74.2167	0.710	0.598	0.824	32.0224	136.3112	0.832	0.748	0.929
MEDI1	40.9902	165.6805	0.645	0.585	0.920	39.9826	159.5383	0.566	0.463	0.906
MEDI2	41.3137	163.4640	0.662	0.716	0.919	40.0473	157.8456	0.684	0.706	0.900
MEDI3	41.4804	168.2551	0.616	0.467	0.921	40.4602	170.5483	0.389	0.365	0.913
MEDI4	41.2672	160.5206	0.757	0.617	0.915	40.1716	154.6563	0.691	0.569	0.900
MEDI5	41.3162	164.5067	0.748	0.643	0.916	40.4080	157.4940	0.672	0.610	0.901
MEDI6	41.2328	161.5058	0.814	0.720	0.913	40.1791	154.9155	0.768	0.690	0.896
MEDI7	41.2843	166.5136	0.721	0.571	0.917	40.3881	159.7692	0.640	0.457	0.902
MEDI8	40.9363	166.3055	0.630	0.599	0.920	40.0821	157.9509	0.619	0.533	0.903
MEDI9	41.3701	165.7718	0.646	0.666	0.920	40.0821	159.9010	0.654	0.640	0.902
MEDI10	41.2206	162.3296	0.755	0.632	0.915	40.1070	157.2379	0.731	0.602	0.898
MEDI11	41.3137	168.4615	0.604	0.510	0.921	40.2637	159.1123	0.661	0.499	0.901
MEDI12	41.5711	171.7689	0.595	0.412	0.921	40.2687	162.2568	0.641	0.475	0.902
MILI1	24.0049	70.6191	0.481	0.357	0.868	26.7313	94.8553	0.417	0.395	0.882
MILI2	24.0466	64.6784	0.733	0.633	0.839	26.0199	80.8974	0.773	0.773	0.845
MILI3	23.9975	65.3489	0.699	0.559	0.843	26.1393	80.7287	0.796	0.726	0.843
MILI4	24.0441	68.6369	0.616	0.460	0.853	26.3333	88.6517	0.647	0.460	0.860
MILI5	24.0466	66.5900	0.675	0.583	0.846	26.0174	82.4062	0.755	0.722	0.848
MILI6	24.1176	66.6741	0.654	0.574	0.848	26.0498	80.6758	0.767	0.788	0.846
MILI7	24.0882	66.2674	0.681	0.496	0.846	26.4378	84.6707	0.715	0.578	0.852
MEDI11	23.7255	72.6075	0.437	0.291	0.871	26.4279	102.2504	0.212	0.111	0.898
OPI1	28.3799	79.7153	0.607	0.412	0.858	28.3881	83.8041	0.545	0.342	0.861
OPI2	28.4069	78.7579	0.544	0.397	0.864	28.2562	83.2684	0.499	0.336	0.865
OPI3	28.5172	79.1889	0.608	0.401	0.858	28.2164	76.5790	0.656	0.489	0.851
OPI4	28.3603	78.4030	0.674	0.461	0.853	28.4055	80.1319	0.645	0.428	0.852
OPI5	28.4877	75.4593	0.668	0.536	0.852	28.6542	83.0298	0.516	0.354	0.864

	Male Sample					Female Sample				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
OPI6	28.1642	86.0196	0.363	0.325	0.878	28.2065	80.5932	0.621	0.456	0.854
OPI7	28.4681	74.0285	0.755	0.661	0.844	28.2662	79.1434	0.698	0.552	0.847
OPI8	28.4436	72.9501	0.694	0.669	0.850	28.2711	78.5472	0.644	0.494	0.852
SUPI8	28.3799	80.7374	0.569	0.462	0.862	28.2214	80.6416	0.617	0.454	0.854
PERI1	31.0858	117.6609	0.454	0.424	0.887	31.5871	112.0585	0.606	0.608	0.862
PERI2	31.3211	112.1890	0.600	0.449	0.877	31.7214	116.7751	0.441	0.271	0.875
PERI3	31.0417	112.9884	0.635	0.515	0.875	31.4453	112.0731	0.620	0.505	0.861
PERI4	31.2770	110.2351	0.650	0.615	0.874	31.4925	108.8491	0.693	0.581	0.856
PERI5	31.5637	113.7502	0.551	0.428	0.881	31.0746	114.6927	0.453	0.493	0.875
PERI6	31.3873	109.6777	0.705	0.603	0.870	31.4453	106.5818	0.733	0.611	0.852
PERI7	31.2402	112.0208	0.618	0.487	0.876	31.1766	116.6745	0.437	0.392	0.875
PERI8	31.4216	113.1585	0.595	0.504	0.878	31.6269	112.3692	0.593	0.526	0.863
PERI9	31.3627	110.7919	0.679	0.610	0.872	31.5100	109.4326	0.692	0.568	0.856
PERI10	31.3578	109.4441	0.734	0.649	0.868	31.5398	108.4186	0.739	0.675	0.852
PNTI1	14.3309	25.6470	0.754	0.574	0.859	14.7786	29.7788	0.781	0.615	0.866
PNTI2	14.3284	28.1327	0.613	0.415	0.890	15.1567	32.2871	0.649	0.482	0.894
PNTI3	14.2647	25.4924	0.784	0.633	0.853	14.7537	29.8320	0.760	0.611	0.871
PNTI4	14.3554	25.6252	0.728	0.589	0.866	14.7214	30.0419	0.742	0.605	0.875
PNTI5	14.3382	25.8558	0.775	0.603	0.855	14.9080	29.9641	0.796	0.649	0.863
PUBI1	10.7843	13.7863	0.694	0.546	0.718	10.8234	14.0410	0.636	0.496	0.687
