

MEASUREMENT INVARIANCE OF THE PCQ-24 ACROSS ETHNIC GROUPS IN SOUTH AFRICA

Nichále Esterhuizen

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



Supervisor: Prof G Görgens

March 2021

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 sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2021

ABSTRACT

In South Africa, psychological assessment is governed by strict legal parameters and practitioners are required to proactively show that their measures are valid, reliable, unbiased and can be fairly applied across groups (Republic of South Africa, 1998, p. 7). Despite the increased research interest regarding culturally appropriate measurement practices in the field, group comparisons are routinely reported in literature and presented as true differences in the construct of interest, without considering, and testing for the possibility of bias in assessment. This is highly concerning as if practitioners are not cognisant of potential biases in measurement, research findings can be misleading with severe impacts on individuals as well as the organisation (Steenkamp & Baumgartner, 1998).

Measurement invariance is introduced as a rigorous manner of testing for measurement bias and meeting the requirements of the Employment Equity Act. In particular, the study aims to investigate the measurement invariance and equivalence of the Psychological Capital Scale (PCQ-24). The PCQ-24 is a well-researched positive psychological construct which has been demonstrated to be valid and reliable in the South African context (e.g. Bernstein & Volpe, 2016; Görgens-Ekermans & Herbert, 2013; Roemer & Harris, 2018; Van Wyk, 2016). To date however, to the knowledge of the researcher and her supervisor, no measurement invariance study has been conducted on the PCQ-24 across ethnic groups in the South African context. The study therefore aimed to evaluate measurement invariance and equivalence according to the taxonomy proposed by Dunbar, Spangenberg and Theron (2011) by testing the significance of difference in fit between subsequent measurement models with increasing constraints placed on the model. The analyses were conducted on archival data which was obtained from previous Masters studies which involved Psychological Capital (PsyCap). The data obtained was sufficient to perform the measurement invariance analyses over Black, White and Coloured ethnic groups.

According to the Dunbar et al. (2011) methodology, the PCQ-24 showed evidence of a lack of construct bias and a lack of non-uniform bias. This indicated that the PCQ-24 measured the same underlying construct across the Black, White and Coloured groups. There was vast evidence however, of uniform bias as well as error variance bias toward the Black and Coloured groups (Dunbar et al., 2011; Wu, Li & Zumbo, 2007). This suggested a group membership main effect, i.e. group membership explained significant variance in item responses, not explained by the latent variable (Fontaine, 2008). Furthermore, the presence of error variance bias in some of the items suggested that there was systematic residual variance influencing respondents' item scores, not attributed to the underlying variable (Wu et al., 2007). In line with previous PsyCap research, the negatively keyed items in the scale proved to be problematic throughout the analyses. Notwithstanding, the results revealed a

potentially more salient problem, specifically for the Black sample. Consequently, the PCQ-24 demonstrated metric, partial scalar, partial conditional probability equivalence. The study offers potential explanations for the sources of bias prevalent in the study, as well as recommendations for future practitioners utilising the PCQ-24. Finally, a contribution of the study is that it provides a recommendation for enhancement of the Dunbar et al. (2011) taxonomy for future measurement invariance studies.

OPSOMMING

In Suid-Afrika word sielkundige assessering deur streng regsparameters beheer en daar word van praktisyns verwag om proaktief aan te toon dat hul maatreëls geldig, betroubaar, onbevooroordeeld is en regverdig toegepas kan word in groepe (Republiek van Suid-Afrika, 2014, p. 7). Ondanks die verhoogde navorsingsbelangstelling met betrekking tot kultuurtoepaslike metingspraktyke in die veld, word groepvergelings gereeld in die literatuur gerapporteer en as ware verskille in die struktuur van belangstelling voorgestel, sonder om die moontlikheid van sydigheid in assessering te oorweeg. Dit is baie kommerwekkend, aangesien praktisyns nie kennis dra van potensiële sydighede in meting nie en kan navorsingsbevindinge misleidend wees met ernstige impak op individue, sowel as die organisasie (Steenkamp & Baumgartner, 1998).

Meting-invariansie word ingestel as 'n noukeurige manier om metingssydigheid te toets en aan die vereistes van die Wet op Billike Indiensneming te voldoen. In die besonder beoog die studie om die meting-invariansie en meting-ekwivalensie van die Psychological Capital Scale (PCQ-24) te ondersoek. Die PCQ-24 is 'n goed nagevorsde, positiewe sielkundige konstruk wat bewys is dat dit geldig en betroubaar is in die Suid-Afrikaanse konteks (bv. Bernstein & Volpe, 2016; Görgens-Ekermans & Herbert, 2013; Roemer & Harris, 2018; Van Wyk, 2016). Tot die kennis van die navorser en haar studieleier is daar tot dusver nog geen meting-invariansie-studie gedoen oor die PCQ-24 in rasgroepe in die Suid-Afrikaanse konteks nie. Die studie het dus ten doel gehad om meting-invariansie en ekwivalensie te evalueer volgens die taksonomie voorgestel deur Dunbar, Spangenberg en Theron (2011) deur die belangrikheid van die verskil in pas tussen die daaropvolgende meetmodelle met toenemende beperkings op die model te toets. Die ontledings is gedoen op argiefdata wat verkry is uit vorige meestersgraadstudies waarby Psychological Capital (PsyCap) betrokke was. Die data wat verkry is, was voldoende om die meting-invariansie-ontledings oor swart, wit en gekleurde rasgroepe uit te voer.

Volgens die Dunbar et al. (2011) se metodiek het die PCQ-24 bewys gelever van 'n gebrek aan konstruksydigheid en 'n gebrek aan nie-eenvormige sydigheid. Dit het aangedui dat die PCQ-24 dieselfde onderliggende konstruk gemeet het in die swart, wit en gekleurde groepe. Daar was egter groot bewyse van eenvormige sydigheid sowel as foutafwykingsydigheid teenoor die swart en gekleurde groepe (Dunbar et al., 2011; Wu, Li & Zumbo, 2007). Dit het 'n hoofeffek van die groeplidmaatskap voorgestel, in ander woorde, het die groeplidmaatskap 'n beduidende variansie in itemresponse verklaar, nie verduidelik deur die latente veranderlike nie (Fontaine, 2008). Verder het die teenwoordigheid van foutafwykingsydigheid in sommige van die items voorgestel dat daar stelselmatige oorblywende variansie was, wat respondente se artikeltellings beïnvloed, nie toegeskryf

aan die onderliggende veranderlike nie (Wu et al., 2007). In ooreenstemming met vorige PsyCap-navorsing was die negatiewe sleutelitems in die skaal tydens die ontledings problematies. Desondanks het die resultate 'n meer opvallende probleem aan die lig gebring, spesifiek vir die swart groep. Gevolglik het die PCQ-24 metrieke, gedeeltelike skalêre, gedeeltelike voorwaardelike waarskynlikheids-ekwivalensie getoon. Die studie bied moontlike verklarings vir die bronne van sydigheid wat in die studie voorkom, asook aanbevelings vir toekomstige praktisyns wat die PCQ-24 gebruik. Laastens is die bydrae van die studie dat dit 'n aanbeveling gee vir die verbetering van die Dunbar et al. (2011) taksonomie vir toekomstige meting-invariansie studies.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my incredible supervisor, Prof Gina Görgens, for supporting me on this journey. Your patience, dedication and willingness to assist is unwavering and I am so grateful to have had you as my supervisor. You helped to push me to do my best and this would not have been possible without you, sincerely thank you!

Secondly, I would like to thank Prof Callie Theron for your guidance and support in conducting the analyses for this study. Your genuine interest in this project has been inspiring and I thank you for all your contributions!

To my amazing support system, my family and friends, thank you all for encouraging and motivating me to continue. All the love and support I received throughout this journey is so appreciated!

Lastly to my amazing boyfriend Nic Labuschagne, thank you for being my Number 1 supporter and for helping me through the early mornings and late nights. You helped me to push through when I needed it most and for that I am so grateful.

None of this would have been possible without all the support I received, so thank you!

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Research Initiating Question.....	8
1.3 Research Objectives.....	8
1.4 Overview of the Study.....	8
CHAPTER 2: LITERATURE REVIEW	9
2.1 Introduction	9
2.2 PCQ-24	9
2.2.1 Latent Dimensions Comprising the PsyCap Construct	10
2.2.2 Psychometric Properties of the PCQ-24	13
2.2.3 Current Research Examining Group Differences in Psychological Capital	26
2.3 Bias, Measurement Invariance and Measurement Equivalence	29
2.3.1 Measurement.....	29
2.3.2 Cross-cultural Measurement	29
2.3.3 Bias in Measurement	29
2.3.4 Measurement Invariance/Equivalence (MI/ME)	33
2.3.5 Taxonomy of Measurement Invariance and Equivalence.....	36
2.3.6 Assessing Measurement Invariance and Measurement Equivalence	38
2.4 Conclusion.....	41
CHAPTER 3: RESEARCH METHODOLOGY	43
3.1 Introduction	43
3.2 Substantive Research Hypothesis	43
3.3 Research Design.....	44
3.3.1 Evaluating Measurement Model Fit.....	46
3.4 Statistical Hypotheses	49
3.4.1 Tests of Practical Significance	51

3.5	Sampling.....	53
3.5.1	Archival Dataset Number One	54
3.5.2	Archival Dataset Number Two	54
3.5.3	Archival Dataset Number Three.....	55
3.5.4	Critical Reflection on using Archival Data	56
3.5.5	Combined Dataset.....	57
3.6	Statistical Analyses.....	59
3.6.1	Preparatory Procedures.....	59
3.6.2	Evaluation of the PsyCap Measurement Model	73
3.7	Ethical Considerations.....	79
3.7.1	Risk-Benefit Analysis	80
3.7.2	Ethical Risks, Guiding Principles and Legislation.....	80
3.8	Summary	82
CHAPTER 4: RESULTS.....		83
4.1	Introduction	83
4.2	Assessing the PCQ-24 Measurement Model	83
4.3	Assessing the PCQ-24 Single-Group Measurement Models via CFA	84
4.3.1	Measurement Model Fit Indices	85
4.3.2	Measurement Model Residuals	87
4.3.3	Parameter Estimates for the Single-group Measurement Models.....	93
4.4	Evaluating the PCQ-24 Multigroup MI and ME.....	104
4.4.1	Configural Invariance	106
4.4.2	Weak Invariance.....	108
4.4.3	Metric Equivalence	109
4.4.4	Strong Invariance	111
4.4.5	Partial Strong Invariance.....	112
4.4.6	Partial Scalar Equivalence	114
4.4.7	Strict Invariance	117

4.4.8	Conditional Probability Equivalence	118
4.4.9	Partial Conditional Probability Equivalence	119
4.3	Assessing the PCQ-24 Measurement Model at the Dimension Level.....	121
4.9	Conclusion.....	126
CHAPTER 5: DISCUSSION.....		127
5.1	Introduction	127
5.2	Summary of Key Findings.....	129
5.3	Implications and Recommendations for Future Research.....	137
5.4	Limitations of the Study	143
5.5	Conclusion.....	145
REFERENCE LIST		146

LIST OF APPENDICES

	Page
Appendix A: Goodness of fit statistics for the PCQ-24 multigroup configural invariance measurement model	166
Appendix B: Goodness of fit statistics for the PCQ-24 multigroup weak invariance measurement model	167
Appendix C: Goodness of fit statistics for the PCQ-24 multigroup strong invariance measurement model	168
Appendix D: Goodness of fit statistics for the PCQ-24 multigroup partial strong invariance measurement model	169
Appendix E: Goodness of fit statistics for the PCQ-24 multigroup strict invariance measurement model	170
Appendix F: Path diagram for the configural invariance measurement model: Black group	171
Appendix G: Path diagram for the configural invariance measurement model: White group	172
Appendix H: Path diagram for the configural invariance measurement model: Coloured group	173
Appendix I: LISREL syntax for the classically parallel PsyCap configural invariance multigroup measurement model	174
Appendix J: Ethical Clearance Letter (REC)	175

LIST OF FIGURES

		Page
Figure 3.1	Analysis performed in testing MI: Operational hypotheses 1 – 5	43
Figure 3.2	Ex post facto correlational design	46
Figure 4.1	Stem-and-leaf plot of standardised residuals for the Black measurement model	88
Figure 4.2	Q-plot of standardised residuals for the Black measurement model	89
Figure 4.3	Stem-and-leaf plot of standardised residuals for the White measurement model	90
Figure 4.4	Q-plot of standardised residuals for the White measurement model	91
Figure 4.5	Stem-and-leaf plot of standardised residuals for the Coloured measurement Model	92
Figure 4.6	Q-plot of standardised residuals for the Coloured measurement model	93

LIST OF TABLES

		Page
Table 1.1	Taxonomy for the interpretation of the magnitude of Cronbach alpha internal consistency reliability coefficients	20
Table 2.1	Degrees of MI	36
Table 2.2	Levels of ME	37
Table 3.1	Analyses performed in testing ME: Operational hypotheses 6 -8	44
Table 3.2	Criteria to evaluate the practical significance of the difference in fit between nested models	51
Table 3.3	Ethnic group by gender frequency distributions	58
Table 3.4	Frequency distribution of respondents by Age	58
Table 3.5	Frequency distribution of respondents by Home language	59
Table 3.6	Degrees of Freedom for the Single-Group and series of Multi-Group MI Models	62
Table 3.7	Preacher and Coffman Statistical Power Calculations	63
Table 3.8	The means, standard deviations and reliability statistics for the PCQ-24 subscales –Black group	67
Table 3.9	The means, standard deviations and reliability statistics for the PCQ-24 subscales – White group	69
Table 3.10	The means, standard deviations and reliability statistics for the PCQ24 subscales – Coloured group	71
Table 3.11	The means, standard deviations and reliability statistics for the PCQ24 subscales – Combined sample	72
Table 3.12	Steps in conducting analyses of MI and ME	75
Table 3.13	GOF Statistics to judge model fit in MI analyses	75
Table 4.1	Test of multivariate normality for continuous variables	84
Table 4.2	Summary of GOF indices: Black sample	85
Table 4.3	Summary of GOF Indices: White sample	86
Table 4.4	Summary of GOF Indices: Coloured sample	86
Table 4.5	Summary: Standardised residuals for Black measurement model	88
Table 4.6	Summary: Standardised residuals for White measurement model	90
Table 4.7	Summary: Standardised residuals for Coloured measurement model	92
Table 4.8	Item content of Resilience subscale: Significant negative factor loadings	95
Table 4.9	Unstandardized lambda-X matrix for the Black measurement model	95
Table 4.10	Completely standardised lambda-X matrix for the Black measurement model	98
Table 4.11	Unstandardized lambda-X matrix for the White measurement model	99

	Page	
Table 4.12	Completely standardised lambda-X matrix for the White measurement model	101
Table 4.13	Unstandardized lambda-X matrix for the Coloured measurement model	102
Table 4.14	Completely standardised lambda-X matrix for the Coloured measurement model	104
Table 4.15	Summary of the multigroup measurement models' goodness of fit statistics	105
Table 4.16	Statistical and Practical significance – Analysis of Metric Equivalence	110
Table 4.17	Summary of Measurement Model GOF Indices - Partial Strong Invariance Analyses	114
Table 4.18	Statistical significance of the scaled chi-square difference statistic: A test of partial scalar equivalence	115
Table 4.19	Practical significance of the CFI, Gamma Hat and McDonald difference statistics: a test of partial scalar equivalence	116
Table 4.20	Statistical and Practical significance – Analysis of Conditional Probability Equivalence	119
Table 4.21	Practical significance indices: Analysis of Partial conditional probability Equivalence	120
Table 4.22	Calculation of the number of freed measurement model parameters in the configural invariance Psycap first-order multigroup measurement model	122
Table 4.23	Calculation of the number of freed measurement model parameters in the configural invariance Psycap second-order multigroup measurement model	122
Table 4.24	Calculation of the number of freed measurement model parameters in the configural invariance PsyCap first-order multigroup measurement model defined as a classically parallel measurement model	124
Table 4.25	Summary of fit statistics for the classically parallel multigroup configural invariance measurement model	125

CHAPTER 1: INTRODUCTION

1.1 Introduction

In South Africa, Industrial Psychologists' scope of practice is set out by the Health Professions Council of South Africa (HPCSA). The scope outlines the domains which the professional is registered to work in, as well as areas of overlap with other professions. In the Industrial / Organisational (i.e. I/O) Psychologists' scope of practice, psychological assessment in particular is governed by strict legal parameters. Measures used by professionals need to be proactively shown to be valid and reliable, unbiased and can be fairly applied across groups, as well as classified by the HPCSA (Republic of South Africa, 1998, p. 7). These regulations stem from the Employment Equity Act 55 of 1998, Section 8 (Government Gazette, 1998), which placed cross-cultural assessment on the agenda in an effort to address the risk associated with assessment in a multicultural society (Meiring et al., 2006). The prominence of fair and unbiased testing in the South African environment is evident in the fact that other countries address issues of bias and fairness in professional organisations of psychologists' codes of conduct, while in SA it was incorporated into national law (Laher & Cockcroft, 2013; Van de Vijver & Rothmann, 2004).

The act places the onus of proof on the professional practitioner to show that the psychological measures they use meet these legal requirements. However, research has indicated that sourcing measures that are appropriate for a multilingual, multicultural South African society and compliant with the Employment Equity Act is a very difficult task (Abrahams, 1996; Meiring et al., 2006; Meiring et al., 2005). Nevertheless, the regulations are vital in ensuring proper test usage in a diverse society.

Many of the psychometric measures classified by the HPCSA have been developed in South Africa, for South Africans. Examples include the Maree Career Matrix (MCM), Learning Potential Computerised Adaptive Test (LPCAT), Meyer Interest Questionnaire (MB-10) and the South African Vocational Interest Inventory (SAVII) (Government Gazette, 2017). However, for various reasons, many of the psychometric measures used today did not originate from SA but were imported from international countries for use on the local population.

Laher and Cockcroft (2013) explain that the use of imported psychological measures previously, had links with South Africa's troubled past. In the 1900s, psychological assessment in SA followed international trends, as measures were imported from abroad and applied to all sectors in society during this time (Foxcroft, 1997). The majority of the measures were imported by SA from Britain and

were used during the Apartheid regime to justify the unequal distribution of resources. These included education and economic opportunities, as well as the exploitation of black labour in line with apartheid policies (Laher & Cockcroft, 2013; Nzimande, 1995; Sehlapelo & Terre Blanche, 1996). One such example was the job reservation policy which ensured employment for White individuals. “Psychological assessments were misused to support their policy as assessments which were developed and standardised on the British educated, white population were administered to illiterate, uneducated or poorly educated Black South Africans. The results were used to justify job reservation and further supported the logic of apartheid, as they indicated the superiority of white intellect over black intellect” (Laher & Cockcroft, 2013, p. 2)

A change in awareness regarding test usage occurred soon after the release of Mandela in 1990 and the democratic election in 1994. There was a stronger demand for the cultural appropriateness of tests, given the transformation and integration of society since 1994 (Meiring et al., 2005). A new system of democracy, mutual respect and freedom of expression was developed, underpinned by the Constitution. The shift in consciousness was thus strongly linked to the regulations regarding test usage promulgated in the EEA (Government Gazette, 1998). A high standard was evidently set, and the implementation thereof was filtered down to the practitioner, whom should now be more conscious of developing and utilising assessments in a fair and unbiased manner (Laher & Cockcroft, 2013; Sehlapelo & Terre Blanche, 1996).

The use of assessments not adapted or suited to the South African context opens up potential risks in assessment. The effects of poor measures in the assessment process can be collectively termed as bias. Van de Vijver and Tanzer (1997) explain that bias occurs when test score differences do not correspond with differences in the underlying construct. These are unwanted, yet systematic sources of variance which influence the scores on a measure. Three types of bias should be considered, namely construct, method and item bias (Van de Vijver & Leung, 1997a, b).

Construct bias is evident when the construct measured is not the same across cultures or when the behaviours characterising the construct are not the same across cultures (Fontaine, 2008; Harzing, 2006). This can stem from various sources, such as the definition of a construct not overlapping across cultures, causing bias (Meiring et al., 2005). The second type of bias, namely method bias, occurs due to method-related issues in a measure (Fontaine, 2008). “This can be caused by an incompatibility of the sample on factors other than the target variable, problems caused by instrument characteristics as well as administrative issues” (Meiring et al., 2005. p.2). Finally, item bias (or differential item functioning [DIF]) occurs at the item level when the meaning of an item is not the same across cultures. This can be due to various reasons, including inapplicability of the item content or poor translation

(Meiring et al., 2005). The result is that individuals with the same standing on the latent variable may differ in terms of their expected score on the measure, because of their group membership rather than actual differences on the latent variable (Berry, 2015; Fontaine, 2008).

According to Van de Vijver and Tanzer (1997), bias can be an indication of a source of cross-cultural differences that need to be investigated, such as stimulus familiarity or response styles. If the influence of nuance factors in measurement are not known or considered by the practitioner, these could lead to inaccurate and potentially unfair decision making with detrimental effects on the individuals involved, as well as the organisation (Theron, 2007). Measurement invariance or equivalence (MI / ME) is therefore introduced as a rigorous manner of testing for bias in measurement and meeting the requirements of the Employment Equity Act.