PUBI2	10.6936	13.9526	0.723	0.577	0.704	10.6368	14.2119	0.715	0.551	0.646
PUBI3	10.9853	16.0096	0.626	0.405	0.755	10.8159	16.6244	0.577	0.353	0.723
PUBI4	10.9632	16.6104	0.458	0.211	0.832	10.3731	16.3043	0.415	0.185	0.808
RELI1	13.8946	33.4016	0.780	0.692	0.909	14.6244	32.8985	0.695	0.635	0.879
RELI2	13.8701	32.1624	0.857	0.765	0.893	14.7363	31.1672	0.840	0.733	0.844
RELI3	13.7696	36.5021	0.644	0.495	0.934	15.0945	37.0833	0.528	0.402	0.912
RELI4	13.9265	32.7906	0.880	0.782	0.888	14.7139	31.5663	0.845	0.731	0.843
RELI5	13.8039	34.2022	0.843	0.713	0.897	14.8507	33.4590	0.793	0.650	0.857
RSKI1	9.6569	18.8009	0.463	0.218	0.800	9.9876	15.2991	0.417	0.214	0.646
RSKI2	9.8848	16.4511	0.650	0.481	0.709	10.4776	15.6267	0.443	0.315	0.631
RSKI3	9.9951	15.4496	0.714	0.543	0.673	10.5224	13.8761	0.546	0.371	0.563
RSKI4	9.9191	17.3915	0.568	0.339	0.751	10.2512	13.7896	0.459	0.238	0.623
SALI1	23.5000	56.0688	0.415	0.199	0.855	24.8856	66.1664	0.613	0.493	0.855
SALI2	23.7525	53.1106	0.542	0.348	0.841	25.4527	67.6000	0.621	0.474	0.854
SALI3	24.1716	50.2457	0.701	0.571	0.822	25.3905	63.0416	0.738	0.684	0.841
SALI4	24.1005	52.2282	0.618	0.500	0.832	25.4179	64.8274	0.718	0.674	0.844
SALI5	23.8971	52.1171	0.609	0.464	0.833	25.3657	66.6714	0.605	0.475	0.856
SALI6	23.6593	51.0606	0.636	0.464	0.830	25.2139	65.2409	0.659	0.541	0.850
SALI7	23.7304	53.0819	0.556	0.419	0.839	25.5746	72.0555	0.421	0.395	0.875
SALI8	23.8358	50.6192	0.664	0.516	0.826	25.5597	67.0101	0.620	0.551	0.854
SCII1	24.5539	50.5671	0.760	0.675	0.844	23.2040	61.7837	0.749	0.671	0.870
SCII2	24.8162	56.8728	0.416	0.289	0.889	23.5323	72.5189	0.405	0.199	0.906
SCII3	24.6078	48.6763	0.810	0.710	0.836	23.1891	59.1213	0.837	0.757	0.858
SCII4	24.5833	57.3444	0.462	0.303	0.881	23.0970	66.8559	0.548	0.354	0.893
SCII5	24.3848	50.6943	0.715	0.567	0.850	23.1269	61.0537	0.781	0.652	0.865

	Male Sample					Female Sample				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
SCII6	24.4779	52.4221	0.672	0.478	0.856	23.2886	62.3405	0.714	0.562	0.874
SCII7	24.5760	50.2841	0.783	0.625	0.841	23.2040	61.6092	0.801	0.688	0.864
SUPI1	25.0123	51.9286	0.484	0.334	0.812	25.0249	54.0143	0.578	0.445	0.782
SUPI2	24.6373	51.1162	0.533	0.379	0.806	24.7786	54.6167	0.542	0.433	0.787
SUPI3	25.0098	50.1866	0.619	0.440	0.795	25.3383	52.9875	0.674	0.525	0.768
SUPI4	25.1789	49.7984	0.589	0.384	0.798	25.3134	55.4227	0.554	0.349	0.786
SUPI5	24.8113	48.3992	0.545	0.327	0.805	24.9229	52.8394	0.590	0.413	0.780
SUPI6	25.1446	52.7481	0.454	0.252	0.816	25.4577	60.8423	0.318	0.174	0.818
SUPI7	24.8358	51.1695	0.521	0.327	0.807	25.4602	61.9199	0.276	0.183	0.823
SUPI8	24.9387	49.4926	0.633	0.472	0.792	25.1368	52.6870	0.690	0.572	0.766
WOODI1	13.9338	23.3298	0.751	0.627	0.802	15.3507	38.9964	0.717	0.669	0.712
WOODI2	13.9167	24.6319	0.702	0.573	0.816	15.3781	40.1110	0.669	0.585	0.726
WOODI3	14.1397	24.1057	0.667	0.457	0.825	15.3532	37.5507	0.664	0.537	0.718
WOODI4	14.0784	26.2887	0.548	0.303	0.855	15.4254	34.0206	0.370	0.142	0.876
WOODI5	13.9216	25.1044	0.680	0.471	0.822	15.5373	39.0023	0.674	0.513	0.721
WRTI1	17.8873	32.4787	0.622	0.442	0.808	17.2413	38.5576	0.683	0.519	0.795
WRTI2	18.0123	32.3954	0.695	0.583	0.793	17.1468	37.9111	0.606	0.468	0.812
WRTI3	17.8995	31.9039	0.677	0.539	0.796	17.5224	39.3374	0.657	0.552	0.801
WRTI4	17.9485	31.5821	0.708	0.535	0.789	17.3060	39.0009	0.680	0.544	0.796
WRTI5	17.9363	32.1384	0.729	0.610	0.786	17.2711	37.3453	0.780	0.634	0.776
WRTI6	17.4730	40.2794	0.251	0.095	0.871	16.8831	45.9240	0.300	0.167	0.868

APPENDIX 2: DESCRIPTIVE STATISTICS FOR MALE SAMPLE: INTEREST MODEL

	Item	ADEVI1	ADEVI2	ADEVI3	ADEVI4	ADVI1	ADVI2	ADVI3	ADVI4	ADVI5	ADVI6	ADVI7	ADVI8	ANI1	ANI2
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.956	3.772	3.576	3.711	3.620	3.534	3.304	3.336	3.500	3.392	3.429	3.627	3.360	3.336
	Std. Deviation	1.642	1.703	1.569	1.447	1.502	1.603	1.582	1.718	1.575	1.389	1.494	1.522	1.575	1.584
	Variance	2.696	2.899	2.461	2.093	2.256	2.569	2.502	2.951	2.482	1.929	2.231	2.318	2.482	2.509
	Skewness	-0.354*	-0.238	-0.043	-0.068	-0.064	-0.034	0.097	0.165	0.114	0.175	0.067	-0.034	0.035	0.115
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.087*	-1.177*	-1.091*	-0.939*	-1.033*	-1.187*	-1.135*	-1.257*	-1.082*	-0.726*	-0.976*	-1.044*	-1.041*	-1.005*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	ANI3	ANI4	ARTI1	ARTI2	ARTI3	ARTI4	ARTI5	ARTI6	ATHI1	ATHI2	ATHI3	ATHI4	ATHI5	ATHI6
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.380	3.363	3.618	3.630	3.539	3.615	3.461	3.453	3.468	3.539	3.326	3.542	3.392	3.662
	Std. Deviation	1.553	1.577	1.590	1.565	1.522	1.531	1.615	1.491	1.645	1.551	1.756	1.682	1.688	1.884
	Variance	2.413	2.487	2.527	2.450	2.318	2.345	2.608	2.224	2.707	2.406	3.085	2.829	2.848	3.551
	Skewness	0.117	0.022	-0.087	-0.024	-0.056	-0.056	0.057	0.149	-0.060	-0.024	0.167	-0.037	0.066	-0.144
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.014*	-1.090*	-1.116*	-1.082*	-1.051*	-1.033*	-1.101*	-0.943*	-1.181*	-1.044*	-1.339*	-1.202*	-1.284*	-1.475*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	ATHI7	CDEVI1	CDEVI2	CDEVI3	CDEVI4	CDEVI5	CDEVI6	CDEVI7	COUNI1	COUNI2	COUNI3	COUNI4	COUNI5	COUNI6
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.569	3.461	3.716	3.647	3.554	3.667	3.493	3.586	3.841	3.767	3.640	3.723	3.395	3.527
	Std. Deviation	1.418	1.724	1.596	1.746	1.520	1.574	1.789	1.454	1.613	1.590	1.452	1.641	1.543	1.477
	Variance	2.010	2.971	2.548	3.050	2.312	2.478	3.199	2.115	2.601	2.528	2.108	2.692	2.382	2.181
	Skewness	-0.054	0.041	-0.209	-0.108	0.036	-0.103	0.022	-0.058	-0.170	-0.056	-0.081	-0.089	0.070	-0.041
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.791*	-1.317*	-1.071*	-1.313*	-1.002*	-1.062*	-1.337*	-0.871*	-1.119*	-1.212*	-0.776*	-1.239*	-0.965*	-0.934*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	CULI1	CULI2	CULI3	CULI4	CULI5	CULI6	FARMI1	FARMI2	FARMI3	FARMI4	FARMI5	FARMI6	FASHI1	FASHI2
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.301	3.230	3.385	3.125	3.400	3.331	3.522	3.613	3.551	3.605	3.600	3.385	3.613	3.500
	Std. Deviation	1.594	1.530	1.559	1.532	1.440	1.569	1.624	1.476	1.576	1.580	1.585	1.740	1.673	1.603
	Variance	2.540	2.340	2.429	2.346	2.073	2.463	2.638	2.179	2.484	2.495	2.511	3.028	2.798	2.570
	Skewness	0.160	0.261	0.075	0.205	0.036	0.089	-0.026	-0.104	-0.059	-0.148	-0.159	0.043	-0.079	-0.007
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.081*	-0.974*	-1.127*	-0.938*	-0.887*	-1.051*	-1.145*	-0.966*	-1.141*	-1.042*	-1.057*	-1.286*	-1.258*	-1.143*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	FASHI3	FASHI4	FASHI5	FASHI6	FINI1	FINI2	FINI3	FINI4	FINI5	FINI6	FINI7	FINI8	INTI1	INTI2
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.488	3.424	3.216	3.446	3.576	3.593	3.402	3.485	3.554	3.498	3.645	3.355	3.495	3.446
	Std. Deviation	1.915	1.945	1.607	1.468	1.506	1.607	1.588	1.531	1.634	1.462	1.508	1.612	1.695	1.630
	Variance	3.666	3.783	2.582	2.154	2.269	2.581	2.521	2.344	2.670	2.138	2.274	2.598	2.875	2.656
	Skewness	0.049	0.035	0.170	0.032	-0.077	-0.101	0.051	0.032	-0.118	-0.062	-0.072	0.135	0.063	0.069