Little (1997, p.55) defines measurement invariance (or equivalence) as the mathematical equality of corresponding measurement parameters for a given factorially defined construct, across two or more groups. A lack of sufficient evidence to support the invariance of an instrument would result in the scientific inferences drawn from the instrument to be considered weak as the observed scores could not be interpreted unambiguously (Dunbar et al., 2011; Van de Vijver & Tanzer, 1997). According to Dunbar et al. (2011), levels of invariance and equivalence need to be met in order for observed scores to be directly comparable across different cultural groups. Four levels of invariance are presented, namely configural invariance, weak invariance, strong invariance and strict invariance. Furthermore, levels of equivalence are also presented, namely metric equivalence, scalar equivalence, conditional probability equivalence and complete equivalence, where each subsequent level of invariance and equivalence places tighter constraints on the measurement model (Dunbar et al., 2011).¹

This research aims to investigate the measurement invariance and equivalence of the Psychological Capital Questionnaire (PCQ-24) across Black, White and Coloured groups in South Africa. It is envisioned that this research will contribute to the body of knowledge regarding the PCQ-24 within the South African environment and inform its responsible use in future research and practical applications.

The PCQ-24 was developed from a body of knowledge termed Positive Organisational Behaviour (POB), which stems from positive psychology advocated by Seligman and Csikszentmihalyi (2000). This approach to psychology advocates a move away from the disease and dysfunction model as it focusses on individual strengths, which buffers the effects of negative stressors (Nelson & Cooper, 2007). This paradigm shift in psychology inspired the work of Fred Luthans, who championed research in

¹ Due to the complexity of the taxonomy by Dunbar et al. (2011) it will be unpacked in detail in Chapter 2.

organisational psychology with a distinct positive psychological focus. Luthans, along with Youssef, Avolio and Avey, among others, contributed to the development of positive psychology in I/O Psychology and shifted the emphasis solely from managing weaknesses, to also building on strengths in the workplace. In particular, they focused on the construct of Psychological Capital (i.e., PsyCap), its measurement and development as well as its relationship with positive psychological outcomes (Avey et al., 2008; Luthans, 2002a; Luthans & Jensen, 2002; Luthans et al., 2007a; Luthans et al., 2015; Luthans & Youssef, 2004). Their work aided in the move from a problem-oriented framework to concentrating on the determinants and conditions necessary for optimal organisational performance and well-being.

Luthans (2002a) recommended that psychological constructs that (a) are open to development in terms of organisational interventions, and (b) could be validly measured, should be studied by researchers because of their potential impact on work performance. Constructs which meet these inclusion criteria, according to Luthans et al. (2007b), include Self-efficacy, Resilience, Hope and Optimism, which have been collectively termed PsyCap.

Luthans et al. (2007a, p. 542) define PsyCap as:

an individual's positive psychological state of development and is characterised by: (i) having confidence (self-efficacy) to take on and put in the necessary effort to succeed at challenging tasks; (ii) making a positive attribution (optimism) about succeeding now and in the future; (iii) persevering towards goals and, when necessary, redirecting paths to goals (hope) in order to succeed; and (iv) when beset by problems and adversity, sustaining and bouncing back and even beyond (resilience) to attain success.

PsyCap therefore includes individuals' psychological resources, which they can draw upon in challenging times of development and growth. Embracing PsyCap thus aids in actualising potential, and assists individuals in being happier, healthier and more productive (Luthans et al., 2007b). Moreover, PsyCap is unique as it moves away from what you know (human capital) and who you know (social capital) to who you really are (your true self) (Nelson & Cooper, 2007).

Psychological Capital is a higher-order construct and is distinctive from other constructs in positive psychology as it is state-like in nature. It differs from trait-like constructs such as the "Big-Five" personality dimensions and implies that PsyCap is open to change and development as opposed to being relatively stable over time (Luthans et al., 2007b). Previous research supports the notion that Self-efficacy, Hope, Optimism and Resilience are developable. These include the work of Bandura (1997) on the development of Self-efficacy, Snyder (2000) on the development of Hope, Carver and

Scheier (2005) on strategies to develop Optimism, and finally Masten and Reed (2002) on Resilience-based developmental interventions.

The malleable characteristic of PsyCap makes it a valuable construct within the workplace. This is due to the fact that it is not only applicable in selection but can also be used in interventions for developmental purposes. Organisations could measure PsyCap during selection as an indication of candidate's ability to cope in adverse circumstances. Moreover, the construct could be utilised in the implementation of training and development programs within the organisation to assist employees in enhancing their Self-efficacy, Resilience, Hope and Optimism.

The PsyCap construct has received growing research interest in the literature and Görgens-Ekermans and Herbert (2013) noted that to date, studies have been conducted on samples from the USA (Luthans et al., 2015; Staples, 2014), Canada (Laschinger & Grau, 2012), China (Cheung et al., 2010; Wang et al., 2012; Wan & Hu, 2017), the United Kingdom (Nigah et al., 2012), Germany (Nolzen, 2018), India (Rani & Chaturvedula, 2018; Tripathi, 2011), South Africa (Bernstein & Volpe, 2016; Du Plessis & Barkhuizen, 2012; Roemer & Harris, 2018; Van Wyk, 2016) and Portugal (Rego et al., 2012). Furthermore, authors have published comprehensive reviews of the PsyCap construct, the psychometric properties of the measure, and its antecedents, outcomes and moderating effects (Dawkins et al., 2013; Nolzen, 2018; Wan & Hu, 2017).

Various studies have shown the positive effects of Psychological Capital on various workplace outcomes. For example, a study by Görgens-Ekermans and Herbert (2013) analysed the relationship between PsyCap, stress and burnout. The findings revealed the four sub-dimensions and the PsyCap total score to be statistically significantly related to perceived stress, as well as with work-related burnout. Furthermore, the results of the moderated regression showed strong evidence for PsyCap as a moderator in the relationship between stress and work-related burnout, as the interaction term was significant at the 0.05 level (Görgens-Ekermans & Herbert, 2013).

Furthermore, a study by Luthans et al. (2008) reported positive results regarding PsyCap's relationship with job satisfaction, performance and organisational commitment, as well as a mediating role in the relationship between a supportive organisational climate and employee performance. This was an important finding to the study because it indicated that PsyCap, in conjunction with a supportive organisational climate, has a positive impact on actual employee performance (Luthans et al., 2008). Similarly, a study by Luthans et al. (2007a) supports the findings regarding the relationship between PsyCap, work performance and job satisfaction. The results indicated that the higher-order PsyCap

construct was a better predictor of performance and job satisfaction than the individual sub-dimensions alone.

Psychological Capital has also been shown to be a significant predictor of desirable and undesirable work attitudes and behaviour. In a study by Avey et al. (2010a), PsyCap was found to add unique significant variance to the dependent variables (organisational cynicism, intention to quit, organisational citizenship behaviour and counterproductive work behaviour) above the control variables (self-evaluation, positive personality traits, person-organisation and person-job fit). PsyCap was therefore shown to be negatively related to organisational cynicism, intention to quit and counterproductive work behaviour and positively related to organisational citizenship behaviours (Avey et al., 2010a). These findings have important implications for the workplace considering that the outcomes mentioned above are key constructs within the field of I/O Psychology. Practitioners will always strive to improve positive attitudes and behaviours (e.g. organisational citizenship) and decrease negative attitudes and behaviours (e.g. cynicism, intention to quit and counterproductive behaviour) in the workplace. The results discussed therefore indicate that focusing on PsyCap in workplace interventions can add value beyond other widely recognised variables, such as extraversion, conscientiousness, core self-evaluations, person-organisation and person-job fit (Avey et al., 2010a).

The importance of psychological capital is further evident as it has also been shown to play a role in employees' psychological well-being (PWB). Youssef-Morgan and Luthans (2015) argue that several theoretical mechanisms link PsyCap to well-being. Firstly, they argue that well-being is predominantly shaped by affective and cognitive appraisals of our general lives as well as specific life events, domains and circumstances (Bakker & Oerlemans, 2012; Youssef-Morgan & Luthans, 2015). Secondly, the resources comprising PsyCap are characterised by the formation of positive appraisals of past and future life events. This is evident as the positive appraisals are "based on positive explanatory styles of the past (optimistic attributions), motivated effort and perseverance in the present (efficacy, resilience and hope- agency), and positive expectancies (optimistic outlook) and intentional goal pursuit (hope- pathways) for the future" (Youssef-Morgan & Luthans, 2015, p. 185). Thus, it is argued that PsyCap's positive appraisals contribute to individuals' well-being.

The second mechanism linking PsyCap to well-being is satisfaction with important life domains, which plays a vital role in well-being (Diener, 2000). Important life domains include health, work and relationships, and PsyCap has been shown to predict satisfaction in these domains (Luthans et al., 2013). Satisfaction with important life domains can lead to improved well-being over time due to inter-

domain crossover and spill over effects between the life domains. For example, if an individual had a lower level of Self-efficacy, but their Resilience was higher, their lower Self-efficacy PsyCap domain would be supported by their higher Resilience PsyCap domain. In this way, the motivation and information feedback drawn from the domains would result in feelings of control, mastery and overall well-being (Youssef-Morgan & Luthans, 2013).

Thirdly, resource theories link PsyCap to well-being as they propose that overall assessments of well-being are based on cognitive evaluations of the availability of personal resources (Wright & Hobfoll, 2004). PsyCap has been considered a psychological resource since its inception and hence an abundance of perceived PsyCap would result in positive appraisals of well-being (Avey et al., 2010b; Luthans et al., 2015; Youssef-Morgan & Luthans, 2015). Similarly, Fredrickson's (2003) Broaden-and-Build Theory is relevant here, as PsyCap and positivity aid in building individuals' psychological resource repositories. The PsyCap resources assist individuals to cope in difficult circumstances and therefore the development thereof will help them to overcome challenges, remain positive and improve their well-being (Youssef-Morgan & Luthans, 2015). This line of reasoning is further supported by the Conservation of Resources Theory (COR) which states that individuals "seek to obtain, retain, and protect resources and that stress occurs when resources are threatened with loss or are lost or when individuals fail to gain resources after substantive resource investment" (Hobfoll, 2002, p.312). The development of PsyCap as a positive psychological resource therefore aids individuals in dealing with adverse circumstances and the resultant positive gain spiral in conjunction with other positive resources, boosts overall well-being (Avey et al., 2010b).

The relationship between PsyCap and well-being is supported by research where PsyCap was shown to be a significant predictor of PWB across various demographic groups (Avey et al., 2010a; Gibson & Hicks, 2018; Mishra & Shafiq, 2018; Roche et al., 2014). These findings support the notion that organisations and employees would benefit from the development of PsyCap. Various studies and literature have demonstrated PsyCap's links to various positive employee outcomes in the workplace, including improved well-being, performance and organisational citizenship behaviours (Avey et al., 2010a; Görgens-Ekermans & Herbert, 2013; Luthans et al., 2007a). Consequently, organisations investing in PsyCap development interventions would benefit from the holistic positive impact on their employees.

Currently, the PCQ-24 is only being used for research purposes in South Africa, however the preceding evidence indicates that it is an important construct in the I/O Psychology environment and validation of the instrument would be valuable to practitioners in the workplace. To the knowledge of the

researcher and her supervisor to date, no study has not been conducted regarding the measurement invariance of the PCQ-24 across ethnic groups in SA.

The current study seeks to address the topic of the MI and ME of the PCQ-24 across Black, White and Coloured ethnic groups in SA. As mentioned previously, the inferences derived from measuring instruments affect the lives of individuals as well as the organisation. In the past, most assessments were developed in western cultures and consequently scientific evidence demonstrating the validity of the measures in a multi-cultural context such as South Africa needs to be shown. This study is not investigating cultural definitions of PsyCap and the potential biases as a result, but rather is interested in evaluating the MI and ME of the PCQ-24 across ethnic groups in SA. It is therefore envisaged that the research will contribute to the body of knowledge regarding cross-cultural assessment in SA the responsible use of the PCQ-24 in cross-cultural applications in particular.

1.2 Research Initiating Question

Consequently, the research initiating question is: Can observed scores from the PCQ-24 be meaningfully compared across ethnic groups in South Africa?

1.3 Research Objectives

The objectives of the study therefore are:

- a) Explore the current body of knowledge regarding the MI and ME of the PCQ-24,
- b) Analyse the gathered data to inspect the MI and ME of the PCQ-24 over Black, White and Coloured ethnic groups in South Africa, at the following invariance levels:
 - a. Configural invariance,
 - b. Weak invariance,
 - c. Strong invariance,
 - d. Strict invariance,
 - e. Metric equivalence,
 - f. Scalar equivalence, and
 - g. Conditional probability equivalence.

1.4 Overview of the Study

The following section will provide the theoretical framework for the study, outlining the construct of PsyCap and the psychometric properties of the instrument. The chapter will also cover an overview of MI/ME and current research regarding the invariance of the PCQ-24 across groups. Chapter 3 will introduce the research methodology that was used in the analysis of the PsyCap measure as well as ethical considerations of the study. Chapter 4 will provide a discussion of the study results of the various MI and ME analyses, followed by Chapter 5, which will discuss the implications of the results and recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The current research is aiming at evaluating the MI of the PCQ-24 across Black, White and Coloured groups in SA, in order to contribute to the body of knowledge regarding the appropriate usage of the measure and inform its use in future research and practical applications. The purpose of this chapter is thus to explicate the various sources of bias, including construct, method and item bias, as well as discussing MI and ME. Firstly, however, this chapter will provide an overview of the PCQ-24, discussing the development of the measure based on the widely researched construct of Psychological Capital. This will include a focus on the dimensions of the instrument, as well as the psychometric properties of the instrument reported in current literature. This will be followed by a critical review of past and current research regarding the measure.

2.2 PCQ-24

The PCQ-24 measures an individual's level of PsyCap via four subscales comprising six items each. These measure the four latent dimensions of PsyCap as conceptualised by the PCQ-24, namely Hope, Self-efficacy, Resilience and Optimism, comprising 24 items in total. The instrument is commonly used on working adults for the purposes of assessment and development, for either personal development or in the organisational context (Mind Garden, Inc., n.d.). The resulting score gives an indication of the individual's PsyCap, which is characterised by an individual having confidence to take on challenging tasks, having a positive attribution toward succeeding, persevering and redirecting paths toward goals as necessary, and bouncing back from adversity to achieve success (Luthans et al., 2007b).

The conceptualisation of the construct of PsyCap was inspired by a paradigm shift in Psychology. In I/O Psychology, this shift was known as Positive Organisational Behaviour (POB). POB is understood as "the study and application of positively oriented human resource strengths and psychological capacities that can be measured, developed and effectively managed for performance improvement in today's workplace" (Luthans 2002b, p.59). In order to be incorporated in the conception of POB, constructs had to meet certain inclusion criteria. These criteria include that while the psychological capacities must be relatively unique and positive, they must also be theory- and research-based, measurable, malleable (i.e. amenable to development) and related to work performance and other positive outcomes (Luthans et al., 2007a). According to Luthans et al. (2007b) the constructs which best met these inclusion criteria, were Hope, Self-efficacy, Resilience and Optimism. The four latent dimensions of the multidimensional construct of PsyCap will subsequently be discussed, followed by the psychometric properties of the instrument.

2.2.1 Latent Dimensions Comprising the PsyCap Construct

2.2.1.1 *Self-efficacy*

The latent dimension of Self-efficacy has its origin in Albert Bandura's (1997) Social Cognitive Theory. Self-efficacy can be defined as an individual's confidence about his/her abilities to mobilise the necessary motivation, cognitive resources and courses of action in order to successfully execute a task (Luthans et al., 2015, p.50). Therefore, Self-efficacy, or confidence, forms part of the driving force which motivates us in our endeavours based on our perceptions or beliefs that we will be successful. It is an awareness of who you are which develops over time and can be improved to envision what you can become (Luthans et al., 2015).

The relationship between Self-efficacy and work performance is well researched (Luthans et al., 2015), showing that it has a very positive impact on work performance. In addition, the benefit of Self-efficacy is amplified by its developmental potential. Bandura (1997) showed that an individual can enhance their Self-efficacy through four sources of efficacy expectations. These include, firstly, experiencing mastery, as success builds confidence. Secondly, Self-efficacy can be increased through social persuasion, as verbal encouragement has been shown to alter self-doubting beliefs. Thirdly, psychological arousal is a source of efficacy expectations, as individuals' emotional states have been shown to contribute to their feelings of efficacy. Finally, through vicarious learning individuals can indirectly learn and build their own confidence by observing the success of others. Developing efficacy can be achieved through targeted organisational interventions, or less informally, through positive feedback, for example (Luthans et al., 2007b).

According to Prinsloo (2013) individuals with high Self-efficacy possess five distinct characteristics which distinguish them from other individuals. These include setting high goals for themselves, thriving on challenges, possessing a high level of motivation, putting in the necessary effort to succeed, and persevering through obstacles. These characteristics are similar to those displayed by an individual with high Hope, relating to the agency and goal directed behaviour, as well as the pathways necessary to complete these goals (Luthans et al., 2007b).

2.2.1.2 *Hope*

The next dimension of PsyCap which meets the POB inclusion criteria is Hope, which has been labelled by Luthans et al. (2015, p.80) as "the will and the way". This is since the PsyCap definition of Hope has two components, both of which are necessary in order to display high levels of Hope when accomplishing goals, namely willpower and pathways (Luthans et al., 2015). The willpower component is evident when individuals feel determined to achieve their goals, set challenging goals for themselves and enjoy engaging in these goals, while the pathway component can be seen when individuals

proactively look for ways, and alternative paths when necessary, to accomplish their goals (Luthans et al., 2015). This understanding is echoed by Snyder et al. (1991, p. 287), well known researchers on Hope in the positive psychological sphere, who defined Hope as, “a positive motivational state that is based on an interactively derived sense of successful (i) agency (goal directed energy) and (ii) pathways (planning to meet goals)”.

Gaining a deeper understanding of this dimension is important as the term Hope is regularly used in everyday language. The common-sense conceptualisation, however, often differs substantially from the connotative meaning it carries as a construct in the POB literature. This is evident as Hope is often confused with wishful thinking or simply an unsubstantiated positive attitude, whereas the abovementioned definitions illustrate that as a scientific construct it encompasses much more.

As a psychological strength, Hope is very valuable to individuals and the relationship between Hope and performance in various life domains is well established (Luthans et al., 2007b). Luthans and Youssef (2004) demonstrated the significance of Hope in the work domain in a study conducted with over 1000 managers and employees. The results showed that Hope correlated positively with job satisfaction, performance and organisational commitment. Luthans and Youssef (2004) however cautioned that these results do not imply structural/causal paths. For example, Luthans and Youssef (2004) argued that Motivation (amongst others) most likely mediates the effect of Hope on all these latent variables.

Luthans et al. (2007b) explain that Hope can be developed in numerous ways, including appropriate goal setting, setting stretch goals (goals which are conducive for development), stepping (breaking challenging, long-term goals into more manageable, short-term goals), and creating employee involvement and opportunities for participation. In addition, Hope can be developed using reward systems, effective allocation of resources and strategic alignment of individuals’ talents and strengths to the position. Hope can also be developed through training which is interactive that can form part of micro interventions targeted to develop psychological strengths (Luthans et al., 2007b).

2.2.1.3 Optimism

The third dimension which met the inclusion criteria is Optimism, which Luthans et al. (2007b) argued is one of the least understood psychological strengths. Generally, an optimist is understood as an individual who is positive and believes that positive events will occur in the future, while pessimistic individuals have more negative thoughts and expect undesirable events will occur. As a psychological strength however, Optimism is defined as much more, an important component being the attributions or reasons which individuals use to explain positive or negative events (Luthans et al., 2007b). For

example, an individual anticipating positive events may still not be regarded as having a high standing on the scientific construct of Optimism if they do not interpret them with an optimistic explanatory style.

An explanatory style relates to the attribution's individuals make regarding their behaviour or events. According to Seligman (1998) an optimistic explanatory style is evident when individuals link a positive incident to personal, permanent causes and interpret negative incidents as situation-specific, temporary and external to the individual. In contrast, pessimists attribute positive events to external factors and negative events in terms of permanent, personal causes; the difference being that optimists recognise their individual efforts towards positive events in their lives (Seligman, 1998).

An important qualification that needs to be imposed on the definition of Optimism, however, is that the attributions need to be flexible and realistic. Luthans et al. (2007b) explain that non-scrutinising and blatantly optimistic people may expose themselves to undesirable side effects if they underestimate potential risks. Thus they advocate for 'realistic and flexible' optimism, meaning that individuals need to try to correctly appraise situations in order to choose when to use optimistic or pessimistic explanatory styles (flexible) and be realistic in their assessments, ensuring that they do not put too high expectations and pressures on themselves as this can lead to negative consequences for the individual (Luthans et al., 2007b). Realistic, flexible optimism thus aids in preventing these negative consequences and can be of great value to individuals in the workplace.

Here, the value of optimism lies in the fact that optimists and pessimists react differently to changing conditions in an organisation. Luthans et al. (2007b) explain that optimists are more likely to see what the future holds, focus their energy on capitalising on opportunities and embrace changes during turbulent times, and react differently to pessimists regarding changes that could result in adverse consequences. For example, if an employee with an optimistic explanatory style had been retrenched, Optimism would present itself as a psychological strength as the employee would more likely attribute the negative event to external causes, such as current economic conditions, rather than to internal causes resulting in self-blame and feelings of inadequacy (Luthans et al., 2007b). Optimism can thus help individuals during difficult times, and can also contribute to increased motivation, aiding long-term performance and success.

Additionally, Optimism meets the POB inclusion criteria due to its malleable character, meaning that it can be developed in individuals for their personal benefit. Luthans et al. (2007b) state that improvement can be achieved through either reinforcing an optimistic explanatory style or by unlearning a pessimistic explanatory style. Schneider (2001) proposed three approaches for

developing realistic optimism in the workplace. These include having leniency for the past (acknowledging the realities of the situation), being appreciative of the present (as any situation has positive aspects, no matter how unfavourable) and seeking opportunities for the future (proactively seeking future opportunities in line with one's strengths and weaknesses). As such, Optimism is a far more comprehensive construct than initially envisaged and can be a powerful tool for business leaders to motivate and inspire employees to choose challenges resulting in greater future performance (Luthans et al., 2007b).

2.2.1.4 Resilience

The last dimension of the PCQ-24 which met the inclusion criteria is Resilience, which is defined as “the capacity to rebound or bounce back from adversity, conflict, failure or even positive events, progress and increased responsibility” (Luthans, 2002b, p. 702). Furthermore, according to Masten and Reed (2002, p. 75) Resilience is understood as “a phenomenon characterised by patterns of positive adaption in the context of significant adversity or risk”. Resilience generally involves rising up against adversity, however, the PsyCap meaning encompasses the ‘bouncing back’ capacity, as well as the will to go beyond the normal or equilibrium point, leading to superior performance (Luthans et al., 2015).

Various factors play a role in an individual's Resilience development, either contributing toward or hindering its development. Luthans et al. (2007b) explain that an individual's Resilience is affected by asset factors that promote Resilience and risk factors that detract from Resilience and therefore collectively determine how individuals deal with adverse situations. These assets can include an individual's temperament, cognitive abilities and positive self-perceptions, while potential risk factors can include dysfunctional or destructive experiences and trauma, or less obvious vulnerability factors such as poor health and stress (Luthans et al., 2007b).

Furthermore, the underlying value system of an individual plays a major role in Resilience as an asset by shaping, guiding and giving meaning to one's actions, emotions and cognitions (Luthans et al., 2007b). Thus, individuals' beliefs and values can aid Resilience by carrying individuals through difficult or overwhelming life events by linking them to a more pleasant future to look forward to. Now that the dimensions have been explored, the next section will discuss the PCQ-24's psychometric properties, and more specifically, its reliability and construct validity.

2.2.2 Psychometric Properties of the PCQ-24

The reliability of an instrument relates to the extent to which variance in the observed item and subscale scores are produced by systematic sources of variance rather than random error (Nunnally,

1978). If the variance in the item and subscale score are only to a limited degree brought about by random error sources of variance, the subscale scores will remain consistent across different measurement occasions. Validity in turn refers whether the inferences derived from the subscale scores of the instrument are permissible (Kaplan & Saccuzzo, 2009).

2.2.2.1 Construct Validity

An instrument is developed to measure individuals' standing on a construct which carries a particular constitutive definition (Kaplan & Saccuzzo, 2009). Firstly, an internal structure is attributed to the construct by the constitutive definition and the construct is constituted by a number of latent dimensions that are structurally and/or correlationally associated to one other in a particular manner (Cronbach & Meehl, 1995). The instrument attempts to assess the construct as constitutively defined indirectly by eliciting observable behaviour in which the latent dimensions express themselves via specific test items. Specific items are therefore earmarked to elicit observable behavioural responses that will reflect test taker's standing on subdimensions of the measured construct (Cronbach & Meehl, 1995). A measurement model (in which particular items are regressed on latent dimensions) is implied by the internal structure of the construct and the design intention which underpins the instrument (Diamantopoulos & Siguaw, 2000). The construct validity of an instrument is firstly examined by inspecting whether the measurement model showed at least close fit in the parameter (i.e. not rejecting the null hypothesis of close fit $H_0: RMSEA < 0.05$), the completely standardised factor loadings are statistically significant ($p < 0.05$) and large ($\lambda_{ij} \geq 0.50$), the completely standardised measurement error variance are statistically significant ($p < 0.05$) and small ($\theta_{\delta ii}$), and the correlations between the latent dimensions are statistically significant ($p < 0.05$) and small ($\phi_{jk} < 0.90$), resulting in a good ratio of signal to noise.

As shown in a review by Dawkins et al. (2013), various studies support the construct validity of the PCQ-24. In particular, two studies conducted by Avey et al. (2008) and Avey et al. (2010a) showed strong evidence for a four-factor model underlying the PCQ-24. In both cases, the authors conducted a confirmatory factor analysis (CFA) of the measurement model, which revealed 'adequate factor analytic fit (SRMR= 0.05; RMSEA= 0.05; CFI= 0.96) (Avey et al., 2010a, p. 444; Avey et al., 2008). Subsequently, similar results have been demonstrated in other studies. Peterson et al. (2011) conducted a study on 179 employees from a financial services organisation in the United States. The study utilised longitudinal data to assess changes in individual's PsyCap over time and the resultant impact on performance. Petersen et al. (2011) explained that CFAs were conducted at each time point to assess the validity of the PCQ-24. They concluded that "the model fit results supported the validity

of the four-dimension structure of Psychological Capital (average CFI= 1.00, average TLI- 1.00, RMSEA ranged from 0.05 to 0.07, and average SRMR= 0.00)" (Peterson et al., 2011, p. 435).

Interestingly, a South African study by Du Plessis and Barkhuizen (2012) did not obtain results consistent with previous findings regarding the four-factor structure of the PCQ-24. The study investigated whether HR practitioners, as "custodians of change and positive behaviour in organisations in SA, embrace the core elements of POB" (Du Plessis & Barkhuizen, 2012, p. 16). A sample of 131 HR practitioners took part in the study, which consisted of 63.36% males and 36.64% females. Moreover, the sample was predominantly White (75.38%), followed by 16.92% Black respondents and only 7.69% Coloured and Asian ethnic groups. More than half of the sample (51.15%) was over the age of 45 years, with 48.85% under the age of 45 years respectively. Furthermore, in terms of home language, the majority of respondents were Afrikaans (49.23%), followed by English (33.08%) and other languages (17.69%).

The PCQ-24 was administered to the sample of 131 participants and an EFA was conducted (as opposed to CFA). The results, however, were very perplexing. Firstly, the authors presented findings regarding the factor analysis, which revealed six possible factors underlying the measure. The authors decided not to interpret the fifth and sixth factors however, stating that the first four accounted for 54% of the total cumulative variance. Secondly, Du Plessis and Barkhuizen (2012) proceeded to rotate the remaining four factors to achieve 'a more interpretable structure', using the orthogonal rotation method. Here, according to the authors the results did not support a four-factor structure of the PCQ-24, as negative correlations were evident among the factors. The authors proceeded to reverse score the negatively keyed items, which they named items "4, 7 and 10 as per the original design of the PCQ" (Du Plessis and Barkhuizen, 2012, p. 23). It is unclear where the authors derived these item numbers from however, as the reverse keyed items in the PCQ-24 are well-known as items 13, 20 and 23 (Avey et al., 2010a; Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013; Luthans et al., 2007b). The impact of this on the subsequent results is therefore unknown.

Thirdly, inspection of the results of the four-factor structure showed that most items did not load on the original intended dimensions. For example, the results showed that 11 items loaded on the first factor (originally Self-efficacy, termed "Confidence/Hope" by Du Plessis and Barkhuizen [2012]), namely items 20, 5, 21, 24, 14, 16, 15, 17, 19, 6 and 8. Some of the items which loaded on "Confidence/Hope", were originally intended to load on Resilience (items 14 – 17) and Optimism (items 19- 21 and 24). A similar observation was evident amongst the other dimensions, with numerous items not loading on their intended dimensions. Lastly, Du Plessis and Barkhuizen (2012, p. 24), stated, "Because of the conclusion that the sample did not differentiate between Confidence and

Hope, these constructs were combined to form Hopeful-Confidence”, and consequently a three-factor structure underlying the measure was suggested by the authors. The results of the three-factor structure were just as perplexing however and did not support the authors’ recommendations. This was evident as items which traditionally loaded on Optimism (PCQ 19- 20, 21 and 24) loaded on the “Hopeful/Confidence” and Resilience subscales. Furthermore, items which traditionally measured Resilience were loaded on the “Hopeful/Confidence” subscale and Optimism subscales (PCQ13 -18). Numerous cross-loadings of the original Self-efficacy and Hope subscales were also evident, and hence the findings of Du Plessis and Barkhuizen’s (2012) study are highly questionable. It is noted however that a very small sample participated in the study (n=131), which was not very representative of the greater population (i.e. mostly White respondents (75.38%), who were largely male (63.36%), Afrikaans speaking (49.23%) and over the age of 45 (51.15%). As a result, the generalisability of the findings is limited.

Subsequent South African studies yielded differing findings to that of Du Plessis and Barkhuizen (2012). Firstly, Görgens-Ekermans and Herbert (2013) conducted a study² in which the CFA of the measurement model showed strong evidence for a four-factor model underlying the PCQ-24. The close fit null hypothesis was not rejected ($p = 0.93$; $p > 0.05$) and was supported by other GOF results (RMSEA= 0.04; CFI= 0.98; SRMR=0.06). The researchers then attempted to perform a CFA to inspect the higher-order factor of PsyCap in the South African environment, however the model would not converge. Alternatively, an EFA was performed on the four-factor model (with the latent variable scores derived from the CFA results) where factors were extracted by considering the scree test as well as ‘eigen value greater than one rule’. One factor was extracted which explained 69.33% of the variance. It was concluded that partial support for a higher-order PsyCap construct on the South African sample was thus obtained (Görgens-Ekermans & Herbert, 2013). Furthermore, the factor loadings were found to be significant but only moderately large with two missing the 0.50 criterion in the completely standardised solution (Görgens-Ekermans & Herbert, 2013).

More recently, a study by Munyaka et al. (2017) which aimed to investigate the relationship between authentic leadership, PsyCap, psychological climate, team commitment and intention to quit, found further support for a four-factor structure underlying the PCQ-24. The sample included 204 junior and

² The sample in this case consisted of 209 employees from a medium-sized organisation in the construction industry in SA. The sample was predominantly male (69.4%) with 29.2% females respectively. Furthermore, in terms of ethnicity, 45% of the sample was White, 38.3% “mixed-race” and 12.4% Black respondents. Görgens-Ekermans and Herbert (2013) commented that a further 2.9% of the sample was Asian, but the remaining 0.5% of the ethnic data for the sample was missing. Lastly regarding first language, the majority of respondents indicated Afrikaans (56.5%), 32.1% indicated English and the remaining 6.7% indicated IsiXhosa (Görgens-Ekermans & Herbert, 2013).

senior management level employees in a South African based, global tyre manufacturing business. The majority of respondents were male (73.4%), with 26.6% female respondents. In terms of ethnicity, the study reported 61.9% White respondents, followed by 19.8% Black respondents and 14.2% 'Mixed race'. The remaining respondents (4%) included Indian, Asian and/or Chinese respondents. The home language distribution of respondents was mostly English (41.6%), followed by Afrikaans (37.4%) and IsiXhosa (16.8%). The remaining respondents indicated their home language as either South Sotho or Chinese (4.2%). In this study, the authors "revalidated" the measure, indicating that four items were possible poor items, which were eventually removed from the measure. The authors neglected to reveal however, which items were excluded (Munyaka et al., 2017). Despite this, the Cronbach's alpha value of the original scale ($\alpha = 0.87$) and revalidated version ($\alpha = 0.84$) were both satisfactory (> 0.70).

A third study, conducted by Kotze (2018), tested the influence of PsyCap, self-leadership and mindfulness on work engagement on South African employees. The sample consisted of 407 employees from a variety of organisations in both the public and private sector. Of the respondents, 52% were female, with 48% male respondents respectively. In terms of ethnic groups, 51.8% of respondents were Black, 39.4% were White, 7.6% Coloured and 1.2% Asian. Moreover, most participants indicated African languages (47.7%) as their language, followed by Afrikaans (41.8%), English (10.3%) and other languages (2%).

The author tested the construct validity of the measurement model by assessing the "outer loadings, average variance extracted (AVE), composite reliability, and Cronbach's alpha" (Kotze, 2018, p. 287). Here, evidence in favour of construct validity would be achieved if the standardised loadings obtained values of 0.70 or higher, while standardised loadings of 0.40 or less would be excluded from the measurement model. Furthermore, the authors stated that the AVE value should be higher than 0.50, while the composite reliability and Cronbach's alpha value should be greater than 0.70 (Kotze, 2018). The results revealed that of all the PsyCap subscales, the Resilience and Optimism subscales did not meet the AVE requirement of 0.50. In addition, the Resilience and Optimism subscales were the only two that missed the Cronbach's alpha value of 0.70, with Resilience obtaining 0.670 and Optimism 0.588 respectively. Consequently, Kotze (2018) excluded the items from the measurement model which obtained low loadings to obtain the minimum AVE of 0.50. Inspection of the modified measurement model showed that items 13 and 16 were removed from the Resilience subscale, and items 20 and 23 were removed from the Optimism subscale. Evidently, three of the four items removed (i.e. 13, 20 and 23) are the three negatively keyed items in the scale, which have been shown to be problematic in previous PsyCap studies. In Kotze's (2018) study therefore, where the majority of participants did not indicate English as their first language, it is not surprising that the negatively keyed

items performed poorly in this case, as literature has shown that participants tend to perform poorer on these items, especially when they are not assessed in their mother tongue (Dalal & Carter, 2015; DiStefano & Motl 2006; Jackson Barnette 2000). Due to the measure being modified by the author however, the reliability results will not be discussed further under the Reliability section, as it is not appropriate to compare the Cronbach's alpha values of the amended subscales to previous South African studies.

The fourth study to be discussed was conducted recently on a Zimbabwean sample (Ngwenya & Pelser, 2020). The study aimed to investigate the impact of employee engagement, employee performance, PsyCap and job satisfaction in the manufacturing sector in Zimbabwe. The sample included 257 randomly selected employees from 8 participating manufacturing sector organisations. The demographics of the sample included 59% male respondents and 41% female respondents. Unfortunately, the authors did not report on the sample ethnic group distributions or first language distributions in this case. Ngwenya and Pelser (2020) indicated that measures with AVE values greater than 0.50 would provide evidence in support of convergent validity. In this case, the PsyCap scale obtained an AVE value of 0.754 which exceeded the 0.50 criterion, thereby supporting the convergent validity of the scale. Furthermore, the authors indicated that all the PsyCap subscales obtained factor loadings above 0.70, obtaining values of 0.883, 0.886, 0.795 and 0.884 respectively. However, it is clear from the results that the authors did not fit an outer model (with PLS) on the item level. They used the PsyCap subscale scores as observed variables. Hence the factor loadings and AVE reflects the higher order structure of the instrument. The authors did not indicate the subscale names but rather X.1 – X.4, so it is not possible to comment on which specific factors obtained the factor loadings mentioned above. The authors concluded that the results showed that the subscales could be regarded as appropriate indicators of the respective constructs, satisfying the requirements of convergent validity (Ngwenya & Pelser, 2020).

Previous international studies have also supported the higher order structure of PsyCap. Of these results included CFAs performed by Luthans et al. (2007a) which supported the higher-order structure of the instrument (RMSEA = 0.046; CFI = 0.93; SRMR = 0.051). These findings were supported by a cross-validation on a second sample (Luthans et al., 2007a). Furthermore, multiple CFAs were reported to analyse statistical differences between various combinations of three factor models, a one-factor model and the higher-order factor, providing further evidence for the higher-order structure of the PsyCap construct. Subsequently, these findings have also been demonstrated in other studies (Luthans et al., 2010).

2.2.2.2 *Discriminant validity*

A highly relevant, but very often ignored, question when validating psychological tests is whether the instruments succeed in measuring these interrelated latent dimensions with sufficient precision as separate, but interrelated constructs. When inspecting the factor structure of the PCQ-24, Du Plessis and Barkhuizen (2012) noted that some items did not load significantly on the factors they intended to reflect. Conceptually sound interpretations based on the four-factor model could thus not be made and factor analysis was subsequently repeated on the data, for three factors. The authors drew the conclusion that the sample did not differentiate between hope and self-efficacy (confidence) and consequently the constructs were joined, creating Hopeful-Confidence (Du Plessis & Barkhuizen, 2012). Despite the changes to the factor structure, multiple items still appeared to cross-load on the factors, which the authors attributed to the synergistic effects of the PsyCap construct. They did acknowledge however that the synergistic effect may have been more evident in the results due to the size of the sample in the study (Du Plessis & Barkhuizen, 2012).

More favourable results regarding the discriminant validity of the PsyCap subscales were obtained by Görgens-Ekermans and Herbert (2013) which confirmed previous international research (e.g. Luthans et al., 2007a). In this study a series of CFAs were performed where increasing parameter constraints were placed on the measurement models and compared to the unconstrained model. The GOF statistics indicated that each of the partially constrained models fitted the data significantly poorer than the unconstrained model, which indicated that each of the PCQ-24 subscales demonstrated discriminant validity (Görgens-Ekermans & Herbert, 2013). The magnitude of the correlations between the subscales supported these findings, as only two intercorrelations marginally surpassed the 0.60 critical value proposed by Kline (1998) for discriminant validity. This was interpreted by the authors as evidence, alongside the series of CFA evidence, that the PsyCap dimensions do not correlate too highly. Collectively, it was concluded that these findings supported the discriminant validity of the PsyCap subscales (Görgens-Ekermans & Herbert, 2013).

2.2.2.3 *Reliability*

According to Tabachnick and Fidell (2013, p. 733), "Reliability is defined in the classic sense as the proportion of true variance relative to total variance (true plus error variance)". A well-known form of reliability is internal consistency reliability, which evaluates whether the various items of an instrument measure the same underlying trait (Kaplan & Saccuzzo, 2009). Consequently, assessments designed to measure a number of traits would be expected to yield lower internal consistency reliability estimates. On the other hand, measures which are intended to measure a single dimension should yield higher internal consistency values. Lower values in this case may indicate that the items

of the measure are not reflecting the same underlying trait. George and Mallery (2003) proposed a taxonomy depicted in Table 1.1 which allows for a consistent interpretation of internal consistency reliability coefficients over research studies.

Table 1.1:

Taxonomy for the interpretation of the magnitude of Cronbach alpha internal consistency reliability coefficients.

Cronbach alpha	Interpretation
$\alpha \geq .90$	Excellent
$.90 > \alpha \geq .80$	Satisfactory
$.80 > \alpha \geq .70$	Acceptable
$.70 > \alpha \geq .60$	Questionable
$.60 > \alpha \geq .50$	Poor
$.50 < \alpha$	Unacceptable

A review by Dawkins et al. (2013) illustrated that reliability results have been fairly consistent across studies conducted in the United States and China (e.g. Avey et al., 2011; Avey et al., 2008; Avey et al., 2009; Cheung et al., 2010; Combs et al., 2012; Luthans et al., 2010; Luthans et al., 2007) where reported reliability alphas of the subscales were above the 0.70 level in most cases. Dawkins et al. (2013) noted however, that although generally adequate findings were obtained for the internal consistency of the subscales, the reliabilities of the Resilience and Optimism subscales appeared to be lower than the Hope and Self-efficacy sub-dimensions overall. Reported alpha coefficients for Optimism ranged between .63 to .69 (Avey et al., 2006; Luthans et al., 2008; Roberts et al., 2011) and from .63 to .66 for Resilience (Combs et al., 2012; Luthans et al., 2008). According to Schmitt and Stults (1985), scale reliability can be reduced by reverse-scored items. The addition of the negatively keyed items on the Resilience and Optimism subscales could therefore be negatively impacting the reliability of these subscales (Dawkins et al., 2013).

The reliability of the PCQ-24 has also been demonstrated in the South African context. The findings reported by Du Plessis and Barkhuizen (2012) indicated that the Cronbach's alpha coefficients of all the subscales were above the 0.70 critical value (albeit for the three-factor model proposed by the authors). Furthermore, Görgens-Ekermans and Herbert (2013) obtained the following alpha results for the reliabilities of the subscales: Hope= 0.81 and Self-efficacy= 0.83, where the item analysis indicated that no items should be flagged as potentially poor items (Görgens-Ekermans & Herbert, 2013). Furthermore, Optimism obtained an alpha coefficient of 0.67 and Resilience 0.69, which evidently marginally missed the 0.70 critical value. These findings corresponded with previous studies indicating consistently lower reliabilities for the two subscales (Avey et al., 2010a; Luthans et al., 2007a). Lastly, the item analysis results indicated that item 20 on the Optimism subscale and item 13

on the Self-efficacy subscale obtained lower correlations with the underlying factors (Görgens-Ekermans & Herbert, 2013).

Subsequent SA studies dated between 2015 - 2020 have yielded results in line with those obtained by Görgens-Ekermans and Herbert (2013). The first study to be discussed was conducted by Langenhoven (2015). The study used the job demands-resources (JD-R) model as a framework to investigate the relationships between job demands, resources, burnout, engagement and intention to quit in call centres in South Africa. In this study, PsyCap was included in the model as a personal resource which aided in fostering work engagement amongst employees and also aided employees in coping with job demands, reducing the level of burnout and intention to quit the organisation (Langenhoven, 2015).

The sample consisted of 223 employees from two South African call centres. The biographic data of the respondents revealed that 64% of respondents were female, with 36% males. Furthermore, in terms of ethnic groups, the majority of respondents were Coloured (62%), followed by African (17.5% and White (15%) respondents. The remaining participants (5.5%) indicated Indian and Asian as their ethnic groups. Information regarding the participants' first language distribution was not available in this case. The PCQ-24 was administered to the 223 employees and the results revealed the overall scale reliability was 0.81 (> 0.70). Furthermore, the reliabilities of the Self-efficacy, Hope, Resilience and Optimism subscales were 0.71, 0.86, 0.68 and 0.61 respectively. Evidently, in this case the Resilience and Optimism subscales fell below the 0.70 criterion, however the Self-efficacy subscale only exceeded the criterion slightly. While the reliability of the Self-efficacy subscale was expected to be higher, it still met the requirements of satisfactory reliability (> 0.70). The finding regarding the Resilience and Optimism subscales, however, was in line with previous research. Langenhoven (2015) stated, "The Resilience subscale showed one poor item, and the Optimism subscale showed two poor items. It was noted that the poor items in both the Resilience and Optimism subscales were reversed items. It was speculated that the reversed nature of these items caused participants some confusion". Therefore, the reverse keyed items were also shown to be problematic in Langenhoven's (2015) study.