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.499*	-1.554*	-1.079*	-0.938*	-1.025*	-1.103*	-1.086*	-0.965*	-1.169*	-0.926*	-1.005*	-1.139*	-1.252*	-1.202*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	INTI3	INTI4	INTI5	LAWI1	LAWI2	LAWI3	LAWI4	LAWI5	LAWI6	LAWI7	LAWI8	LAWI9	LAWI10	LEADI1
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.500	3.664	3.635	3.647	3.772	3.520	3.588	3.463	3.571	3.466	3.571	3.495	3.711	3.603
	Std. Deviation	1.641	1.792	1.652	1.673	1.447	1.734	1.636	1.678	1.648	1.511	1.539	1.672	1.592	1.773
	Variance	2.693	3.211	2.729	2.799	2.093	3.007	2.675	2.814	2.717	2.284	2.368	2.796	2.535	3.144
	Skewness	0.005	-0.115	-0.091	-0.133	-0.184	-0.020	-0.088	-0.011	-0.128	0.019	-0.045	-0.014	-0.216	-0.071
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.177*	-1.396*	-1.217*	-1.151*	-0.881*	-1.341*	-1.163*	-1.260*	-1.148*	-0.932*	-1.052*	-1.231*	-1.014*	-1.372*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	LEADI2	LEADI3	LEADI4	LEADI5	LEADI6	MATHI1	MATHI2	MATHI3	MATHI4	MATHI5	MATHI6	MATHI7	MECHI1	MECHI2
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.956	3.493	3.632	3.625	3.838	4.074	4.115	4.147	4.002	4.277	4.208	4.150	3.569	3.855
	Std. Deviation	1.565	1.806	1.573	1.526	1.408	1.627	1.609	1.682	1.676	1.688	1.604	1.597	1.794	1.432
	Variance	2.450	3.263	2.474	2.328	1.984	2.648	2.589	2.828	2.808	2.849	2.573	2.550	3.219	2.050
	Skewness	-0.328*	0.026	-0.181	-0.076	-0.220	-0.387*	-0.533*	-0.548*	-0.382*	-0.620*	-0.522*	-0.497*	-0.060	-0.213
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.965*	-1.393*	-1.106*	-1.031*	-0.815*	-1.087*	-0.831*	-0.954*	-1.096*	-0.929*	-0.859*	-0.917*	-1.381*	-0.913*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	MECHI3	MECHI4	MECHI5	MECHI6	MECHI7	MECHI8	MECHI9	MEDI1	MEDI2	MEDI3	MEDI4	MEDI5	MEDI6	MEDI7
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.975	3.745	3.882	4.042	3.422	3.944	3.792	4.037	3.713	3.547	3.760	3.711	3.794	3.743
	Std. Deviation	1.676	1.523	1.518	1.727	1.587	1.589	1.553	1.637	1.720	1.554	1.678	1.500	1.532	1.447
	Variance	2.810	2.318	2.306	2.984	2.520	2.525	2.411	2.679	2.957	2.416	2.817	2.250	2.346	2.093
	Skewness	-0.342*	-0.151	-0.223	-0.389	0.027	-0.280*	-0.255*	-0.390*	-0.214	-0.022	-0.165	-0.139	-0.170	-0.194
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.163*	-0.941*	-1.042*	-1.165*	-1.038*	-1.114*	-0.983*	-1.085*	-1.238*	-1.118*	-1.197*	-0.905*	-0.979*	-0.745*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	MEDI8	MEDI9	MEDI10	MEDI11	MEDI12	MILI1	MILI2	MILI3	MILI4	MILI5	MILI6	MILI7	OPI1	OPI2
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	4.091	3.657	3.806	3.713	3.456	3.434	3.392	3.441	3.395	3.392	3.321	3.350	3.571	3.544
	Std. Deviation	1.633	1.629	1.594	1.569	1.398	1.656	1.643	1.653	1.543	1.599	1.631	1.614	1.455	1.666
	Variance	2.667	2.653	2.540	2.461	1.954	2.743	2.701	2.734	2.382	2.558	2.661	2.607	2.118	2.774
	Skewness	-0.449*	-0.155	-0.233	-0.124	0.099	0.032	0.107	-0.001	0.030	0.052	0.132	0.180	-0.091	-0.010
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.948*	-1.106*	-1.045*	-1.093*	-0.768*	-1.269*	-1.171*	-1.200*	-0.966*	-1.103*	-1.079*	-1.106*	-0.878*	-1.151*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	OPI3	OPI4	OPI5	OPI6	OPI7	OPI8	PERI1	PERI2	PERI3	PERI4	PERI5	PERI6	PERI7	PERI8
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.434	3.591	3.463	3.787	3.483	3.507	3.699	3.463	3.743	3.507	3.221	3.397	3.544	3.363
	Std. Deviation	1.495	1.437	1.669	1.422	1.615	1.805	1.662	1.707	1.578	1.727	1.711	1.651	1.677	1.652
	Variance	2.236	2.065	2.785	2.021	2.609	3.258	2.761	2.913	2.491	2.983	2.929	2.726	2.814	2.728
	Skewness	0.073	-0.031	0.039	-0.094	0.003	0.006	-0.140	0.036	-0.139	-0.004	0.212	0.067	-0.017	0.105