Secondly, a study conducted by Bernstein and Volpe (2016) investigated the implications of sex role identity (SRI) and Psychological Capital for organisations, based on research indicating that SRI and PsyCap are important for organisational well-being. The study was based on 478 participants, who were gathered using a convenience sampling technique. The sample consisted of individuals working in SA organisations, with the final sample comprising 27.2% males and 72.8% females respectively. The majority of participants were Caucasian (61.9%), followed by Black African (23.2%), Coloured (8.2%), Indian (5.9%) and other ethnic groups (0.8%). Inspection of the reliability statistics showed that the PsyCap scale obtained an overall reliability coefficient of 0.89. Moreover, the subscales obtained

reliability coefficients of 0.85 (Self-efficacy), 0.82 (Hope), 0.69 (Resilience) and 0.62 (Optimism). Bernstein and Volpe (2016) suggested that the results met the minimum recommended level of 0.60, however based on the Nunnally criteria of good internal consistency, it is evident that the Resilience and Optimism subscales missed the 0.70 criterion, in line with findings by Gørgens-Ekermans and Herbert (2013) and Langenhoven (2015).

Similar findings were shown in a third study by Van Wyk (2016), who investigated the role of PsyCap in protecting the psychological well-being of call centre employees in SA. The sample consisted of 201 call centre employees from a number of different organisations. Most of the sample was female (73.1%), with 26.9% males respectively. Furthermore, Van Wyk (2018) indicated that 52.7% of the sample was Coloured, followed by 37.8% Black Africans. The remaining participants (comprised of White and Indian ethnic groups) accounted for less than 10% of the sample. The author did not report on the first language distribution of the sample. Regarding the psychometric properties of the measure, the following reliability coefficients were reported for the subscales of the PCQ-24: Self-efficacy (0.797), Hope (0.834), Resilience (0.686) and Optimism (0.621). Furthermore, the author stated that item 13 was flagged for removal on the Resilience scale. In addition, item 20 was flagged for removal in the Optimism subscale (evidently, two of the three negatively keyed items in the measure). Both items would result in significant increases in the Cronbach's alpha values from 0.686 to 0.746 and 0.621 to 0.668 for the Resilience and Optimism subscales respectively. Although item 23 was not flagged to be removed in this case, Van Wyk (2016) did note upon inspection of the PCQ-24 measurement model, that all items obtained statistically significant factor loadings, except item 23 which obtained a loading below 0.40 (0.259). These findings were therefore in line with results by Gørgens-Ekermans and Herbert (2013), who also flagged the negatively keyed items as poor items in the scale.

A fourth study conducted by Roemer and Harris (2018) aimed to analyse the relationship between perceived organisational support (POS), employee well-being and PsyCap. A sample of 159 South African employees was recruited for the study utilising a convenience, snowball sampling technique. The final sample consisted of employees from varying professions and industries and was made up of 39.6% males and 60.4% females (Roemer & Harris, 2018). The authors did not describe the participants' ethnicity, however the first language of respondents was reported. The majority of participants indicated Afrikaans as their first language (46.5%), followed by English (39.6%). IsiXhosa and IsiZulu speaking participants made up 13.2% and 0.6% of the sample respectively. Inspection of the results revealed that the overall reliability of the PCQ-24 was 0.90, indicating high reliability (Nunnally, 1987). The subscales obtained reliability coefficient values of 0.79 (Self-efficacy) and 0.84

(Hope), while the Resilience and Optimism subscales again revealed lower internal consistency scores, at 0.64 and 0.67 respectively. Furthermore, Roemer and Harris (2018) commented that the Cronbach's alpha values could be improved substantially by removing items 13 and 20 (two of the three negatively keyed items in the scale). The subscale reliabilities would increase from 0.64 to 0.70 for Resilience and from 0.67 to 0.72 for Optimism if the two items were removed, meeting the criterion of acceptable reliability (Nunnally, 1987). The authors consequently decided to remove both items from the subsequent analyses and concluded that "items 13 and 20 can be regarded as problematic items when being used with South African participants." (Roemer & Harris, 2018, p.8).

Furthermore, a fifth study conducted by Ramsden (2019) investigated the relationship between PsyCap, job satisfaction and work engagement at a Meridian Holdings establishment. The sample consisted of 118 support staff who were mostly women (62%), with 38% males participating respectively. In terms of the age distribution of respondents, the majority of respondents were aged between 25 – 34 years (46%), 31% between 35 – 44 years and 13% between 45 – 54 years of age. The remaining participants were aged between 18 – 24 years (6%), 55 – 64 years (4%) and finally 65 years and older (1%) (Ramsden, 2019). The researcher did not report on respondents' ethnicity or language distributions. The item analysis revealed the following reliability statistics for the PsyCap subscales: Self efficacy= 0.86, Hope= 0.86, Resilience= 0.77 and Optimism= 0.66. The subscale reliabilities in this case could be considered satisfactory as the Optimism was the only subscale which marginally missed the 0.70 cut off value. It is evident, however, that again the Resilience and Optimism subscales obtained lower internal consistency reliabilities in comparison to the other subscales, which is in line with previous research. Ramsden (2019) commented however, that although the subscale reliability was satisfactory in this case, the reverse keyed items (items 13, 20 and 23) did obtain lower item-total correlations in relation to the other items in the subscales. The author concluded that this could indicate that the respondents were confused by the negatively keyed items, which literature has shown can impact the reliability of the scale (Ramsden, 2019).

Finally, a recent study by Sepeng et al. (2020) aimed to assess the role of Psychological Capital as a mediator in the relationship between authentic leadership, organisational citizenship behaviour and intention to quit. The research was conducted on a sample of 633 healthcare employees in the public sector. The authors reported that the majority of participants were "Black (87.9%), Sesotho-speaking (44%) females (79.6%)". In addition, the mean age of participants was reported as 42 years old. The results revealed that the overall reliability coefficient of the PsyCap scale was 0.88 (Sepeng et al., 2020). In this case, the authors created a parcel comprising the four subscales (thereby creating a unidimensional PsyCap construct). Consequently, Cronbach's alpha values were not available for the

individual subscales of the PCQ-24. Despite this, the authors concluded that the scale revealed satisfactory reliability. (> 0.70).

In other cases, however, studies have yielded differing results to those presented above, in some cases more favourable and in others less favourable than the general trend of findings. Firstly, Simons and Buitendach (2013) studied the influence of PsyCap, work engagement and organisational commitment amongst call centre employees. Their sample consisted of 106 employees from a SA call centre, which was made up of 70.8% women and 29.2% men. Furthermore, most of the sample was aged between 25-35 years old (47.2%), followed by the 24 years and younger age group (37.3%). Lastly, 15.1% of respondents were aged between 36-45 years old (Simons & Buitendach, 2013). The authors did not report on respondents' ethnic distribution or first language. In terms of the psychometric properties of the PCQ-24, the overall reliability of the scale was 0.91, indicating high reliability (Nunnally, 1987). Furthermore, in terms of the individual subscales, the Self-efficacy scale obtained an alpha value of 0.87, Hope: 0.76; Resilience 0.90 and Optimism 0.72. Interestingly in this case, the Resilience subscale obtained the highest alpha value out of all the subscales. This finding is not in line with previous studies, which consistently show that the Self-efficacy and Hope subscales perform better than the Resilience and Optimism subscales. It is noted however, that the study was completed on a relatively small sample (106 employees), which the authors concurred, "limits the generalisation of the results to a larger population of call centre employees" (Simons & Buitendach, 2013, p. 10).

The second study to be discussed was conducted by Herholdt (2015), which inspected the antecedents of organisational citizenship behaviours and work engagement amongst nurses in South Africa. A sample of 199 nurses from private hospitals based in Gauteng and the Western Cape participated in the study. The sample comprised mostly women (97%), with only 3% men represented. Furthermore, a relatively equal number of respondents were aged between 30-39 (29%) and 40-49 (28%) years old. A further 21% of respondents were aged between 50-59, while 16% were aged between 20-29. Only 5% of the sample were aged 60-69 years old (Herholdt, 2015). The ethnicity and first language distribution of respondents was not reported in the study. The results of the reliability analysis indicated the following Cronbach's alpha values for the PsyCap subscales: Self-efficacy= 0.87, Hope= 0.87, Resilience= 0.61 and Optimism= 0.49. The Resilience and Optimism subscales clearly performed poorer than the Self-efficacy and Hope subscales, with Optimism falling far below the criterion of satisfactory reliability (0.70). Herholdt (2015) explained that removing item 13 in the Resilience subscale would increase the alpha value significantly to 0.76. In addition, while item 20 was flagged as potentially poor in the Optimism subscale, removal of this item would not result in a drastic increase in subscale reliability. The author concluded that the results could be attributed to the negatively

keyed items in the scale. Despite these findings, the overall scale reliability of the PCQ-24 was 0.82, which the author concluded indicated acceptable internal consistency reliability (Herholdt, 2015).

Thirdly, a study by Van der Merwe (2016) investigated the predictors of intention to quit amongst sales employees in the Financial Services Industry. The sample included 102 personal financial advisers; the majority of which were male (75%). Furthermore, most respondents were aged between 18-30 years old (45%), followed by the 31-45-year-old age group (38%). The remaining 17% of the sample was 46 years or older (Van der Merwe, 2016). In this case the author did not report on the sample ethnicity or first language distribution. The results of the analysis of the measurement model revealed the overall reliability of the scale as 0.92. Furthermore, the reliability of the subscales was 0.96 (Self-efficacy), 0.94 (Hope), 0.90 (Resilience) and 0.91 (Optimism) respectively, indicating very good reliability well above the 0.70 critical value (Nunnally, 1987). Clearly, these results are not in line with the trend of previous South African PsyCap studies. Unfortunately, as the author did not report in ethnicity or language distribution of respondents it is not possible to infer why these results could have been observed in this case. A limitation of the study should be noted however, namely that the sample size utilised in the study was very small ($n= 102$), comprising of mostly male respondents (75%). The generalisability of the results is therefore limited (Van der Merwe, 2016).

Lastly, a study by Kanengoni et al. (2018) investigated the relationship between PsyCap and various workplace outcomes including well-being, organisational commitment and job satisfaction, among South African church ministers. In total, 191 church ministers across all nine provinces participated in the study, most of which were male (94%). Furthermore, the authors reported that most participants held a degree (54.3%), followed by 27.7% holding a diploma and a further 4.8% holding a post-graduate degree. Kanengoni et al. (2018) reported the overall scale reliability of the PCQ-24 as 0.88. Unfortunately, the authors did not report on the reliability per subscale, so it is not possible to assess whether any of the subscales performed better than others. It was highlighted however, that this sample was not representative of female respondents, consisting of 94% males respectively; the generalisability of results is therefore limited (Kanengoni et al., 2018).

In the preceding studies, limitations were noted in each study relating to small sample sizes (± 100 respondents), where the authors concurred that the generalisability of results was limited as a result (Kanengoni et al., 2018; Simons & Buitendach, 2013; Van der Merwe, 2016). Furthermore, a review of the sample demographics in each case also revealed that the samples were not fully representative of the respective populations, as was highly evident in the gender distributions for example (i.e. 97% female: Herholdt, 2015, 94% male: Kanengoni et al., 2018; 75% male: Van der Merwe, 2016). Therefore, considering the findings presented in the preceding studies which were in line with

International trends, overall sufficient evidence was provided in support of the reliability of the measure in the South African context. The following section will highlight current research analysing differences in PsyCap across groups.

2.2.3 Current Research Examining Group Differences in Psychological Capital

Comparisons over different groups in terms of their PsyCap routinely seem to take place and have been reported by various practitioners, however in the absence of invariance results, one would not know if the differences are due to real latent mean differences or due to bias in measurement (Steenkamp & Baumgartner, 1998). This section will provide a critical review of studies where findings illustrated supposed differences between groups in terms of their PsyCap.

The first study to be discussed was conducted by Avci and Erdem (2017). This was a cross-national study which investigated the PsyCap of 336 employees by nationality and status. The sample included “security employees in a military organization consisting of seventeen countries, namely Turkey, U.S.A., Germany, Portuguese, Slovenia, Austria, Poland, Greece, Swiss, Ireland, Hungary, Romania, Croatia, Czech Republic, Sweden, Italy, Finland, Denmark, Norway, Britain, Netherlands, Ukraine, Canada, Lithuanian, Bulgaria, Albania and Kosovo” (Avci & Erdem, 2017, p. 205). Analysis of variance (ANOVA) tests were conducted and showed that regarding nationality, statistically significant differences were evident in Optimism ($p > 0.05$), Hope ($p > 0.05$) and Self-efficacy ($p > 0.05$) indicating that the Turkey participants’ Hope and Self-efficacy differed significantly from European respondents, and that Turkish respondents’ Optimism differed significantly from Northern and Western European nations. Finally, statistically significant differences were not found for Resilience ($p < 0.05$), implying that Resilience did not differ significantly across nationalities (Avci & Erdem, 2017).

Secondly, a study by Barmola (2013) investigated differences in PsyCap scores over gender in a sample of Indian students (50 males and 50 females). The results indicated no statistically significant gender differences on Resilience, Self-efficacy and Optimism, although significant differences were found on Hope ($t = 2.77$, $p < 0.01$). Barmola (2013) concluded that the findings indicate that gender differences may be observed among adolescents in terms of their levels of Hope, even in the presence of the small sample which was used in this study. Again, it could be argued that in the absence of invariance results, these findings could be ambiguous and as a result of other factors other than true latent group differences (Steenkamp & Baumgartner, 1998).

Thirdly, an Indian study was conducted by Rani and Chaturvedula (2018) with a sample of 32 female and 100 male officers of the Indian armed forces, with ages ranging from 21 to 60 years of age. The results indicated the following results between gender groups: Hope ($t = 1.062$; $p = 0.290$), Self-efficacy ($t = 1.067$; $p = 0.292$), Optimism ($t = 1.781$; $p = 0.077$), Resilience ($t = 0.085$; $p = 0.933$) and PsyCap total

score ($t = 1.252$; $p = 0.213$) (Rani & Chatuvedula, 2018). The authors (i.e. Rani & Chatuvedula, 2018) proceeded to conclude that none of the data provided evidence of statistically significant gender differences among the PsyCap dimensions, as well as in the total score (Rani & Chatuvedula, 2018).

The fourth study to be discussed is a South African study conducted by Du Plessis and Barkhuizen (2012), who performed a series of MANOVA tests to analyse the relationships between various demographic variables (including age, language and ethnicity) and the PsyCap subscales³. The sample included 131 HR Practitioners which consisted of roughly 63% males and 36% females, as well as 75% White, 17% Black and 8% other ethnic groups (Du Plessis & Barkhuizen, 2012). The results of the MANOVAs indicated that statistically significant differences ($p < 0.05$) existed on the PsyCap construct in terms of ethnicity, as well as self-reported language. In particular, the Black and White groups were found to differ significantly on Hopeful-Confidence ($p = 0.042$), with the White group scoring higher than the Black group on Hopeful-Confidence on average (mean difference= 3.6085). Moreover, the Afrikaans and traditional language groups were found to differ significantly on Resilience ($p = 0.019$), where traditional language speakers reported higher levels of resilience than Afrikaans speakers (mean difference= 2.712) (Du Plessis & Barkhuizen, 2012).

A fifth study by Staples (2014) investigated generational differences in PsyCap. The sample consisted of 347 participants who represented various ages, genders, ethnicities and employment levels in the United States (US) workforce. The results of the ANOVA tests indicated that generations did differ in terms of their Psychological Capital. In particular, Baby Boomers presented higher scores for overall PsyCap compared to the younger generations (Staples, 2014). This supported findings in prior research that older generations differ from younger generations in terms of their psychological resources (Benson & Brown, 2011; Beutel & Witting-Berman, 2008). Furthermore, the results indicated that the oldest generation presented higher Optimism in comparison to the younger generations, while no observable differences were evident between Generation X and Y (Staples, 2014). The author concluded that this corresponded with research conducted by McMurray et al. (2010) which showed that significant differences in Optimism were only observed in older generations (Staples, 2014). A limitation of the study which should be noted is that a convenience, non-probability sampling method was used and the findings in the study therefore cannot be generalised to the population.

³ It should be noted that the MANOVAs were performed on the three-factor PsyCap scale as proposed by Du Plessis and Barkhuizen (2012) discussed previously. The subscales tested included Hopeful-Confidence, Optimism and Resilience, rather than the original four-factor structure of PsyCap.

Lastly, a recent study by Lamba (2017) studied gender-based differences in PsyCap within a Research and Development (R&D) organisation⁴. The sample used in the study comprised of 140 working professionals aged between 22 and 55 years old, which consisted of 70 female and 70 male participants respectively (Lamba, 2017). The PCQ-24 data was analysed using a 't-test', which revealed significant differences between genders in terms of their PsyCap. In particular, the author concluded by stating, "the findings of the present research paper (demonstrate) that male professionals have a significantly higher level of Self-efficacy as compared to the female professional" (Lamba, 2017, p. 17), arguing that the findings are in line with previous research (Fallan & Opstad, 2016; Kirman & Jahan, 2015; Kumar & Lal, 2006; Narasimha & Reddy, 2017; Reyes et. al 2017, as cited in Lamba, 2017). The author fails to acknowledge however, that any form of measurement bias could have accounted for the observed group differences highlighted in the results.

Therefore, it is evident that the authors discussed make no reference to bias in their interpretation of results, or any investigation into potential construct, method or item bias in the assessment. In this situation, it is unclear whether the findings are as a result of actual group differences on the construct of interest, or whether bias was present in the measure (Steenkamp & Baumgartner, 1998). Consequently, the research findings are weakened when practitioners fail to acknowledge the issue of measurement invariance (Horn, 1991). This is highly concerning as when the invariance of a measure is not considered by practitioners, research findings can be misleading. This can result in severe implications on individuals and the organisation as our understanding of the nature of the construct across groups as well as our inferences regarding the group difference in the context of organisational functioning are possibly confounded (Steenkamp & Baumgartner, 1998).

The purpose of measurement is to provide valid information in order for appropriate decisions to be made regarding interventions. The quality of the decisions made therefore depend on the quality of the information received from instrument (Theron, 2017). Errors in the measurement process would affect the validity of the information obtained, which is termed bias. The following section will discuss bias in measurement, as well as introducing measurement invariance and equivalence.

⁴ The author does not state the country which the study participants originate from and refers generally to "Research & Development organisations". In light of the study findings (and in the absence of MI evidence), this renders the conclusions even more concerning, as the results are not presented for a specific country or region, but rather generalised to R & D organisations in general.

2.3 Bias, Measurement Invariance and Measurement Equivalence

2.3.1 Measurement

Psychological Capital is an abstract psychological phenomenon and as such, cannot be directly quantified or observed. I/O Psychologists measure individual's distinguishing attributes, such as PsyCap, using measuring instruments which indirectly measure the observed behaviour via test stimuli (Kaplan & Saccuzzo, 2009). This is possible as the test items stimulate a reaction from the individual (based on their past experiences) in such a nature that their behavioural responses are dependent on the individual's standing on the construct. The goal of the measuring instrument is therefore translating the intangible individual differences into quantified terms so that the information received can be used to inform decisions within the organisation (Vandenberg & Lance, 2000). The validity of the measuring instruments is therefore critically important as the quality of the interventions will hinge on the quality of the information received. Appropriate information from measuring instruments will enable effective decision-making and prediction of future behaviour (Theron, 2017). Considering South Africa's diverse population found in the workplace, an important measuring issue to be considered is the implications of cross-cultural measurement.

2.3.2 Cross-cultural Measurement

Increasing processes of globalisation and migration in the last decade have resulted in many countries across the world becoming multicultural populations (Van de Vijver & Rothman, 2004). The case is no different in South Africa, also known as the rainbow nation, due to the large variety of cultural groups and languages found in the country. Considering the diversity in today's society, and the obligations set out in the EEA, the cross-cultural applicability of measures has become an important topic in research in the field of psychometric measurement (Van de Vijver & Poortinga, 1997; Cheung & Rensvold, 2002). According to Theron (2007) psychological measurement can be considered applicable across groups, if observed scores on the measuring instrument can be interpreted in the same manner across groups; if the measuring instrument succeeds in assessing the construct as it was constitutively defined across groups; and if the inferences derived regarding the construct of interest, given a particular observed score, are the same across groups. An absence of the above may indicate potential bias in the assessment.

2.3.3 Bias in Measurement

Consider the following example, a teacher uses the following question in a class test to assess the student's mathematical ability: Calculate the following, a rugby team scores two tries and one conversion, what is their final score? To some extent, the question would give an indication of the student's mathematical ability, however there is another factor present in this case which could cause

systematic variance other than the construct of interest, namely gender. This is due to the fact that participant's knowledge of rugby would influence their ability to answer the question above mathematical ability alone, which indicates that there is bias in the assessment. Consequently, bias is evident when observed scores on the indicators of a particular construct do not correspond with changes in the construct of interest (Van de Vijver & Tanzer, 1997). Van de Vijver and Leung (2011) proposed a taxonomy to facilitate the examination of bias. Three sources of measurement of bias can be differentiated, namely construct, method and item bias.

2.3.3.1 Construct bias

Construct bias is evident when the same underlying construct is not measured across groups (Van de Vijver & Leung, 2011). Examples of this include the Chinese construct of 'Interpersonal Relatedness' in an indigenous Chinese personality instrument (CPAI-2) which does not load on any factors of the Western Big Five personality model (Cheung, Cheung, Wada & Zhang, 2003). Furthermore, Western intelligence tests traditionally do not include dimensions of social intelligence, which are more common in non-Western conceptions of intelligence (Van de Vijver & Tanzer, 2004). Consequently, if the construct 'social intelligence' were assessed across groups of individuals, the measure may succumb to construct bias (De Kock, 2018). Construct bias therefore refers to inconsistency in the attribute being measured across groups (Sireci, 2011). If construct bias is evident, further group comparisons on the construct are inappropriate due to the lack of shared meaning between two or more groups (Van de Vijver & Leung, 2011).

Three potential sources of construct bias can be identified, namely the differential suitability of item content, an incomplete overlap of construct definitions and inadequate sampling of behaviours associated with the target construct across cultures (Berry et al., 2002; Byrne & Watkins, 2003). A differential suitability of item content occurs when the items used as indicators of the target construct differ in terms of their level of appropriateness across groups (Byrne & Watkins, 2003). This can be due to differences in terms of the appropriateness of emotional expression across cultures for example, which can cause individuals from different groups to respond to item content differently (Ekman & Friesen, 1975). Furthermore, an incomplete overlap of the definition of a construct is evident when the expression of the construct manifests slightly differently in two or more cultures, i.e. the behaviours associated with the target variable are not identical (Berry et al., 2002). Similarly, the third source of construct bias relates to an inadequate sampling of behaviours associated with the target construct (Byrne & Watkins, 2003). Therefore, even if the same underlying construct exists; the test developer may not account for the full scope of the definition of the construct in their descriptors or items for the scale. This could potentially translate to construct bias as an individual's score on the

measure may not be an accurate reflection of their standing on the latent variable (Van de Vijver & Leung, 2011).

2.3.3.2 *Method bias*

Method bias derives from the measurement process or the method of measurement (Van de Vijver & Leung, 2011). It is known to affect all items in a test, rather than single items and examples of this in cross-cultural assessment include differential familiarity with item formats, test administrator bias and response sets, such as acquiescence (Sireci, 2011). Van de Vijver and Tanzer (2004) suggest three forms of method bias, namely instrument bias, administration bias and sample bias. Sample bias is evident in cases when samples are incomparable on factors other than the construct of interest, such as education levels between cultural groups, which can confound actual differences in the construct of interest (Van de Vijver & Tanzer, 2011). In particular, research has shown that a negative relationship exists between verbal ability and negatively keyed items (Marsh, 1986). As a result, respondents who are less proficient in the language being tested in, are likely to experience more difficulty responding to negatively keyed items. Furthermore, Schuttelworth-Edwards et al. (2004) explain that if participants do not understand the item content, they are less likely to respond in line with their standing on the latent variable. This could potentially be significant in the context of the present study, considering that the PCQ-24 scale comprises three negatively keyed items. The language proficiency of the sample respondents should therefore be considered, especially in cases where respondents do not indicate English as their first language. This will be highlighted in the results section.

Instrument bias includes issues that stem from characteristics of the measuring instrument. An example of this is stimulus familiarity, i.e. cultures that are more familiar with the topics or objects utilised in the assessment may be influenced positively because of it, above another culture, resulting in differences in scores unrelated to the target variable (Van de Vijver & Tanzer, 2011). The third type of potential method bias is administration bias. This can be the result of communication issues between administrator and test taker or a lesser understanding of the testing language hampering the collection of appropriate data (Van de Vijver & Tanzer, 2011).

Various sources of method bias can be identified in literature, namely differential stimulus familiarity, differential social desirability and differential response styles (Allen, 2017; Van de Vijver & Tanzer, 2004). Differential stimulus familiarity can occur when participants differ in terms of their awareness or understanding of the assessment process/ technique. This form of bias would be prevalent for example, if individuals from different cultures were assessed using a computerised testing technique, and only one of the groups was familiar with computers. The group that had previously only been

tested using a paper and pencil technique, therefore, would be disadvantaged, inducing method bias (Van de Vijver & Tanzer, 2004). The second source of method bias to be discussed is differential social desirability. This source of bias is a concern as a culturally specific response set may be elicited, due to the differences in social desirability across cultures (Byrne & Watkins, 2003; Hofstede, 1980). In a South African study, Odendaal (2015) investigated whether differences existed in social desirability scores of different ethnic groups, as a result of differing levels of cognitive ability. The results demonstrated that the Sotho- and Nguni- speaking participants scored lower on the General reasoning assessment, while scoring higher on the Social Conformity scale (the measure of social desirability in the Occupational Personality Profile [OPP]) in comparison to the English and Afrikaans speaking participants. Therefore, interestingly, social desirability was negatively related to general reasoning ability and in addition, the relationship was moderated by ethnicity (Odendaal, 2015). Significant differences in social desirability can therefore be observed across cultures (Middleton & Jones, 2000).

Furthermore, Van de Vijver and Tanzer (2004) explain that method bias can often lead to a shift in average test scores. They state, “stimulus familiarity and social desirability tend to influence all items of an instrument and, hence, they will induce a change in average scores. Such a change may occur independently of possible cross-cultural differences on the target variable” (Van de Vijver & Tanzer, 2004, p. 121). The last source of method bias to be considered, is differential response styles. According to Allen (2017, p. 1475), response styles are distinctive ways of responding to questionnaire surveys that are unrelated to the content of the actual survey items. One form of response style is known as acquiescence, i.e. where respondents may agree with all items in a measure, even when items have meanings which contradict each other (Allen, 2017; Suárez-Alvarez et al., 2018; van Sonderen et al., 2013). On the other hand, some participants may favour more extreme responses, either strongly agreeing or disagreeing with most scale items. Lastly, participants may select mostly moderate responses to statements (Allen, 2017; Suárez-Alvarez et al., 2018; van Sonderen et al., 2013). Response style bias has been widely studied in cross-cultural research (i.e. Hofstede, 1980, 2001; Sekeran, 1983). For example, Sekeran’s (1983) study showed that cultural factors influenced Asian participants response styles, as the cultural convention of being courteous lead to a desire to please investigators, which impacted item responses. Evidently, observed group differences can be greatly affected by different forms of method bias, which could impact the validity of inferences derived from a measure (Van de Vijver & Tanzer, 2004).

2.3.3.3 Item bias

When the effects of bias factors present in one or many items, item bias or DIF is said to be present (Berry, 2015). Item bias can be evident as a result of poor translation of an item or if an item is

inappropriate for a particular context, resulting in a lack of equivalence (Van de Vijver & Leung, 1997; Fontaine, 2008). Thus, an item will be considered biased if individuals who have the same standing on the latent variable, differ on their expected scores on the item, i.e. individuals with equal total scores on intelligence should have the same responses/experience to items; different means would indicate item bias (Van de Vijver & Tanzer, 2011).

When investigating item bias, three forms can be distinguished, namely uniform, non-uniform and error-variance bias (Dunbar et al., 2011; Wu et al., 2007). Uniform bias is present when the intercepts of the regression of the indicators on the latent variable, differ across groups (Wu et al., 2007). This provides evidence of a significant group membership main effect, i.e. group membership explains significant variance in item responses, not explained by the latent variable (Byrne & Watkins, 2003; Van de Vijver & Poortinga, 1997). Non-uniform bias on the other hand, occurs when the slopes of the regression of the indicators on the latent trait differ across groups. Evidence of non-uniform bias therefore shows a significant group membership x latent variable interaction effect; indicating that group membership moderates the effect of the latent dimension on the item responses (Dunbar et al., 2011; Wu et al., 2007). Lastly, error variance bias indicates that factors, unrelated to the target variable, are explaining significant variance in item responses, not explained by the construct in question (Wu et al., 2007). If prevalent, these forms of item bias can be concerning as they can potentially impact the inferences derived from a measure (Theron, 2007). The following section will highlight the approach that was employed to identify the various forms of bias in the PCQ-24.

2.3.4 Measurement Invariance/Equivalence (MI/ME)

“Whereas bias refers to construct-irrelevant sources of variance, equivalence may be understood as a lack of group-related bias” (De Kock, 2018, p.4). The term equivalence is commonly used in the context of cross-cultural issues and the meaning of constructs, while invariance is used from a pure measurement perspective (De Kock, 2018). MI and ME testing provides a rigorous manner for assessing measurement bias across groups, which is a prerequisite in order for meaningful comparisons to take place (Vandenberg & Lance, 2000). While some authors use the terms MI and ME interchangeably, Dunbar et al. (2011) make an important distinction between the two, summarised in their invariance / equivalence taxonomy. According to Dunbar et al. (2011) measurement invariance research examines whether a multi-group measurement model with varying degrees of parameter constraints fits the data for two or more groups. The constraints vary from none, to some and finally all the parameters are constrained to be equal, where strictly speaking, the measurement model should demonstrate exact fit to claim that MI has been achieved. Despite this ideal, exact fit is seldom obtained in social science research (Brown & Cudeck, 1993). Alternatively, a less stringent stance is

that a multi-group measurement model demonstrates at least close fit, where certain parameters are constrained to be equal. In this case, the position that measurement invariance has been demonstrated is permissible as the close fit null hypothesis was not rejected. Therefore, the multi-group measurement model reproduced the observed covariance matrix to an acceptable degree (Dunbar et al., 2011).

Measurement equivalence, within the Dunbar et al., (2011) taxonomy, on the other hand refers to whether a multi-group measurement model with some of its parameters constrained to be equal across groups fits the data significantly poorer than a multi-group measurement model with fewer parameters constrained to be equal across groups (Dunbar et al., 2011, p.6). If this was the case, it would imply that significant differences are evident in one or more measurement model parameters across the groups, resulting in a poorer fit when these are constrained to be equal.

Although the awareness of the importance of testing MI and ME is slowly increasing in cross-cultural research, a review of literature conducted by Vandenberg and Lance (2000) revealed that a lack of consensus still exists regarding procedures for testing MI/ME. Specifically, inconsistencies emerged regarding the various tests conducted to establish different forms of MI, terminology used to describe various MI tests and the sequence in which the tests were conducted. In response to this, Vandenberg and Lance (2000) integrated their findings to propose a consolidated paradigm for conducting MI tests. According to the authors, there was general consensus in literature that the omnibus test of the equality of covariance matrices is the necessary first step in MI testing (Alwin & Jackson, 1981; Bagozzi & Edwards, 1998; Byrne et al., 1989; Cole & Maxwell, 1985; Horn & McArdle, 1992; Jöreskog, 1971; Rock et al., 1978; Schaie & Hertzog, 1985; Steenkamp & Baumgartner, 1998). This test assesses the, “null hypothesis of invariant covariance matrices” (Vandenberg & Lance, 2000, p. 12). Researchers (e.g., Alwin & Jackson, 1981; Bagozzi & Edwards, 1998; Cole & Maxwell, 1985; Jöreskog, 1971; Mulaik, 1975; Steenkamp & Baumgartner, 1998) argued that MI would be established if the covariance matrices did not differ across groups, indicating overall measurement equivalence; hence, further tests of MI and ME would be unwarranted if this level of MI was achieved. Therefore, if invariant covariance matrices were not established, further investigation would ensue (Vandenberg & Lance, 2000).

Furthermore, Vandenberg and Lance (2000) found substantial agreement in literature that the configural invariance test should be the second step of MI testing. This is since the configural model is necessary in order to evaluate subsequent, more restricted MI models. More importantly, the configural invariance test assesses whether the same underlying construct is operationalised within the respective groups. Therefore, should this level of MI not be achieved, it would suggest that the

items are not measuring corresponding constructs across groups, making further cross-group comparisons unsubstantiated (Bialosiewicz et al., 2013; Gunn, 2016). Vandenberg and Lance (2000) explained however, that following the omnibus and configural invariance tests, little agreement existed in literature regarding the sequencing of tests of MI and ME.

Of the tests reviewed, metric invariance, scalar invariance and invariance of unique variances were suggested by Vandenberg and Lance (2000) as the next steps in assessing MI. In particular, metric invariance assesses whether the factor loadings of the items are invariant across groups (Bialosiewicz et al., 2013). Achieving metric invariance would indicate that factor loadings do not differ across groups and thus group comparisons in terms of factor variances and covariances would be warranted (Bialosiewicz et al., 2013). Should metric invariance not be attained however, it would imply the presence of non-uniform bias (Barendse et al., 2015; Fontaine, 2008).

Moreover, scalar invariance is a test of invariant item intercepts across groups (Vandenberg & Lance, 2000). Should scalar invariance be achieved, comparisons of groups in terms of their latent mean scores on the measure would be justified (Bialosiewicz et al., 2013; Van de Schoot et al., 2012). As configural, metric and scalar invariance would have been shown, it could be considered sufficient evidence to suggest that the same underlying construct is measured across the groups and any observed differences in scores relate to actual differences in the construct of interest (Selig, Card & Little, 2008). Should a lack of invariance at the scalar level be found however, it would suggest the presence of uniform bias (Barendse et al., 2015; Fontaine, 2008).

Lastly, invariance of unique variances (Also termed strict invariance/ full-uniqueness/ strict factorial invariance [Steyn & de Bruin, 2020]), relates to residual or error variance in items across groups (Bialosiewicz et al., 2013; Vandenberg & Lance, 2000). This test therefore assesses the presence of systematic errors relating to the measurement of the construct, which can include unique errors associated with certain items (Steyn & de Bruin, 2020). According to Vandenberg and Lance (2000), the test of invariance of unique variances is seldom attained in practice. Moreover, as it is a highly restrictive model, authors have suggested that this level of MI is unreasonable (Bialosiewicz et al., 2013; Byrne, 2009; Chen et al., 2009; Vandenberg & Lance, 2000). Due to the lack of a standardised process for testing MI, the present study decided to implement the taxonomy presented by Dunbar et al. (2011). The taxonomy was based on the work of Mels (2010) and Meredith (1993), and aimed to contribute toward a more consistent understanding and method of assessing MI.

2.3.5 Taxonomy of Measurement Invariance and Equivalence

Table 2.1 depicts the taxonomy which Dunbar et al. (2011) recommended for the analysis of measurement invariance, i.e. testing the extent to which multi-group measurement models fit the data across groups when increasing constraints are introduced.

Table 2.1:

Degrees of MI

Configural invariance	Weak invariance	Strong invariance	Strict invariance	Complete invariance
A multi-group measurement model in which the structure of the model is constrained to be the same across groups, fits multi-group data.	A multi-group measurement model in which the structure of the model is constrained to be the same across groups and in which the factor loading matrix (Λ^x) is constrained to be the same across groups, fits multigroup data.	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups and in which the vector of regression intercepts (ξ^x) is constrained to be the same across groups, fits multigroup data.	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups, in which ξ^x is constrained to be the same across groups and in which the measurement error variance-covariance matrix (Θ_δ) is constrained to be the same across groups, fits multi-group data	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups, in which ξ^x is constrained to be the same across groups, in which Θ_δ is constrained to be the same across groups and in which the latent variable variance covariance matrix (Φ) is constrained to be the same across groups, fits multigroup data.

Note: Degrees of Measurement Invariance. Reprinted from "A cross-validation study of the Performance Index", by Dunbar et al. 2011, *Management Dynamics*, 20(3), p. 2-24.⁵

As a result of a lack of agreement regarding accepted terminology for the levels of ME, Dunbar et al. (2011) presented the stages depicted in Table 2.2. These can assist practitioners in analysing the extent to which a multi-group measurement model with increasing parameter constraints fits significantly poorer than a multi-group measurement model where these are freely estimated, while the structure is constrained across groups. From the terms used in Table 2.2, scalar and metric equivalence are generally accepted terms in literature. The term 'conditional probability equivalence' however was created by Dunbar et al. (2011), referring to strong evidence that the groups did not differ in terms of error (residual) variance of the regression of χ_i on ξ_1 .

⁵ Permission to reprint the table was obtained from the editor of the journal.

Table 2.2:*Levels of ME*

Metric equivalence	Scalar equivalence	Conditional probability equivalence	Full equivalence
A multi-group measurement model in which the structure of the model is constrained to be the same across groups and in which the factor loading matrix (Λ^x) is constrained to be the same across groups does not fit multi-group data poorer than a multi-group measurement model in which the structure of the model is constrained to be the same across groups, but all model parameters are freely estimated (i.e., the configural invariant multi-group model).	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups and in which the vector of regression intercepts (ξ^x) is constrained to be the same across groups does not fit multigroup data poorer than a multigroup measurement model in which the structure of the model is constrained to be the same across groups, but all model parameters are freely estimated.	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups, in which ξ^x is constrained to be the same across groups and in which the measurement error variance covariance matrix (Θ_δ) is constrained to be the same across groups does not fit multi-group data poorer than a multi-group measurement model in which the structure of the model is constrained to be the same across groups but all model parameters are freely estimated.	A multi-group measurement model in which the structure of the model is constrained to be the same across groups, in which Λ^x is constrained to be the same across groups, in which ξ^x is constrained to be the same across groups, in which Θ_δ is constrained to be the same across groups and in which the latent variable variance-covariance matrix (Φ) is constrained to be the same across groups does not fit multi-group data poorer than a multi-group measurement model in which the structure of the model is constrained to be the same across groups, but all model parameters are freely estimated

Note: Degrees of Measurement Equivalence. Reprinted from "A cross-validation study of the Performance Index", by Dunbar et al. 2011, *Management Dynamics*, 20(3), p. 2-24.⁶

Similarly, the forms of equivalence will be evaluated hierarchically, starting from metric equivalence, once the configural invariance (baseline model) has demonstrate close fit. In line with this taxonomy, obtaining a certain level of MI would therefore indicate that a measurement model with a certain degree of constrained parameters shows close fit in that it provides a satisfactory account of the observed covariance matrices. Furthermore, obtaining a certain level of ME would indicate that a measurement model with a certain degree of constrained parameters does not fit the data poorer than the baseline model (Dunbar et al., 2011).

Evidently, the Dunbar et al. (2011) taxonomy provides a comprehensive approach for assessing MI and ME in comparison to the recommendations by Vandenberg and Lance (2000). This is due to the fact that the tests cover invariance / equivalence at the factor structure (configural/baseline), factor

⁶ Again, permission to reprint the table was obtained from the editor of the journal.

loading (weak/metric), intercepts (strong/scalar) and error variance (strict/conditional probability) levels (Dunbar et al. 2011; Vandenberg & Lance, 2000). The Dunbar et al. (2011) methodology, however, does not account for the first MI test recommended by Vandenberg and Lance (2000), namely the omnibus test of the equality of covariance matrices. Dunbar et al. (2011) questioned the value of the omnibus test, explaining that it has been shown to indicate full equivalence in cases where further MI tests have indicated a lack of equivalence at the metric, scalar and conditional probability levels. As a result, confidence in the findings of the omnibus are significantly weakened and subsequent MI tests are still necessary in order to confirm that MI has been achieved at the factor loading, intercept and error variance level. Performing the omnibus test therefore yields little meaning to the researcher, and in line with Dunbar et al. (2011) methodology, the omnibus test of the equality of covariance matrices was not performed in the present study.

2.3.6 Assessing Measurement Invariance and Measurement Equivalence

This begs the question, however, how achievement of a respective level of MI / ME is judged? MI and ME can be assessed by using CFA through a series of analyses (Dunbar et al. 2011). More stringent analyses are only conducted when the preceding analyses provided successful outcomes, i.e. the measurement models obtained at least close fit (Vandenberg & Lance, 2000). Various statistics are employed when comparing the constrained and unconstrained multi-group measurement models. The first is the chi square (χ^2) difference test which evaluates the statistical significance of the difference between two nested measurement models. However, as a result of the sensitive nature of the test, minor variances may result in rejection of the null hypothesis (Cheung & Rensvold, 2002; Dunbar et al., 2011; Van de Vijver & Leung, 2011).

Consequently, many practitioners use other goodness-of-fit indices (GOF) for investigating MI and ME. A study by Cheung and Rensvold (2002) indicated that reporting changes in the CFI, Gamma Hat and McDonald non-centrality index were sufficient to indicate significant differences in multi-group measurement models. In conjunction to these, Meade et al. (2008) suggest reporting changes in the χ^2 likelihood ratio test. Due to the sensitivity of the chi square difference test to sample size however, all the aforementioned GOF statistics will be reported in the present study. A comprehensive review of the respective GOF indices and cut-off values which were utilised in the current study will be discussed in Chapter 3. Should close fit not be achieved and the alternative GOF indices indicate mediocre to poor model fit; the sources of the potential misfit would need to be investigated. This would require the inspection of partial measurement invariance and equivalence (Vandenberg & Lance, 2000).

2.3.6.1 Partial Measurement Invariance and Equivalence

In MI and ME studies, the ideal would be to obtain strict invariance and conditional probability equivalence (Dunbar et al., 2011). Achieving these results, however, are not commonly achieved in practise, as assessments may be invariant or equivalent in certain samples, but not in others (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). As discussed, multigroup measurement models need to demonstrate configural invariance prior to further analyses of MI and ME taking place. As it is occasionally difficult to obtain weak invariance and metric equivalence, many authors have recommended less stringent evaluations, favouring partial MI and partial ME, thereby allowing cross-group comparisons to still take place (Byrne et al., 1989; Vandenberg & Lance, 2000). Consequently, partial MI and ME enables cross-group evaluations which would not have been suitable otherwise. It should be noted that partial invariance or equivalence will only be warranted should the corresponding level of MI or ME not be achieved, in which case the source of variance needs to be identified (Vandenberg & Lance, 2000). This process will enable the researcher firstly to identify items suffering from uniform, non-uniform or error variance bias and secondly to decide which parameters should be freed to be estimated in order to enable the evaluation of invariance/ equivalence (Vandenberg & Lance, 2000).