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.934*	-0.775*	-1.196*	-0.798*	-1.106*	-1.401*	-1.298*	-1.259*	-1.118*	-1.301*	-1.204*	-1.165*	-1.250*	-1.204*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	PERI9	PERI10	PNTI1	PNTI2	PNTI3	PNTI4	PNTI5	PUBI1	PUBI2	PUBI3	PUBI4	RELI1	RELI2	RELI3
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.422	3.426	3.574	3.576	3.640	3.549	3.566	3.691	3.782	3.490	3.512	3.422	3.446	3.547
	Std. Deviation	1.632	1.612	1.537	1.460	1.513	1.579	1.484	1.671	1.604	1.426	1.606	1.717	1.719	1.628
	Variance	2.662	2.599	2.363	2.132	2.290	2.494	2.202	2.794	2.574	2.034	2.580	2.947	2.955	2.651
	Skewness	0.065	0.071	-0.039	-0.084	-0.049	-0.036	-0.039	-0.160	-0.254*	-0.054	0.028	0.075	0.058	-0.089
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.085*	-1.154*	-0.940*	-0.940*	-1.044*	-1.039*	-0.929*	-1.215*	-1.089*	-0.859*	-1.155*	-1.301*	-1.302*	-1.187*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	RELI4	RELI5	RSKI1	RSKI2	RSKI3	RSKI4	SALI1	SALI2	SALI3	SALI4	SALI5	SALI6	SALI7	SALI8
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.390	3.512	3.495	3.267	3.157	3.233	3.735	3.483	3.064	3.135	3.338	3.576	3.505	3.400
	Std. Deviation	1.627	1.545	1.666	1.695	1.743	1.692	1.421	1.474	1.462	1.419	1.446	1.498	1.449	1.490
	Variance	2.646	2.388	2.776	2.874	3.037	2.862	2.018	2.172	2.138	2.014	2.092	2.245	2.098	2.221
	Skewness	0.061	0.039	-0.004	0.139	0.169	0.249	-0.098	-0.001	0.353*	0.274*	0.080	0.013	0.010	0.100
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-1.152*	-1.033*	-1.282*	-1.274*	-1.326*	-1.213*	-0.871*	-0.961*	-0.762*	-0.659*	-0.857*	-0.905*	-0.883*	-0.899*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	SCII1	SCII2	SCII3	SCII4	SCII5	SCII6	SCII7	SUPI1	SUPI2	SUPI3	SUPI4	SUPI5	SUPI6	SUPI7
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	4.113	3.850	4.059	4.083	4.282	4.189	4.091	3.498	3.873	3.500	3.331	3.699	3.365	3.674
	Std. Deviation	1.562	1.642	1.638	1.480	1.625	1.547	1.548	1.484	1.470	1.404	1.496	1.727	1.456	1.488
	Variance	2.439	2.697	2.684	2.190	2.640	2.394	2.397	2.201	2.161	1.971	2.237	2.983	2.119	2.215
	Skewness	-0.383*	-0.230	-0.408*	-0.473*	-0.607*	-0.502*	-0.420*	0.043	-0.287*	-0.043	0.117	-0.177	0.120	-0.094
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.965*	-1.162*	-1.044*	-0.684*	-0.805*	-0.880*	-0.910*	-0.987*	-0.869*	-0.865*	-0.972*	-1.261*	-0.910*	-1.054*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

	Item	SUPI8	WOODI1	WOODI2	WOODI3	WOODI4	WOODI5	WRTI1	WRTI2	WRTI3	WRTI4	WRTI5	WRTI6
N	Valid	408	408	408	408	408	408	408	408	408	408	408	408
	Mean	3.571	3.564	3.581	3.358	3.419	3.576	3.544	3.419	3.532	3.483	3.495	3.958
	Std. Deviation	1.447	1.574	1.490	1.609	1.530	1.465	1.627	1.513	1.597	1.581	1.489	1.393
	Variance	2.093	2.478	2.219	2.589	2.342	2.147	2.647	2.288	2.549	2.501	2.216	1.942
	Skewness	-0.025	-0.152	-0.046	0.074	0.086	0.026	-0.069	0.100	-0.040	0.028	0.014	-0.347*
	Std. Error of Skewness	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121	0.121
	Kurtosis	-0.896*	-1.090*	-0.967*	-1.088*	-1.101*	-0.983*	-1.168*	-0.987*	-1.101*	-1.027*	-0.995*	-0.618*
	Std. Error of Kurtosis	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241	0.241

* significant ($p < 0.05$)

APPENDIX 3: DESCRIPTIVE STATISTICS FOR FEMALE SAMPLE: INTEREST MODEL

	Item	ADEVI1	ADEVI2	ADEVI3	ADEVI4	ADVI1	ADVI2	ADVI3	ADVI4	ADVI5	ADVI6	ADVI7	ADVI8
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.731	3.900	3.512	3.438	3.530	3.381	3.092	3.179	3.348	3.114	3.271	3.507
	Std. Deviation	1.537	1.640	1.595	1.492	1.416	1.535	1.542	1.591	1.530	1.446	1.471	1.518
	Variance	2.362	2.688	2.545	2.227	2.005	2.356	2.378	2.531	2.342	2.092	2.163	2.305
	Skewness	-0.122	-0.218	0.062	0.092	-0.029	0.175	0.362*	0.264*	0.117	0.241*	0.285*	-0.049
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.034*	-1.177*	-1.103*	-0.999*	-0.766*	-0.995*	-0.855*	-0.975*	-1.038*	-0.837*	-0.783*	-0.977*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	ANI1	ANI2	ANI3	ANI4	ARTI1	ARTI2	ARTI3	ARTI4	ARTI5	ARTI6	ATHI1	ATHI2
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.654	3.697	3.682	3.769	3.333	3.306	2.990	3.458	3.308	3.453	3.639	3.614
	Std. Deviation	1.753	1.739	1.761	1.725	1.457	1.572	1.609	1.541	1.581	1.526	1.637	1.545
	Variance	3.075	3.025	3.100	2.976	2.123	2.472	2.588	2.374	2.498	2.328	2.680	2.387
	Skewness	-0.123	-0.095	-0.147	-0.142	0.175	0.147	0.377*	0.080	0.165	0.173	-0.138	-0.071
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.297*	-1.294*	-1.323*	-1.306*	-0.954*	-1.100*	-1.043*	-1.004*	-1.022*	-0.990*	-1.139*	-1.000*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	ATHI3	ATHI4	ATHI5	ATHI6	ATHI7	CDEVI1	CDEVI2	CDEVI3	CDEVI4	CDEVI5	CDEVI6	CDEVI7
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.684	3.677	3.751	3.177	3.714	3.552	3.512	3.522	3.562	3.403	3.515	3.381
	Std. Deviation	1.849	1.758	1.826	1.747	1.453	1.638	1.554	1.656	1.592	1.616	1.667	1.554
	Variance	3.419	3.092	3.334	3.054	2.110	2.682	2.415	2.744	2.536	2.610	2.779	2.416
	Skewness	-0.123	-0.146	-0.218	0.224	-0.130	0.040	0.017	0.060	-0.049	0.117	-0.012	0.098
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.418*	-1.290*	-1.370*	-1.303*	-0.845*	-1.170*	-1.050*	-1.202*	-1.091*	-1.108*	-1.177*	-0.987*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	COUNI1	COUNI2	COUNI3	COUNI4	COUNI5	COUNI6	CULI1	CULI2	CULI3	CULI4	CULI5	CULI6
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.400	3.346	3.604	3.284	3.490	3.495	3.440	3.259	3.199	3.127	3.095	3.366
	Std. Deviation	1.614	1.751	1.533	1.775	1.580	1.500	1.577	1.617	1.520	1.609	1.579	1.602
	Variance	2.605	3.065	2.349	3.151	2.495	2.251	2.486	2.616	2.309	2.590	2.495	2.567
	Skewness	0.023	0.154	-0.067	0.173	0.034	0.016	0.070	0.321*	0.234	0.298*	0.321*	0.044
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.170*	-1.317*	-0.972*	-1.339*	-1.091*	-1.072*	-1.099*	-1.072*	-0.902*	-1.022*	-1.020*	-1.118*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	FARMI1	FARMI2	FARMI3	FARMI4	FARMI5	FARMI6	FASHI1	FASHI2	FASHI3	FASHI4	FASHI5	FASHI6
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.861	4.000	3.891	4.007	3.983	3.756	3.231	3.104	3.425	3.303	3.077	3.393
	Std. Deviation	1.764	1.739	1.705	1.585	1.701	1.894	1.654	1.633	1.653	1.559	1.505	1.432
	Variance	3.113	3.022	2.906	2.511	2.895	3.586	2.737	2.667	2.734	2.431	2.266	2.050
	Skewness	-0.242*	-0.395*	-0.269*	-0.432*	-0.351*	-0.120	0.175	0.290*	0.074	0.145	0.269*	0.127
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.280*	-1.224*	-1.187*	-0.929*	-1.143*	-1.481*	-1.220*	-1.165*	-1.201*	-1.034*	-0.959*	-0.839*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	FINI1	FINI2	FINI3	FINI4	FINI5	FINI6	FINI7	FINI8	INTI1	INTI2	INTI3	INTI4
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.749	3.644	3.898	3.659	3.806	3.420	3.595	3.604	3.600	3.371	3.311	3.149
	Std. Deviation	1.702	1.740	1.798	1.743	1.783	1.579	1.637	1.707	1.632	1.693	1.668	1.836
	Variance	2.897	3.028	3.234	3.038	3.179	2.494	2.681	2.913	2.665	2.867	2.783	3.369
	Skewness	-0.164	-0.071	-0.333*	-0.090	-0.262*	0.088	-0.092	-0.059	-0.047	0.097	0.163	0.294*