Analyses of partial MI and ME involves the lessening of equality constraints placed on the measurement models. Through this process, invariant parameter estimates are freed to be estimated until the measurement model obtains close fit and the fit between the partially constrained measurement model and the baseline model is no longer significant. While Byrne et al. (1989), recommended the first procedure for evaluating partial MI and ME nearly thirty years ago, disagreement still exists regarding best practices in partial MI and ME analyses (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). In particular, researchers differ in terms of the statistical criteria that should be considered when lessening invariance constraints [e.g. goodness-of-fit indices, expected parameter changes and modification indices. Furthermore, some have argued that the statistical criteria are not applied in a consistent manner (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). The impact of the lack of agreement is a deficit of literature regarding partial MI and ME as well as the conceptual and statistical implications thereof (Putnick & Bornstein, 2016). Differing opinions also exist regarding the number of invariant items that can be freed while claiming partial MI or ME. Some authors proposed freeing parameter constraints until only the reference indicator and one other item per factor remain (Steenkamp and Baumgartner, 1998). Vandenberg and Lance (2000) argue however, that a drawback of this method is that inappropriate evaluations of mean group differences will be conducted on non-comparable measures. Conversely, they propose that constraints should only be lessened under the following conditions, when: (a) a thorough theoretical argument supports the decision; (b) only few problematic items have been

identified; and (c) when cross-validation evidence supports the feasibility of lessening the constraints on the model (Vandenberg & Lance, 2000).

Literature suggests that invariant items can be detected in numerous ways (Byrne et al., 1989; Cheung & Rensvold, 1999, Davis, 2014; Putnick & Bornstein, 2016; Steenkamp & Baumgartner, 1998; Yoon & Kim, 2014). These methods assist in the identification of items suffering from various forms of bias (non-uniform, uniform and error variance bias). The first approach is applicable to multigroup measurement models which comprise of numerous different constructs, and aids in the identification of indicators which lack invariance (Byrne et al., 1989; Cheung & Rensvold, 1999). This approach specifies that the measurement model for each construct is tested separately by constraining the relevant factor loadings, intercepts or error variances, while all the other measurement model terms are freely estimated (Cheung & Rensvold, 1999). A comparison is then performed between the individual models and the configural invariance model in terms of their fit to the data. If the respective measurement model fits (practically or statistically) significantly poorer than the baseline model, it indicates that the respective construct contains one invariant item at the minimum and the affected subscale should be reported by the researcher (Cheung & Rensvold, 1999).

The process then continues by evaluating the measurement models at the item level (Byrne et al., 1989). This is achieved by creating measurement models for each individual subscale in order to establish each item's contribution to the lack of invariance. Again, the model's fit is compared to the configural model and it is assessed whether the model's fit is significantly poorer, either from a statistical or practical perspective. Items can be sequentially released in two ways, namely the forward method and the backward method. The former method involves an iterative process through which item constraints are added, while the latter method involves releasing item constraints on a model which has all its items constrained initially (Putnick & Bornstein, 2016; Yoon & Kim, 2014).

The second method to be discussed assesses the significance of factor loadings between the single group measurement models (Cheung & Rensvold, 1999). This method is questioned however, as significant differences can be observed in parameter estimates even in cases where they are both statistically significant. Furthermore, it is also possible for similar significance values to be observed across the groups while only one is found to be statistically significant (Cheung & Rensvold, 1999; Davis, 2014). Thirdly, items lacking invariance can be identified through examining the factor loadings of items which demonstrated large differences between groups in terms of the configural invariance model (Cheung & Rensvold, 1999). The factor loading, intercept or error variance with the largest difference between the groups is freed to be estimated and the fit of the model is consequently assessed. Significance tests are not applied to evaluate item invariance here, instead the model fit is

evaluated, and should close fit not be obtained, the items will continue to be freed sequentially until close fit is attained (Cheung & Rensvold, 1999). Achieving partial weak MI and partial metric ME enables the evaluation of partial strong MI and partial scalar ME, where the same methodology can be utilised. Items that have been freed to be estimated at a preceding level of measurement invariance or equivalence will continue to be freed in the following levels of analyses (Steenkamp & Baumgartner, 1998).

The present study utilised the process discussed in the third technique, whereby the configural invariance model's absolute differences in the completely standardised factor loadings (or intercepts/error variances) were calculated and rank ordered. Through this process, the items reflecting the largest absolute differences in their factor loadings, intercepts or error variances in the respective analyses were freed to be estimated to allow the evaluation of partial MI and partial ME where necessary (Cheung & Rensvold, 1999).

Once particular items have been shown to be problematic, the focus then shifts to how the researcher should approach these. Different methods of dealing with problematic items can be considered, such as accepting and interpreting data from non-invariant items as cross-group data; deleting items which may be resulting in the model not obtaining close fit and consequently lacking invariance, and lastly, retaining problematic items by allowing partial invariance and partial equivalence to be evaluated (Cheung & Rensvold, 1999). While ignoring the items causing the lack of invariance might seem appealing, practitioners who make inferences based on respondents' results do so on the basis of their dimension scores, not on the item level (Theron, 2007). Therefore, the cumulative impact of the items lacking invariance on the dimension level needs to be considered. The noninvariant items may have differing effects on the different groups, for example, particular items may benefit the Black group, causing the results to favour one group over the others. The invariant items could also act inconsistently, resulting in the bias cancelling out on the dimension level for a particular group, and these cumulative effects should be investigated and understood by the practitioner (Theron, 2007). As the aim of the present study was to evaluate the MI and ME of the PCQ-24, the third option of dealing with biased items was favoured, namely retaining problematic items through evaluating partial MI and partial ME. This technique enables cross-group comparisons in the latent means in measures which demonstrate bias (Cheung & Rensvold, 1999).

2.4 Conclusion

This section provided an overview of the PCQ-24, its development, subscales and psychometric properties. It also covered current research regarding the scale focusing on group differences. Furthermore, the chapter introduced bias in its various forms, followed by a discussion of MI and ME

and the taxonomy that will be used in evaluating MI and ME. The following section will focus on the research methodology that will be used in assessing the invariance and equivalence of the Psychological Capital scale.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

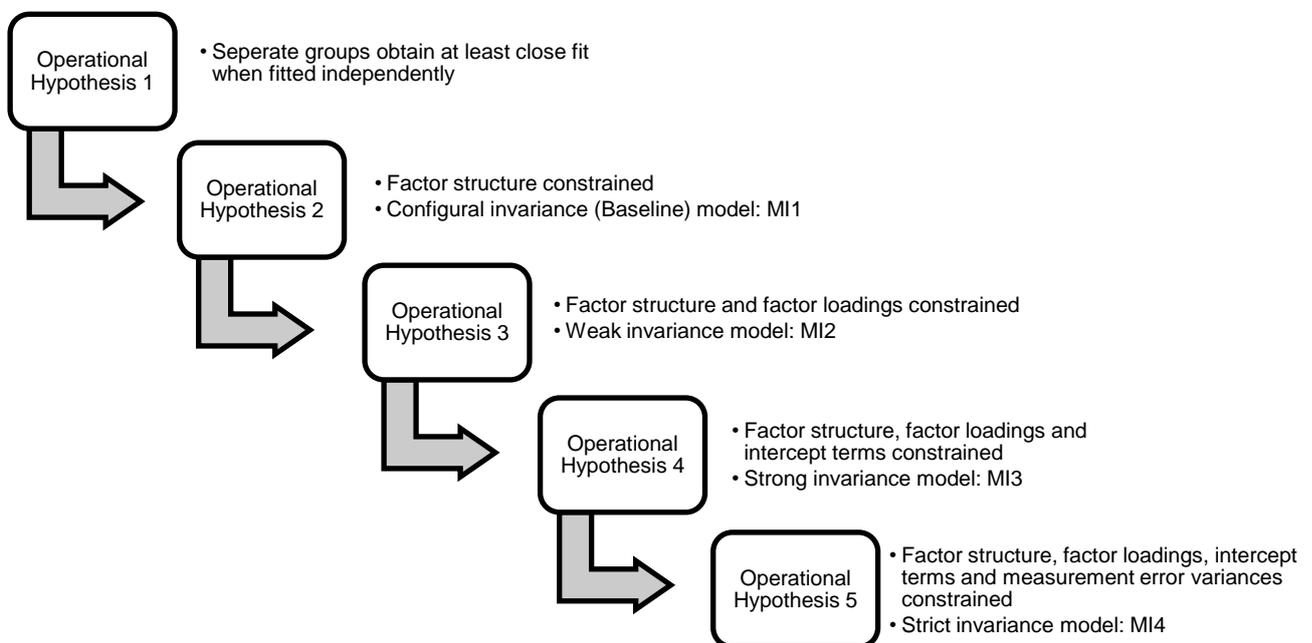
This chapter will discuss the research methodology used in the present study. Firstly, the substantive research hypotheses will be discussed, followed by the research design that was applied in testing the substantive research hypotheses. The chapter will also provide the statistical hypotheses, discuss the sample for the study, as well as the statistical analyses that were used to assess the MI and ME of the PCQ-24.

3.2 Substantive Research Hypothesis

The overarching substantive research hypothesis tested in the present study is that the PCQ-24 measures the Psychological Capital construct in a valid and reliable manner, and that the measure is not biased, i.e. the factor loadings, intercepts of the regression of the indicator variables on the latent variables, as well as the measurement error variances, are invariant across Black, White and Coloured groups in South Africa. Consequently, the operational hypotheses that were tested in the study are depicted below.

Figure 3.1

Analyses performed in testing MI: Operational hypotheses 1 - 5



The operational hypotheses involved in testing for MI are performed through five sequential analyses, whereby the results for both the single- and multi-group measurement models are required to obtain reasonable fit in the preceding step, to enable the next analyses to be performed (Dunbar et al., 2011). The Goodness of Fit (GOF) indices used in assessing model fit are described in paragraph 3.3.1. If the single group models attained close fit, a multi-group measurement model was fitted to the data, in which only the factor structure has been constrained across groups. If this model demonstrates close fit, stricter constraints are placed on the model in the subsequent levels of MI analyses. These additional constraints include the factor loadings, intercept terms and the measurement error variance terms, which are constrained to be equal across groups while the rest of the measurement model parameters are estimated without restrictions (Dunbar et al., 2011).

Achieving a particular level of MI in the study enables the evaluation of the corresponding ME level to be performed. In these analyses, the Configural invariance (MI1) model is used as the baseline model which the subsequent MI levels are assessed against. The sequence of analyses tested in operational hypotheses 6 – 8 are depicted in Table 3.1 below. In operational hypotheses 6 – 8, each successive level of MI is compared to the MI1 (configural invariance) model and the level of equivalence is considered to be achieved if the corresponding invariance model does not fit the sample data significantly poorer than the baseline model (Dunbar et al., 2011). The GOF indices that were considered in evaluating ME are discussed in detail in paragraph 3.4.1.

Table 3.1

Analyses performed in testing ME: Operational hypotheses 6 - 8

MI Models	Measurement models compared	ME Models tested	Hypotheses tested
Configural Invariance model (MI1)	-	-	-
Weak invariance model (MI2)	MI2 – MI1	Metric equivalence: ME1	Operational hypotheses 6
Strong invariance model (MI3)	MI3 – MI1	Scalar equivalence: ME2	Operational hypotheses 7
Strict invariance model (MI4)	MI4 – MI1	Conditional probability equivalence: ME3	Operational hypotheses 8

3.3 Research Design

The research design determines the strategy and overall structure that will be utilised in testing the operational hypotheses in a study, including the procedures followed, the data collected, and the data analysis conducted by the researcher (Leedy & Ormrod, 2005). The intention of the research design is to assist in answering the research initiating question by producing unambiguous evidence to test the operational hypotheses (Kerlinger & Lee, 2000). The extent to which this is achieved however,

depends on the chosen design's ability to control variance, which requires, "(1) Maximising variances in variables from the hypothesis; (2) Controlling variance of extraneous or unwanted variables that may have an effect on the experimental outcomes; and (3) Minimising sampling errors or random variance, including errors of measurement" (Kerlinger & Lee, 2000; p. 456). The research design is therefore of paramount importance as the credibility of the research findings rest on the interpretation of the results derived from the study.

The most appropriate research design, based on the intention of the study, was decided upon after reviewing the three most prominent designs in social sciences, which include experimental design, *ex post facto* (or retrospective) quasi-experimental design, as well as *ex post facto* correlational design (Babbie, 2013; Cohen et al., 2000; Kerlinger & Lee, 2000). Of the three potential designs, experimental designs are praised for their ability to control variance as these studies are conducted in controlled environments which involve randomly assigning sample participants and experimental manipulation of variables (Babbie, 2013; Kerlinger & Lee, 2000).

The present study did not warrant the use of an experimental design as the goal of the study was not to manipulate the PsyCap constructs via experimentation. An *ex post facto* correlational design was rather used, through observing manifestations retrospectively. This research design was appropriate as the goal of the present study was not to change variables, but rather to compare different Black, White and Coloured groups' data from the PCQ-24. A limitation of this design is that it does not enable the researcher to manipulate variables via experimentation, however there are many advantages to using the *ex post facto* correlational design. These include that it can be used in circumstances where an experimental approach is not feasible or appropriate. *Ex post facto* designs are also a useful exploratory tool, due to the valuable information they produce regarding the nature of phenomena (Cohen et al., 2000).

The focus of the current study was not to investigate a traditional structural model, but rather a series of multi-group measurement models. According to Moyo and Theron (2011) measurement models assume that the regression of indicator variables on latent variables follow a positive, statistically significant slope. Here, the dependent variables are the observable indicator variables and the independent variables are the person constructs. The operational hypotheses in the study, therefore, investigated whether the factor structure, item responses and the slopes thereof, regress on the PsyCap dimensions in the same manner across the Black, White and Coloured groups. The hypotheses associated with the PsyCap measurement models were tested via the *ex post facto* correlational design shown in Figure 3.2.

Figure 3.2*Ex post facto correlational design*

$[X_{11}]$	$[X_{12}]$...	$[X_{1j}]$...	$[X_{1,24}]$
$[X_{21}]$	$[X_{22}]$...	$[X_{2j}]$...	$[X_{2,24}]$
		
$[X_{i1}]$	$[X_{i2}]$...	$[X_{ij}]$...	$[X_{i,24}]$
:	:	...	:	...	:
$[X_{n1}]$	$[X_{n2}]$...	$[X_{nj}]$...	$[X_{n,24}]$

In the design depicted above, the individual items of the PCQ-24 were used as indicators, with 4 exogenous latent variables ξ and 6 items per subscale, ranging from X_{i1} to $X_{i,24}$; $i=1, 2, \dots, n$. X_{ij} refers to participant i 's score on item j .

When using this research design, the measurement model is tested by defining the study variables to be tested (i.e. the items and indicators of the PCQ-24) and calculating the observed inter-item covariance matrices via LISREL 8.8. LISREL obtains estimates for the freed measurement model parameters with the aim of accurately reproducing the observed covariance matrices (Diamantopoulos & Sigauw, 2000). In the event that the fitted measurement model was not able to accurately reproduce the observed covariance matrix, it would indicate that measurement model was not able to provide an acceptable account of the observed covariance matrix (Byrne, 1998; Kelloway, 1998).

3.3.1 Evaluating Measurement Model Fit

As no single statistical significance test can be used to evaluate the fit of a model in isolation, measurement model fit is evaluated by considering several different criteria. Schermelleh-Engel et al. (2003) explain that for each estimation procedure conducted by the researcher, several GOF indices are produced by LISREL, which are used as the criteria to judge the extent which the model fits the data. The GOF indices that were evaluated to judge the fit of the measurement models in the present study are discussed below.

3.3.1.1 χ^2 Test Statistic

The χ^2 test assesses the research hypotheses by evaluating the suitability of a structural equation model (Schermelleh-Engel et al., 2003). Generally, a measurement model is claimed to differ significantly from the data when the χ^2 values are high, when compared to the number of degrees of freedom. Furthermore, Loehlin and Beaujean (2017) explain that a χ^2 statistic which is relatively equal to the number of degrees of freedom in the model (for reasonable sample sizes) is indicative of satisfactory model fit. The exact fit null hypothesis for the single-group measurement models will be

accepted if the p-values related to the χ^2 statistic are ≥ 0.05 . Furthermore, the χ^2 difference test is used in the interpretation of comparisons of different multigroup measurement models. The null hypothesis of no difference in fit posits that no difference in fit exists between two nested multigroup measurement models. If the probability of observing the normal theory chi-square difference in a multigroup sample under the null hypothesis of no difference in fit was smaller than or equal to 0.05, the null hypothesis would be rejected in favour of the hypothesis that the fit of the multigroup measurement models differ in the parameter (Schermelleh-Engel et al., 2003; Loehlin & Beaujean, 2017).

The χ^2 statistic however has limitations which need to be considered, namely; the assumptions on which the test is based – obtaining a sufficiently large sample and a multivariate normal distribution – are seldom met in practice. Secondly, as more parameters are added to the model, the χ^2 statistic is known to decrease. As a result, models which are highly complex with a larger number of model parameters may display lower χ^2 values as their degrees of freedom are lower. In these cases, the χ^2 statistic may indicate a satisfactory fitting model, not because the model is correctly specified but rather because the model is over-parameterised (Schermelleh-Engel et al., 2003). A further limitation is the χ^2 statistic's sensitivity to the size of the sample. As the sample size increases, the χ^2 value also rises. Consequently, the test may display significant differences resulting in the rejection of models in cases where only trivial differences existed. In cases with small enough samples on the other hand, the statistic may display non-significant differences even though the discrepancy between the model and the data are highly evident (Schermelleh-Engel et al., 2003; Loehlin & Beaujean, 2017). In light of the shortcomings of the χ^2 test, further GOF indices are considered.

3.3.1.2 Root Mean Square Error of Approximation (RMSEA)

The RMSEA is another GOF index which is used to assess overall model fit (Schermelleh-Engel et al., 2003). When interpreting the RMSEA, values of 0.05 or less are considered to be indicative of good fit, while values of 0.08 or less are generally considered to indicate reasonable fit. Moreover, RMSEA values between 0.08 and 0.10 are considered to indicate mediocre fit, with values greater than 0.10 interpreted to indicate poor fit (Browne & Cudeck, 1993; Schermelleh-Engel et al., 2003; Loehlin & Beaujean, 2017). Furthermore, when considering the RMSEA we also inspect the p-value for RMSEA for close fit. When the p value is greater than 0.05, then close fit has been achieved, hence we do not reject the null hypothesis of close fit (Schermelleh-Engel et al., 2003).

3.3.1.3 Root Mean Square Residual (RMR) and Standardized RMR (SRMR)

Also known as 'fitted residuals', the RMR and SRMR display the inconsistencies between the model and the data once the parameters have been estimated in LISREL. RMR values which are close to zero

are considered indicative of good fit, however Schermelleh-Engel et al. (2003) explain that due to the RMR's dependency on the size of the variance and covariance of the variables, one cannot state that a particular RMR value suggests good or bad fit without considering the scales of the variables. To combat this shortcoming, the SRMR was created by dividing the residuals by the standard deviations. Similar to the RMR, SRMR values close to zero are considered to indicate good fit (Schermelleh-Engel et al., 2003). Hu and Bentler (1995) explain however that as a general rule, values of 0.08 and less can be considered to indicate good fit, while values less than 0.10 are considered acceptable.

3.3.1.4 Normed Fit Index (NFI) and Non-normed Fit Index (NNFI)

The Normed Fit Index (NFI) and the Comparative Fit Index (CFI) relate to a model's fit in relation to the fit of a baseline model. With comparison indices, the baseline model is a highly restrictive model, and the aim is for the target model in the study to be an improvement from the baseline (Schermelleh-Engel et al., 2003).

When considering the NFI index, higher values suggest that the target model is an improvement from the baseline model. While values can range from zero to one, and one being the highest possible value, Kaplan (2000) explains that values of ≥ 0.95 can be considered good fit, while values of ≥ 0.97 will be considered very good fit. Moreover, values of 0.90 and above will still be considered acceptable. A limitation of this index however is that it may underestimate the fit of a model in cases with small sample sizes (Schermelleh-Engel et al., 2003). Bentler and Bonnett (1980) addressed this limitation, by creating the Nonnormed Fit Index (NNFI). NNFI adjusted the NFI by including the degrees of freedom in the model. This adjustment corrected the issue of underestimation of model fit in cases where models fit extremely well, however since the index is not normed, the values can vary greater than the 0 – 1 range (Tabachnick & Fidell, 2013). Schermelleh-Engel et al. (2003) therefore advise that values of ≥ 0.95 are interpreted as acceptable while values of ≥ 0.97 are considered to indicate good fit. Again, values of ≥ 0.90 will still be considered acceptable.

3.3.1.5 Comparative Fit Index (CFI)

Lastly, the CFI will be considered, which is another index that is less affected by sample size (Hu & Bentler, 1995; Schermelleh-Engel et al., 2003). According to Hair et al. (2010, p. 643), the CFI is one of the most widely used fit indices due to its many desirable properties, including its relative (although not complete) insensitivity to model complexity. The CFI is normed and therefore values range from 0 to 1. Generally, while values of ≥ 0.90 are considered acceptable, values of ≥ 0.95 are considered good, and values of ≥ 0.97 are considered to indicate very good fit (Hu & Bentler, 1995; Schermelleh-Engel et al., 2003; Tabachnick & Fidell, 2013).

3.4 Statistical Hypotheses

The operational hypotheses were tested by using structural equation modelling via LISREL 8.8. The PCQ-24 was designed in such a manner that an individual's score on the various PsyCap dimensions influence their scores on corresponding items. The relationships of these corresponding items and dimensions are represented in the PsyCap measurement model. Furthermore, hypotheses were formulated regarding the pattern of responses on the measure across the three samples.