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.194*	-1.292*	-1.258*	-1.312*	-1.312*	-1.047*	-1.145*	-1.250*	-1.149*	-1.299*	-1.189*	-1.352*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	INTI5	LAWI1	LAWI2	LAWI3	LAWI4	LAWI5	LAWI6	LAWI7	LAWI8	LAWI9	LAWI10	LEADI1
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.425	3.624	3.794	3.624	3.883	3.597	3.915	3.592	3.550	3.759	3.754	3.286
	Std. Deviation	1.724	1.643	1.539	1.799	1.712	1.779	1.825	1.700	1.464	1.879	1.640	1.686
	Variance	2.973	2.699	2.369	3.238	2.931	3.164	3.329	2.891	2.143	3.530	2.690	2.843
	Skewness	0.064	-0.045	-0.162	-0.028	-0.282*	-0.053	-0.297*	-0.038	-0.031	-0.210	-0.193	0.240*
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.284*	-1.151*	-0.986*	-1.413*	-1.229*	-1.389*	-1.342	-1.231*	-0.973*	-1.443*	-1.137*	-1.170*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	LEADI2	LEADI3	LEADI4	LEADI5	LEADI6	MATHI1	MATHI2	MATHI3	MATHI4	MATHI5	MATHI6	MATHI7
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.945	3.080	3.194	3.393	3.460	4.037	4.025	3.978	4.047	3.886	3.955	3.871
	Std. Deviation	1.685	1.662	1.569	1.526	1.398	1.635	1.738	1.885	1.615	1.870	1.693	1.799
	Variance	2.840	2.762	2.461	2.329	1.955	2.674	3.022	3.553	2.609	3.498	2.866	3.235
	Skewness	-0.319*	0.351*	0.209	0.147	0.099	-0.404*	-0.336*	-0.392*	-0.395*	-0.311*	-0.277*	-0.255*
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.135*	-1.079*	-1.072*	-1.036*	-0.833*	-0.986*	-1.225*	-1.334*	-1.007*	-1.361*	-1.200*	-1.379*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	MECHI1	MECHI2	MECHI3	MECHI4	MECHI5	MECHI6	MECHI7	MECHI8	MECHI9	MEDI1	MEDI2	MEDI3
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	4.127	4.192	3.729	4.080	4.047	3.953	3.990	3.905	4.080	3.876	3.811	3.398
	Std. Deviation	1.908	1.774	1.639	1.807	1.673	1.794	1.925	1.760	1.800	1.771	1.606	1.497
	Variance	3.642	3.148	2.687	3.265	2.798	3.217	3.706	3.098	3.241	3.137	2.578	2.240
	Skewness	-0.496*	-0.526*	-0.153	-0.420*	-0.454*	-0.309*	-0.397*	-0.263*	-0.455*	-0.225	-0.169	0.102
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.274*	-1.155*	-1.179*	-1.258*	-1.038*	-1.353*	-1.375*	-1.281*	-1.187*	-1.306*	-1.133*	-0.934*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	MEDI4	MEDI5	MEDI6	MEDI7	MEDI8	MEDI9	MEDI10	MEDI11	MEDI12	MILI1	MILI2	MILI3
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.687	3.450	3.679	3.470	3.776	3.776	3.751	3.595	3.590	3.291	4.002	3.883
	Std. Deviation	1.759	1.647	1.598	1.589	1.737	1.552	1.547	1.581	1.448	1.697	1.933	1.900
	Variance	3.094	2.712	2.553	2.524	3.017	2.409	2.392	2.501	2.098	2.880	3.738	3.610
	Skewness	-0.109	0.095	-0.092	0.107	-0.189	-0.141	-0.123	-0.020	-0.051	0.212	-0.384*	-0.283*
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.322*	-1.167*	-1.113*	-1.071*	-1.249*	-1.016*	-1.063*	-1.141*	-0.845*	-1.240*	-1.412*	-1.440*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	MILI4	MILI5	MILI6	MILI7	OPI1	OPI2	OPI3	OPI4	OPI5	OPI6	OPI7	OPI8
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.689	4.005	3.973	3.585	3.473	3.604	3.644	3.455	3.206	3.654	3.595	3.590
	Std. Deviation	1.653	1.870	1.960	1.796	1.453	1.601	1.795	1.549	1.584	1.561	1.525	1.671
	Variance	2.733	3.496	3.842	3.226	2.110	2.564	3.222	2.398	2.508	2.436	2.326	2.791
	Skewness	-0.055	-0.401*	-0.369*	-0.009	0.091	-0.037	-0.127	0.034	0.210	-0.065	-0.138	-0.042
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.181*	-1.304*	-1.443*	-1.385*	-0.811*	-1.041*	-1.335*	-0.987*	-0.991*	-1.087*	-0.911*	-1.194*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	PERI1	PERI2	PERI3	PERI4	PERI5	PERI6	PERI7	PERI8	PERI9	PERI10	PNTI1	PNTI2