The first statistical hypotheses formulated relate to the fit of the measurement model, which assess the degree to which the hypothesised measurement model corresponds with the empirical data (Diamantopoulos & Sigauw, 2000; Schermelleh-Engel et al., 2003). As discussed, the overarching substantive research hypothesis tested in the present study is that the PCQ-24 measures the Psychological Capital construct in a valid and reliable manner and is not biased against Black, White or Coloured groups in South Africa. Furthermore, operational hypothesis 1 claimed that the single-group measurement models can reproduce the observed covariances among the individual items in the Black, White and Coloured samples closely. The hypothesis stating that the proposed measurement models provide an exact account of the process that generated the observed inter-item covariance matrix in the parameter can be depicted as follows, where Σ represents the population covariance matrix and $\Sigma(\Theta)$ represents the covariance matrix reproduced by the fitted model (Kelloway, 1998).

$$H_{01i}: \Sigma = \Sigma(\Theta); i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$$

$$H_{a1i}: \Sigma \neq \Sigma(\Theta); i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$$

Alternatively, Browne and Cudeck (1993) proposed formulating the exact fit null hypothesis in the following manner:

$$H_{01i}: \text{RMSEA} = 0; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$$

$$H_{a1i}: \text{RMSEA} > 0; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$$

The abovementioned hypotheses however are impractical as they propose that the measurement model is able to reproduce the observed covariance matrix perfectly in the population (Schermelleh-Engel et al., 2003) Alternatively, it has been suggested that a more sensible approach is assessing whether the model fits reasonably well in the parameter (Kaplan, 2000, p. 111), by testing the null hypothesis of "close fit" (Browne & Cudeck, 1993, p. 146). Consequently, operational hypothesis 1 was tested through the following statistical hypothesis:

$$H_{02}: \text{RMSEA} \leq .05; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$$

$H_{a2}: RMSEA \geq .05; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$

If the proposed measurement model approximated close or at least reasonable fit for the three groups, the following null hypotheses would be tested under operational hypotheses 2-8, evaluating the slope, intercepts and error variance of the regression of the items on the PsyCap latent variables, as obtained when fitting a series of multi-group measurement models.

Operational hypothesis 2 was tested by assessing whether the multi-group configural invariance model demonstrates close fit.

$H_{03}: RMSEA \leq .05$

$H_{a3}: RMSEA \geq .05$

If H_{03} was not rejected (i.e. a finding of configural invariance), operational hypothesis 3 could be tested i.e. whether the multi-group weak MI model demonstrates close fit.

$H_{04}: RMSEA \leq .05$

$H_{a4}: RMSEA \geq .05$

If H_{04} was not rejected (i.e. a finding of weak invariance), operational hypothesis 4 could be tested, i.e. whether the multi-group strong MI model demonstrates close fit.

$H_{05}: RMSEA \leq .05$

$H_{a5}: RMSEA \geq .05$

If H_{05} was rejected (i.e. a finding of strong invariance), operational hypothesis 5 could be tested, i.e. whether the multi-group strict MI model demonstrates close fit.

$H_{06}: RMSEA \leq .05$

$H_{a6}: RMSEA \geq .05$

If $H_{0j}; j= 4, 5, 6$ (or $H_{0ij}; i < 24; j=4,5,6$) was not rejected, i.e. a finding of weak, strong and strict invariance, ME was evaluated via operational hypotheses 6-8, through testing the practical significance of the difference between the configural invariance model and the other multi-group MI models.

3.4.1 Tests of Practical Significance

In the MI/ME literature, the method commonly employed in assessing the difference in fit between two nested models is the χ^2 difference test, where a substantial decline in fit is illustrated by significant changes in the χ^2 difference between the respective multi-group measurement model and the baseline model (French & Finch, 2006; Vandenberg & Lance, 2000). Cheung and Rensvold (2002) argue, however, that a double standard is created when researchers employ a number of GOF indices to evaluate model fit, and then consider only the χ^2 difference test to evaluate the difference between two nested models. Furthermore, use of the χ^2 test in isolation may be misleading due to the test's sensitivity to sample size, influencing the accuracy of the χ^2 test in large samples (French & Finch, 2006).

Consequently, Cheung and Rensvold (2002) suggested the use of alternative GOF indices which can be used to assess whether the difference between two nested models is practically significant. These GOF indices are considered superior to the χ^2 test, as they are not affected by sample size. In a study, Cheung and Rensvold (2002) set out to evaluate 20 different GOF indices, as they explained that while various studies had assessed the performance of GOF indices in the context of single-group data, at that stage none had evaluated "how GOFs change when between-group constraints are added to a measurement model" (Cheung & Rensvold, 2002, p. 233). The authors recommended the use of three indices which appeared to be fairly robust against small errors of approximation in the context of cross-group measurement models, namely the McDonald's Non-centrality index (Mc), the Gamma Hat index and the Comparative Fit Index (CFI) (Cheung & Rensvold, 2002). The critical values proposed to detect whether the null hypotheses should be rejected are stipulated in Table 3.2 below.

Table 3.2

Criteria to evaluate the practical significance of the difference in fit between nested models

Practical fit indices	Critical values
McDonald's Non-centrality index	A change of $\leq -.02$
Gamma hat index	A change of $\leq -.001$
Comparative fit index (CFI)	A change $\leq -.01$

Other authors (e.g., Meade et al., 2008), however, argued that a limitation of the Cheung and Rensvold (2002) study was that it did not evaluate the performance of alternative fit indices in cases where measures were not perfectly invariant. Therefore, although the indices recommended by Cheung and Rensvold (2002) were promising, the power of the indices to identify a lack of invariance (LOI) was unclear (Meade et al., 2008). Hence, Meade et al. (2008) set out to assess how various alternative fit indices in MI tests performed, using data with varying levels of LOI (from minor to more serious). Their

study investigated numerous GOF indices, including the CFI, Gamma-Hat, Noncentrality Parameter (NCP), Incremental Fit index (IFI), Relative Fit index (RNI), critical N and RMSEA.

Firstly, their findings indicated that when detecting a LOI at the configural invariance level, all the GOF indices were able to identify a sizeable LOI, with the McDonald's Non-centrality index performing the best. Secondly, the CFI, Mc, gamma-hat, RMSEA, RNI and IFI were demonstrated to be the most sensitive to a LOI, while being the most robust measures in terms of changes in sample size, the number of items, total factors, interactions amongst variables, as well as sampling error. These indices were also shown to outperform the χ^2 difference test under these conditions (Meade et al., 2008). Thirdly, they advised that reporting on numerous GOF indices would be unnecessary due to redundancy of the information gained (such as the CFI, IFI, gamma-hat and RNI). Lastly, in terms of the cut-off values, they suggested a value of 0.002 as the critical value for the CFI to detect metric and scalar invariance⁷, while different values were proposed as the cut-off for the Mc index for different study conditions, based on the number of factors and number of items in a measure. (Meade et al., 2008). Of all the indices examined, they advised against the use of the RMSEA to assess differences in fit. Overall, the findings of Meade et al., (2008) agreed with those of Cheung and Rensvold (2002), demonstrating that the CFI and Mc indices are the most promising for assessing measurement equivalence.

Consequently, while the present study will report on the χ^2 difference test for the nested models, the decision to accept or reject the null hypotheses under operational hypotheses 6-8 was based on the Cheung and Rensvold (2002) criteria. To assess the criteria, the CFI was read from the LISREL fit statistics output, and the Gamma Hat and Mc were calculated for the multi-group configural invariance model (H_a) and the respective multi-group invariance model (H_i). The difference in model fit was considered practically insignificant (i.e. the respective level of equivalence would be demonstrated) if the change between the partially constrained H_i model and the unconstrained H_a model met the criteria stipulated in Table 3.2⁸.

⁷ Although the authors (Meade et al., 2008) refer to the metric and scalar invariance models, in the context of the present study these would be considered the metric and scalar equivalence models. The terms invariance and equivalence are often used interchangeably in literature, however it must be noted that in the present study, the term equivalence is used when referring to the difference between two nested models.

⁸ It is acknowledged that the null hypotheses formulated for $H_{06} - H_{08}$ do not represent true statistical hypotheses that speculate a value for the population. This is because the decision rule created by Cheung and Rensvold (2002) is not based on a calculation of the probability of observing the sample findings conditional on the parameter, under the null hypothesis. Rather, the criteria are met when the inference can be made that the two multi-group measurement models do not differ significantly based on the three practical fit indices (Cheung & Rensvold, 2002).

Metric equivalence was tested via operational hypothesis 6, i.e. the practical significance of the difference in fit between the MI1 model and the multi-group weak MI model.

H₀₇: The multi-group weak MI model and multi-group MI1 model fits equally well in the parameter.

H_{a7}: The multi-group weak MI model fits poorer than the multi-group MI1 model in the parameter.

Operational hypothesis 7 assessed scalar equivalence, i.e. the practical significance of the difference in fit between the MI1 model and the multi-group strong MI model.

H₀₈: The multi-group strong MI model and multi-group MI1 model fits equally well in the parameter.

H_{a8}: The multi-group strong MI model fits poorer than the multi-group MI1 model in the parameter.

Lastly, conditional probability equivalence was tested via operational hypothesis 8, i.e. the practical significance of the difference in fit between the MI1 model and the multi-group strict MI model.

H₀₉: The multi-group strict MI model and multi-group MI1 model fits equally well in the parameter.

H_{a9}: The multi-group strict MI model fits poorer than the multi-group MI1 model in the parameter.

Based on the findings of the preceding statistical hypotheses, the study will establish the extent to which the PCQ-24 provides a valid, reliable and unbiased measure of the PsyCap construct across Black, White and Coloured groups in South Africa.

3.5 Sampling

The data for this study was obtained from previous research studies conducted within the Department of Industrial Psychology, at Stellenbosch University, where Psychological Capital was included as a variable in these studies. The archival data over these studies was pooled into one datasheet, containing only the matched demographic characteristics contained in the different archival datasets, as well as responses on the 24 items of the PCQ-24. The archival data includes a sample of 418 Black middle managers (PhD study of Shayne Roux; Roux, 2014), 199 responses from participants employed in a wide variety of industries collected via social media platforms (master's study of Maritsa Boers, mostly White respondents; Boers, 2014), and 209 responses collected in a construction company from all levels in the company (mostly White and Coloured respondents; master's study of Marthine Herbert; Herbert, 2011). The data collection procedures and sample demographics of the three respective studies will now be discussed in more detail⁹.

⁹ Reference will be made to the ethics about using archival data under Ethical Considerations, paragraph 3.8.

3.5.1 Archival Dataset Number One

3.5.1.1 Data collection procedure

The PsyCap data in Roux's (2014) study was collected via an online questionnaire. The study was conducted within three participating organisations, two of which were in the Fast-Moving Consumer Goods (FMCG) Industry, and the other in the Service Industry (Roux, 2014). In these organisations, the survey was announced to the employees in advance by introducing the goals of the study, whilst stressing that participation was voluntary. Emphasis was also placed on explaining that responses were confidential and personal information would remain anonymous. Furthermore, the participants were informed regarding their rights relating to participating in the study, that no potential risks were expected as a result of participating and were given the opportunity to confirm their voluntary consent to participate (Roux, 2014). The participants were sent links to complete the survey via email, which contained questions measuring the 10 constructs used in Roux's study. Lastly, to increase the response rate, an additional message was sent to participants as a reminder to complete the questionnaire (Roux, 2014).

3.5.1.2 Sample characteristics

Roux's (2014) sample consisted of 41% females and 59% males. Moreover, the ethnic distribution of participants was 45% African, 38% Coloured and 17% Indian. In terms of age distribution, most of the sample indicated 31 – 40 years as their age group (44%). This was followed by the 41 – 50 age group (25%), and the 18 – 30 age group (20%). Only 11% indicated 51 – 60 as their age group, and 1% indicated 61 – 65 years (Roux, 2014). Furthermore, the majority of participants indicated English as their first language (39%) followed by Afrikaans participants (18%) and Zulu (12%) (Roux, 2014). In terms of the distribution of respondents' qualifications, most respondents had a Diploma (33%), followed by a Bachelor's degree (22%), Grade 12 or equivalent (20%), Post High School Certificate (10%), Honours degree (9%) and a Master's degree (6%). Only 1% of respondents had obtained a Doctoral degree. Regarding occupational level, most respondents were in non-supervisory/managerial positions (36%). These were followed by middle managers (27%), junior management (17%), supervisors (13%) and senior management (13%) (Roux, 2014).

3.5.2 Archival Dataset Number Two

3.5.2.1 Data collection procedure

The data for Boers's (2014) study was collected via a social media platform, *Facebook*. A convenience sample was thus used, as the researcher posted an invitation to participate in the study on the platform from her *Facebook* account. The link was sent to the *Facebook* friends of the principle investigator as well as the supervisor, due to the former approach yielding insufficient numbers

(Boers, 2014). The link to participate in the survey was accompanied by an informed consent agreement which explained the participants' rights relating to the research, confidentiality and voluntary participation (Boers, 2014). As the study aimed to obtain responses from a wide range of industries, certain inclusion criteria were communicated to prospective participants. These criteria included that the participants needed to be: "a permanent resident of South Africa; at least 18 years or older; employed full-time in the formal job market; employed for at least six months in their present job; and willing to share their information for research purposes" (Boers, 2014: p. 94).

3.5.2.2 Sample characteristics

Due to the data collection method used, i.e. survey links sent to 'Facebook friends' on the social media platform, the resulting sample was relatively homogenous. The sample consisted of consisting of 64% females and 36% males. Furthermore, most of the sample were White (89%) respondents, followed by 7% Coloured respondents and 4% African, Indian and Asian groups. The first language distribution of the sample was reported as 84% Afrikaans speaking, 15% English speaking, and 1% Xhosa speaking participants (Boers, 2014). In terms of age, the majority of respondents indicated 20 – 29 as their age group (74%). A further 16% of respondents were from the 30 – 39 age group, followed by the 50 – 59 age group (6%), the 40 – 49 age group (3%) and 60+ age group (1%). In addition, 69% of the sample had obtained tertiary qualifications, 45% of which were postgraduate qualifications (Boers, 2014). The remaining 31% of the sample included respondents who held qualifications ranging from Grade 10 to Diploma level (Boers, 2014). The sample was therefore evidently skewed in comparison to the average education level of general South Africans. Regarding industries, Boers (2014) reported that her sample was spread relatively evenly across 17 industries. These included: "Accounting, Agriculture, Construction, Customer services, Design, Education, Engineering, Finance, Health, Human resources, Information technology, Legal, Logistics, Marketing, Media, Sales, and Science", (Boers, 2014; p. 125 – 126) among others.

3.5.3 Archival Dataset Number Three

3.5.3.1 Data collection procedure

The study by Herbert (2011) was conducted within a medium-sized construction company which operates in the property development and construction industry in South Africa. Approval to conduct the study was obtained from the organisational development director of the business, enabling the data collection procedure to commence. An e-mail was sent to the employees which included the informed consent form, a cover letter explaining the study, a letter from the organisation encouraging the employees to participate, as well as the research questionnaire (Herbert, 2011). The researcher also distributed hard copy questionnaires to sites where the employees did not have access to e-mail.

Participants could send their responses back via e-mail¹⁰, fax or post, and to ensure their anonymity was protected, the researcher printed out the received questionnaires and removed participants' personal information. In this way, the participants' confidentiality was maintained (Herbert, 2011).

3.5.3.2 Sample characteristics

The sample mostly consisted of male (69%) respondents, with 31% female respondents respectively. Furthermore, the respondents were mostly White (46%), followed by the Coloured (38%) respondents. The remaining participants were in the Black, Indian and Asian groups (16%) (Herbert, 2011). In terms of their first language, most participants indicated Afrikaans (57%), followed by English (32%) and Xhosa (7%) (Herbert, 2011). The age distribution of respondents was mostly in the 18 – 30 age group (42%). The 31 – 40 age group followed at 27%, as well as the 41 – 50 age group (20%), the 51 – 60 age group (10%), and lastly the 60+ age group (1%). The participants in the study came from various departments and varying job levels in the company. The majority (i.e., 34.9%) were from the Construction department, followed by the Finance department (15.3%), and then Building (10%). (Herbert, 2011). The job levels in the company were relatively well represented, with Non-managerial employees at 30.1%; Junior management at 24.4%; Middle management at 25.4%; Senior management at 10.5%, and Executive management at 4.8%. In terms of education, the majority of participants had obtained a National Diploma (32.5%), followed by Grade 12 or equivalent (24.9%). A relatively equal number of respondents had obtained a Post-school certificate and Bachelor's degree, with 13.9% and 12.4% respectively, while a smaller portion of respondents held postgraduate qualifications (Herbert, 2011).

3.5.4 Critical Reflection on using Archival Data

As the total sample being used for the present study is archival or secondary data, the advantages and disadvantages of this method should be reflected on. One of the disadvantages of utilising secondary data in research is that originally, the data was collected for different research purposes. As a result, the possibility exists that the data may not be ideal to answer the research problem in question. The advantages of secondary data however include that the researcher may gain access to a wider sample base which they may not have had if they collected the data themselves. Furthermore, some authors suggest that utilising secondary data results in research being conducted at a greater speed and at lower costs (Hox & Boeije, 2005). However, even though the primary data collection step is removed from the research process, some (e.g. Jones, 2010) have argued that using archival data still requires considerable time to source data which is suitable for the research purpose, as well as to prepare the

¹⁰ Participants were informed that by returning the questionnaire through this mechanism, their identity would be revealed to the researcher. The informed consent form explicitly stated this and participants were required to indicate their consent to utilize this feedback mechanism.

data for the subsequent analyses to be performed. Therefore, it may not accurate to consider one method more time efficient than the other. A large disadvantage of the current dataset is that different administration methods were utilised, as not all the data was gathered online. This could potentially have introduced method bias into the process (Van de Vijver & Poortinga, 1997; Van de Vijver & Rothmann, 2004). According to Van de Vijver & Poortinga (1997, p. 30), “method bias occurs when a cultural factor that is not relevant to the construct studied affects most or all items of a test in a differential way across the cultures studied”. Three forms of method bias are identified, namely instrument bias, sample bias and administration bias (Van de Vijver & Poortinga, 1997). Traditionally, the administration process of a measure is standardised and controlled, ensuring that the instructions, time keeping, testing conditions and scoring procedure do not vary for individuals across testing conditions. As the instrument was administered differently to the three initial samples which the data was gathered from, administration bias could potentially be present (Byrne & Watkins, 2003; Van de Vijver & Rothmann, 2004). The implications and limitations of this will be reflected on in the limitations section of this study. Furthermore, it is noted that in the present study, the participants’ responses on the PCQ-24 were collected for different research purposes. Despite this, the combined data was suitable for the analysis as relatively equal amounts of respondents’ data was obtained across the Black, White and Coloured groups. Therefore, the researcher could draw valid inferences regarding the performance of the measure across the three groups. A consideration which is required however, is whether the total sample is representative of the South African general population, which will be discussed in the following section.

3.5.5 Combined Dataset

The sampling method utilised in all three studies from which the combined data was sourced, was a convenience sampling technique (Boers, 2014; Herbert, 2011; Roux, 2014). Therefore, the dataset can be regarded as a non-probability sample. A limitation of this sampling technique is that the sample does not provide a precise and accurate representation of the sampling population. Consequently, when generalising the findings of the research to the population, this should be done with caution (Babbie, 2013). The sample demographics of the combined sample including the ethnic group, gender, age and home language distributions are shown in Tables 3.3 – 3.5 below.

Table 3.3
Ethnic group by gender frequency distributions

		Gender			Total
		Male	Female	Missing	
Ethnic group	Black	136	82	-	218
	White	133	137	2	272
	Coloured	133	113	-	246
Total		402	332	-	736

The total combined sample consisted of 736 respondents, comprised of Black, White and Coloured racial groups. In the combined sample, 402 (55%) of respondents were male and 332 (45%) of respondents were female. These demographics are relatively in line with the demographics of the SA population, with 51% females and 49% males reported in the 2016 Census survey (Statistics South Africa, 2016). Regarding ethnicity, from the table it is evident that a relatively even distribution of respondents was obtained across the three groups, with 218 (30%) Black, 272 (37%) White and 246 (33%) Coloured respondents respectively. The age distribution of the sample is shown in Table 3.4 below. The majority of the sample were aged between 18–30 years' old (39%), followed by the 31–40 age group (36%).

Table 3.4
Frequency distribution of respondents by Age

Age	Number of respondents	Number of respondents as % of total
18-30	289	39.3%
31-40	267	36.3%
41-50	109	14.8%
51-60	63	8.5%
60+	8	1.1%
Total	736	100%

Table 3.5 below illustrates the language distribution amongst the sample. It is evident that the majority of respondents indicated Afrikaans as their first language (47.42%), followed by English (24.59%). A total of 27.98% of the sample indicated an African language as their first language, with Zulu at 6.93%, Sotho at 7.47% and Xhosa at 6.25%. The remaining 7.33% consisted of Tsonga, Swazi, Tswana, Venda, Ndebele and other languages. Evidently, the first language distribution of the sample is not in line with the language distribution of the South African general population. That is, Zulu is the first language of the majority of South Africans (23%), followed by IsiXhosa (16%), Afrikaans (14%) and English (10%).

Table 3.5
Frequency distribution of respondents by Home language

Home Language	Number of respondents	Number of respondents as % of total
Afrikaans	349	47.42%
English	181	24.59%
Xhosa	46	6.25%
Zulu	51	6.93%
Sotho	55	7.47%
Tsonga	7	0.95%
Swazi	7	0.95%
Tswana	26	3.53%
Venda	5	0.68%
Ndebele	2	0.27%
Other	7	0.95%
Total	736	100%

3.6 Statistical Analyses

The following section outlines the procedures used in preparing the data for analyses, as well as conducting the actual analyses. The preparatory procedures will be discussed first, including specifying and identifying the measurement model and how missing data was handled. The statistical analyses section will further discuss variable type, model fit and the evaluation of MI and ME.