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.371	3.236	3.512	3.465	3.883	3.512	3.781	3.331	3.448	3.418	3.801	3.423
	Std. Deviation	1.661	1.719	1.629	1.689	1.855	1.750	1.740	1.667	1.654	1.629	1.626	1.565
	Variance	2.758	2.954	2.654	2.853	3.440	3.063	3.029	2.781	2.737	2.653	2.644	2.449
	Skewness	0.126	0.166	0.019	0.073	-0.271	0.013	-0.174	0.114	0.071	0.090	-0.218	0.097
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.202*	-1.247*	-1.157*	-1.214*	-1.367*	-1.324*	-1.285*	-1.137*	-1.138*	-1.167*	-1.080*	-1.042*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	PNTI3	PNTI4	PNTI5	PUBI1	PUBI2	PUBI3	PUBI4	RELI1	RELI2	RELI3	RELI4	RELI5
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.826	3.858	3.672	3.393	3.580	3.400	3.843	3.881	3.769	3.410	3.791	3.654
	Std. Deviation	1.652	1.658	1.584	1.731	1.587	1.415	1.758	1.806	1.744	1.665	1.697	1.589
	Variance	2.728	2.750	2.510	2.997	2.519	2.001	3.090	3.263	3.041	2.771	2.879	2.526
	Skewness	-0.207	-0.238	-0.035	0.093	0.004	0.010	-0.233	-0.301*	-0.155	0.146	-0.233	-0.038
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.194*	-1.127*	-1.106*	-1.290*	-1.065*	-0.781*	-1.281*	-1.293*	-1.274*	-1.192*	-1.184*	-1.065*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	RSKI1	RSKI2	RSKI3	RSKI4	SALI1	SALI2	SALI3	SALI4	SALI5	SALI6	SALI7	SALI8
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.759	3.269	3.224	3.495	4.095	3.527	3.590	3.562	3.614	3.766	3.405	3.420
	Std. Deviation	1.671	1.562	1.682	1.854	1.652	1.515	1.668	1.566	1.625	1.638	1.533	1.566
	Variance	2.792	2.441	2.828	3.438	2.729	2.295	2.781	2.451	2.641	2.683	2.351	2.454
	Skewness	-0.117	0.155	0.251*	0.016	-0.468*	-0.061	-0.104	0.014	-0.094	-0.188	0.102	0.037
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.249*	-1.007*	-1.130*	-1.438*	-0.934*	-0.950*	-1.195*	-1.011*	-1.124*	-1.072*	-0.964*	-1.060*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	SCII1	SCII2	SCII3	SCII4	SCII5	SCII6	SCII7	SUPI1	SUPI2	SUPI3	SUPI4	SUPI5
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.903	3.575	3.918	4.010	3.980	3.818	3.903	3.751	3.998	3.438	3.463	3.853
	Std. Deviation	1.725	1.495	1.771	1.707	1.724	1.744	1.648	1.654	1.670	1.566	1.566	1.740
	Variance	2.976	2.235	3.138	2.913	2.972	3.042	2.716	2.736	2.791	2.451	2.454	3.028
	Skewness	-0.230	0.006	-0.213	-0.463*	-0.330*	-0.207	-0.277*	-0.190	-0.348*	0.018	0.050	-0.224
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-1.270*	-0.969*	-1.385*	-1.045*	-1.188*	-1.302*	-1.150*	-1.176*	-1.136*	-1.087*	-1.072*	-1.288*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	SUPI6	SUPI7	SUPI8	WOODI1	WOODI2	WOODI3	WOODI4	WOODI5	WRTI1	WRTI2	WRTI3	WRTI4
N	Valid	402	402	402	402	402	402	402	402	402	402	402	402
	Mean	3.318	3.316	3.639	3.910	3.883	3.908	3.836	3.724	3.433	3.527	3.152	3.368
	Std. Deviation	1.532	1.514	1.564	1.633	1.604	1.872	3.029	1.707	1.654	1.866	1.621	1.615
	Variance	2.347	2.291	2.446	2.665	2.572	3.505	9.175	2.914	2.735	3.482	2.628	2.607
	Skewness	0.190	0.112	-0.130	-0.242*	-0.271*	-0.333*	8.074*	-0.141	0.041	2.440*	0.215	0.060
	Std. Error of Skewness	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122
	Kurtosis	-0.922*	-1.025*	-1.068*	-1.190*	-1.072*	-1.345*	96.276*	-1.289*	-1.206*	22.350*	-1.092*	-1.120*
	Std. Error of Kurtosis	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428	0.2428

	Item	WRTI5	WRTI6
N	Valid	402	402
	Mean	3.403	3.791
	Std. Deviation	1.613	1.648
	Variance	2.600	2.714
	Skewness	0.106	-0.254*
	Std. Error of Skewness	0.122	0.122
	Kurtosis	-1.136*	-1.098*
	Std. Error of Kurtosis	0.2428	0.2428

* significant ($p < 0.05$)