3.6.1 Preparatory Procedures

3.6.1.1 Model specification

The measurement model in the present study was derived from the design of the PCQ-24 and the internal structure of PsyCap implied by its constitutive definition. This measurement model illustrates the hypothesised relationships between the latent PsyCap factors and the respective items of the instrument. Inferences regarding the Black, White and Coloured groups' level of PsyCap would only be justified if the measurement models fit the data to an acceptable degree, the factor loadings are significant and large, and the error variances are significant and small. The null hypothesis $H_{01i}; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$ and $H_{02i}; i=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$ was tested by fitting the single-group measurement model as displayed in Equation 1 to the three respective groups:

$$\mathbf{X} = \boldsymbol{\zeta} + \boldsymbol{\lambda}^x \boldsymbol{\xi} + \boldsymbol{\delta} \text{-----} [1]$$

Where:

- \mathbf{X} is a 24x1 column vector of observed item scores
- $\boldsymbol{\zeta}$ is a column vector of the intercept terms
- $\boldsymbol{\lambda}^x$ is a 24x1 matrix of factor loadings
- $\boldsymbol{\xi}$ is a 1x4 column vector of latent PsyCap dimensions

- δ is a 24x1 column vector of measurement error components consisting the combined effect of random measurement error and systematic non-relevant influence (Jöreskog & Sörbom, 1996)

Two additional matrices are required in single-group measurement equations for invariance studies, namely the symmetrical variance-covariance matrix and the diagonal variance-covariance matrix. The symmetrical matrix explains the correlations or covariance between the latent variables (ξ), while the diagonal matrix implies that the error terms are uncorrelated. When the off-diagonal elements are freed, the correlations can reveal additional common factors which may not have been depicted in the original PCQ-24 model. As the present study was confirmatory in nature, freeing of off-diagonal elements could not be substantively justified.

The null hypotheses H_{03} and H_{0j} ; $j=4,5,6,7,8,9$ were tested by fitting the multi-group measurement model as displayed in Equation 2 to the combined data of the three Black, White and Coloured groups:

$$X^g = \zeta^g + \lambda^{X^g} \xi^g + \delta^g \text{-----}[2]$$

Where:

- X^g is a 24x1 column vector of observed item scores for group g ; $g=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$
- ζ^g is a column vector of the intercept terms; $g=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$
- λ^{X^g} is a 24x1 matrix of factor loadings; $g=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$
- ξ^g is a 1x4 column vector of latent PsyCap dimensions; $g=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$
- δ^g is a 24x1 column vector of measurement error components consisting the combined effect of random measurement error and systematic non-relevant influence; $g=1_{\text{White}}, 2_{\text{Black}}, 3_{\text{Coloured}}$ (Jöreskog & Sörbom, 1996)

Here, the symmetrical variance-covariance matrices also describe the covariance or correlations between the latent PsyCap dimensions, while the diagonal matrix describes the error variance terms related with the indicator variables. As this matrix is diagonal in nature, it assumes that the error variance terms are uncorrelated.

3.6.1.2 Model Identification

Tabachnick and Fidell (2013, p. 714) explain that in SEM, a model is specified, parameters for the model are estimated using sample data, and the parameters are used to produce the estimated population covariance matrix; but only models that are identified can be estimated. The problem of identification refers to whether sufficient information is available in order for a unique solution of estimated model parameters to be obtained (Diamantopoulos & Sigauw, 2000; Tabachnick & Fidell,

2013). A model needs to be identified in order for unique values for model coefficients to be determined. Two recommendations are provided in this regard. Firstly, Diamantopoulos and Sigauw (2000) specify that a definite scale should be assigned, and secondly, the model parameters to be estimated should not exceed the number of unique covariance items in the observed covariance matrix (Diamantopoulos & Sigauw, 2000). Should the parameters exceed the unique covariance items, LISREL will not be able to estimate the measurement model parameters (Tabachnick & Fidell, 2013). Appropriate model identification therefore enables researchers to obtain a unique solution for the measurement model parameters to be estimated. Whether a model meets the requirements for identification can be judged using Equation 3 below:

$$t \leq s / 2 \quad [3]$$

Where:

- t = the number of parameters to be estimated
- s = the number of variances and covariances amongst the observable variables, calculated as $(p)(p+1)$
- p = the number of observed variables

If $t > s/2$ it needs to be concluded that the model is under-identified, which means that the model and the combined data failed to obtain unique estimates for the measurement model parameters (Diamantopoulos & Sigauw, 2000; Tabachnick & Fidell, 2013). Alternatively, if $t = s/2$ then the model is considered just-identified. In this case, it is possible to obtain a single unique solution for the parameter estimates, however no variance-covariance information would remain to test the resultant model solution due to the model having zero degrees of freedom. Finally, if $t < s/2$, it would mean that the model is over-identified indicating that the equations available outweigh the model parameters to be estimated, resulting in positive degrees of freedom. Therefore, the derived model solution can be tested as variance-covariance information remains (Diamantopoulos & Sigauw, 2000; Tabachnick & Fidell, 2013). The measurement models in the present study were over-identified since the degrees of freedom ranged from 246 – 570. The degrees of freedom for the single-group and all the fitted multi-group MI models are depicted in Table 3.6.

Table 3.6*Degrees of Freedom for the Single-Group and series of Multi-Group MI Models*

Model / Hypothesis	# Lambda's	# Tau's	# Theta-delta's	# Phi's	Total # of parameters to be estimated	# Indicator variables	# Groups	# Unique information pieces	DF
Single group measurement model ¹¹	20	24	24	10	78	24	1	324	246
Configural invariance [H03]	40	48	48	20	156	24	2	648	492
Weak invariance [H04]	20	48	48	20	136	24	2	648	512
Strong invariance [H05]	20	24	48	20	112	24	2	648	536
Strict invariance [H06]	20	24	24	20	88	24	2	648	560
Complete invariance [H07]	20	24	24	10	78	24	2	648	570

¹¹ Three separate single group measurement models were tested, i.e. one for each of the three ethnic groups. As the single-group measurement model that was tested was the same for the Black, White and Coloured groups, the degrees of freedom were the same.

3.6.1.3 Power Analysis

When discussing the sampling population, an important consideration is how large the sample should be. Moreover, the researcher must consider the statistical power related to the test of exact and close fit and the number of freed model parameters in the model (Babbie & Mouton, 2001). A model is considered to have sufficient statistical power when it is likely to reject a null hypothesis, in cases where the null hypothesis is in fact false. The statistical power of a sample is of critical importance as it indicates whether outcomes of null hypothesis testing can be interpreted unambiguously (Diamantopoulos & Siguaaw, 2000). Therefore, if statistical power was low, the outcome of a null hypothesis not being rejected could be interpreted as a result of the actual situation, while in fact it is due to the statistical power.

According to Bentler and Chou (1987, p. 91):

The ratio of sample size to the number of free parameters may be able to go as low as 5:1 under normal and elliptical theory, especially when there are many indicators of latent variables and the associated factor loadings are large. Although there is even less experience on which to base a recommendation, a ratio of at least 10:1 may be more appropriate for arbitrary distributions. These ratios need to be larger to obtain trustworthy z-tests on the significance of parameters, and still larger to yield correct model evaluation chi-square probabilities.

The ideal statistical power value is ≥ 0.80 , or otherwise stated, the researcher has 80% probability of obtaining a significant finding when it is in fact true (Tabachnick & Fidell, 2013). Power estimates for the test of close fit for the Black, White and Coloured single-group measurement models were obtained using the Preacher and Coffman (2006) software (available at <http://www.quantpsy.org/rmse/rmse.htm>). Estimates were only calculated for the single-group measurement models due to the fact that the software does not permit power evaluations for multi-group samples. Table 3.7 below presents the sample sizes and degrees of freedom for the respective single-group measurement models tested:

Table 3.7
Preacher and Coffman Statistical Power Calculations

Hypothesis/Model	Alpha	RMSEA H0	RMSEA Ha	N	df	Power
Black Group [H ₀₁]	.05	.05	.08	218	246	0.99
Black Group [H ₀₂]	.05	.05	.06	218	246	0.54
White Group [H ₀₁]	.05	.05	.08	272	246	0.99
White Group [H ₀₂]	.05	.05	.06	272	246	0.65
Coloured Group [H ₀₁]	.05	.05	.08	246	246	0.99
Coloured Group [H ₀₂]	.05	.05	.06	246	246	0.60

It is evident that for each single-group measurement model at RMSEA value of 0.08 for H01, a 99% power value was returned by the Preacher and Coffman (2006) syntax. Considering the statistical power level recommendation by Tabachnick and Fidell (2013), this is a very positive result. At an RMSEA value of 0.60 for H02, the Black, White and Coloured measurement models produced statistical power values of 54%, 65% and 60% respectively. Evidently, the measurement models display higher statistical power at H01 level, i.e. an RMSEA of 0.08, where a finding of good fit for the three single-group measurement models could confidently be taken as unambiguous evidence in favour of the models' fit.

3.6.1.4 Treatment of Missing Values

Missing values can be due to various reasons in research, and they must be dealt with prior to analysing the data. For example, in experimental designs which track participants over a period of time and involve a number of tests and re-tests, missing values can arise as a result of respondents not being able or willing to continue in the process. Secondly, in studies involving surveys, it is common for participants to skip some of the questions or choose the 'not certain' response option (Babbie, 2013).

Researchers can employ different methods to deal with missing data, depending on the severity of the issue (Babbie, 2013). Firstly, list wise deletion can be applied, which results in a dataset with only complete cases (Mels, 2003). It should be noted, however, that this can have the effect of reducing sample size depending on the length of the questionnaire and the extent of the problem (Babbie, 2013). Alternatively, a multiple imputation procedure can be used which involves estimation of missing values which are derived from the original sample, rather than deleting cases with missing values (Tabachnick & Fidell, 2013). This process is carried out using LISREL 8.8. For multiple imputation to be used successfully, the data has to be continuous and have a multivariate normal distribution. Full Information Maximum Likelihood (FIML) also assumes a multivariate normal distribution and is more complex than the multiple imputation procedure as the method involves imputing missing values based on the respondents' previous responses. (Hair et al., 2010; Jöreskog & Sörbom, 2003). If the assumption of multivariate normality is not met, the problem of missing values can be solved using imputation by matching. This procedure involves missing values being substituted with real values, by deriving them from other cases with comparable response styles (Jöreskog & Sörbom, 2003).

When missing values are present in the dataset, selecting an approach to deal with the issue requires careful consideration. Hair et al., (2010) advise that if the missing values are missing at random, comprise only 10% or less of the total dataset and fairly high factor loadings are observed (≥ 0.07), any of the above-mentioned approaches will be suitable. If a high number of missing values is prevalent

however, then the strengths and weaknesses of the various approaches need to be considered, in the context of the study in question.

For example, in terms of listwise deletion, this solution is easy to implement using any SEM program and is effective when the sample size is known. However, this method can increase the possibility of non-convergence unless the sample size is large (> 250) and factor loadings are large (> 0.6). The possibility also exists that the procedure can result in factor loading bias, as well as bias in the estimates of relationships among factors (Hair et al., 2010). Multiple imputation procedures appear to have fewer problems with non-convergence and also show the least bias effects in cases of random missing data. This method becomes more beneficial in situations where the cases of missing data increase, and the factor loadings and sample size decrease (Hair et al., 2010). Lastly, FIML is advantageous as the missing data is corrected during the estimation process, however with his method, the researcher has less control regarding how the missing data is corrected. Furthermore, the impact of the missing data on the estimates is also unknown. Despite the potential drawbacks of this method, it still produces relatively less bias than the other missing data procedures (Hair et al., 2010).

In the present study, archival data was used. Upon inspection of the combined data it was evident that the dataset contained no missing values. In the studies of Roux (2014) and Boers (2014), online surveys were utilised in the data collection procedures, which resulted in no missing values in these datasets. To this end, Boers (2014, p. 94 – 95) stated that, “Another benefit of the data collection method was that due to the online administration of the questionnaire, no missing data was evident in the final dataset.” Furthermore, Roux (2014) reflected on the decision to use an electronic survey, with reference to the possibility of missing data by stating that it, “minimises or even eliminates the problem of missing values in a dataset.” (Roux, 2014. p. 125). Furthermore, regarding data collection procedures, he mentioned: “Respondents had to complete the entire questionnaire.” (Roux, 2014. p. 127) implying that no missing data was present in the final dataset as all questions had to be answered¹². The hard copy data collection procedure utilised in the Herbert (2011) study resulted in missing values. The missing values were addressed with the Imputation by Matching procedure, conducted with LISREL 8.8 (Herbert, 2011). While Herbert’s (2011) original sample consisted of 209 participants, the final sample was marginally less (n= 202). The author commented that this was due to the imputation by matching procedure, “when appropriate imputations of missing cases are not possible, the sample size is reduced” (Herbert, 2011, p. 143). The final sample was thus reduced by 7

¹² It is acknowledged that under normal ethical guidelines, participants may refuse to answer survey questions. The fact that this was not the case in the studies of Boers (2014) and Roux (2014) should be interpreted within the boundaries of the ethical clearances that were obtained for these two studies.

cases. Herbert's (2011) imputed dataset was merged with the other two datasets to form the combined dataset for this study.

3.6.1.5 Item Analysis

Prior to fitting the measurement models, a series of item analyses were conducted to assess the assumption that the items of the four PCQ-24 subscales have common sources of variance. According to classical measurement theory, an observed score on a measure is the product of an individual's true score, and measurement error. As the random error in the measure is reduced, the more the observed score reflects an individual's true score (Kline, 2005). In terms of the design of the PCQ-24, the items were developed to reflect the four PsyCap constructs. The 24 items are expected to provide relatively uncontaminated measures of the sub-dimensions of Self-efficacy, Hope, Optimism and Resilience respectively (each represented by 6 items), where each set of items reflects the corresponding underlying variable. Item analysis assists in this process, as the researcher can assess the performance of items compared to other items in the measure and identify items which do not provide clear representations of the subscales as intended (Kline, 2005).

This procedure thus contributes to improved test reliability. Items which interfere with the internal consistency of a particular subscale are flagged as possible poor items and if the item is significantly poor, i.e. the reliability of the measure would improve significantly if deleted, the researcher will consider removing the item from the relevant subscale (Kline, 2005). However, the deletion of poor items was not within the scope of the present study, as the aim was to assess the performance of the full original PCQ-24 measure across Black, White and Coloured groups. The results from the item analyses, would however serve as valuable information to consider when the MI and ME analyses are conducted, in terms of identifying potential sources of a lack of invariance or equivalence in the various levels of analyses.

Item Analysis: Black group

The results of the item analyses for the Black group (Table 3.8) revealed satisfactory reliability values for the Self-efficacy and Hope subscales (i.e., above 0.70; Nunnally, 1987). Much less satisfactory results emerged for the Resilience and Optimism subscales however, with alpha values of 0.448 and 0.376 respectively. These results generally followed the same trend observed in the PsyCap literature, as many authors have commented on the fact that the internal consistency reliability for the Resilience and Optimism subscales appear over various studies to be consistently lower than values obtained for Hope and Self-efficacy (Avey et al., 2010a; Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013; Luthans et al., 2007a). However, in these studies the reliability statistics obtained for Optimism and Resilience ranged between 0.63 to 0.69, missing the 0.70 cut off value to a small degree. The results

in the present study obtained for Resilience (0.448) and Optimism (0.376) are far below previous findings in literature, hence the reliability statistics for the Resilience and Optimism subscales are particularly poor for the Black group.

Table 3.8

The means, standard deviations and reliability statistics for the PCQ24 subscales – Black group

PCQ- 24 Subscales	No of items	M	SD	α
Self-efficacy	6	30.28	3.958	0.791
Hope	6	30.26	3.787	0.795
Resilience	6	28.77	3.325	0.448
Optimism	6	26.85	3.429	0.376

Further inspection of the results for the Self-efficacy and Hope subscales revealed no items that could be flagged as possible poor items. In particular, the Self-efficacy ($\alpha = 0.791$) inter-item correlations ranged from 0.216 (item 1) to 0.640 (item 3) and squared multiple correlations ranged from 0.150 (item 1) to 0.503 (item 3). Although item 1 obtained a squared multiple correlation which was not in line with the trend observed for the other items, the alpha value would increase only marginally if deleted ($\Delta = 0.01$) and hence this item was not flagged as a poor item. For the Hope subscale ($\alpha = 0.795$), the inter-item correlations ranged from 0.217 (item 9) to 0.638 (item 10), with squared multiple correlations ranging from 0.212 (item 9) to 0.535 (item 11). Item 9 correlated slightly lower than the other items, with a squared multiple correlation of 0.212, while the remaining items ranged from 0.249 to 0.534. Only a slight increase in subscale reliability would be observed if deleted however ($\Delta = 0.004$), which is why this item was not considered to be flagged as a poor item.

Contrary to the results for the Hope and Self-Efficacy sub-scales, the inter-item correlation matrix of the Resilience subscale ($\alpha = 0.448$) revealed inconsistent correlations, ranging from -0.188 to 0.322 respectively. Item 13 (“When I have a setback at work, I have trouble recovering from it, moving on”) obtained the lowest inter-item correlations in the subscale, ranging from -0.188 to 0.035 and correlating negatively with two other items. This item also obtained the lowest squared multiple correlation (0.050) while the other items ranged from 0.151 to 0.225. Furthermore, the Cronbach’s alpha value would increase significantly from 0.448 to 0.626 if deleted, and hence the item was flagged as a possible poor item. It is important to note that item 13 is the only negatively keyed item in this subscale, a practice which some researchers¹³ have consistently argued to possibly be detrimental to

¹³ Findings regarding the impact of negatively keyed items in psychometric assessments will be discussed further in paragraph 3.6.1.4.4.

measurement accuracy, particularly when the respondent is not tested in their first language (Dalal & Carter, 2015; DiStefano & Motl 2006; Jackson Barnette 2000; Suárez-Alvarez et al., 2018).

The results for the Optimism subscale ($\alpha = 0.376$) revealed that, items 20 ("If something can go wrong for me workwise, it will") and 23 ("In this job, things never work out the way I want them to") were also flagged as potentially problematic, with inter-item correlations ranging from -0.025 to 0.372 and -0.061 to 0.372 respectively. Furthermore, item 24 ("I approach this job as if "every cloud has a silver lining") also obtained lower inter-item correlations, which varied from -0.173 to 0.310. Evidently, items 20, 23 and 24 performed poorer than rest of the items in the subscale, which ranged from 0.284 to 0.515; however, the results suggested rather poor internal consistency for the Optimism subscale in the Black group.

Furthermore, these items also obtained lower squared multiple correlations, at 0.191 (item 20), 0.182 (item 23) and 0.164 (item 24), while the remaining items ranged from 0.204 to 0.344. However, if deleted, only marginal changes in the alpha would be evident for the subscale. The alpha would increase from 0.376 to 0.457 if item 20 was deleted, to 0.379 if item 23 was deleted, and to 0.381 if item 24 was deleted. These changes would not result in a great improvement in the subscale reliability or bring it closer to 0.70. It is further noted that items 20 and 23 are the other two negatively keyed items on the PCQ-24 (bringing the total of 3 negatively keyed items in the full PCQ-24 scale), which have consistently been shown to be problematic in literature. Based on the overall results, it is clear that the internal consistency of the Resilience and Optimism subscales for the Black group was exceptionally problematic.

Item Analysis: White group

The results of the item analysis conducted on the White sub-sample, indicated that the Self-efficacy and Hope subscales achieved Cronbach's alpha values of 0.878 and 0.849 respectively, indicating very good reliability (Nunnally, 1987). The inter-item correlations for Self-efficacy ranged from 0.344 (item 1) to 0.715 (item 3) with squared multiple correlations ranging from 0.284 (item 1) to 0.661 (item 4). No items were flagged as possible poor items in this subscale. Similarly, the Hope subscale obtained positive findings with inter-item correlations ranging from 0.279 (item 9) to 0.706 (item 11). Furthermore, the squared multiple correlations ranged from 0.297 (item 9) to 0.615 (item 11) and thus no items were highlighted as potentially poor in the process.

The Resilience subscale obtained a fairly lower reliability result ($\alpha = 0.710$) compared to the Hope and Self-Efficacy subscales. Although this is still considered satisfactory in terms of the 0.70 critical value (Nunnally, 1987), the result is in line with the general trend in PsyCap literature; that the Self-efficacy

and Hope subscales are more internally consistent than the other subscales (Avey et al., 2010a; Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013; Luthans et al., 2007). A large variance was evident in the inter-item correlation matrix of the subscale with correlations ranging from 0.044 (item 13) to 0.556 (item 16). Item number 13 (“When I have a setback at work, I have trouble recovering from it, moving on”) again stood out as a potentially problematic item, as it had the lowest inter-item correlations (0.44 to 0.213), as well as the lowest squared multiple correlation (0.056), which fell outside the general trend observed for the squared multiple correlations on this subscale (ranging from 0.289 for item 14 to 0.442 for item 18). Moreover, the reliability of the subscale would increase considerably from 0.710 to 0.792, if item 13 was removed.

The Optimism subscale ($\alpha = 0.692$) marginally missed the 0.70 cut off value for good internal consistency (Nunnally, 1987). Although the reliability was significantly higher for the White group than the Black group ($\alpha = 0.376$), the results are still in line with previous PsyCap research (i.e. Avey et al., 2010a; Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013) illustrating lower subscale reliability than the Self-efficacy and Hope subscales. Similar to the Black group, items 20 and 23 appeared to obtain inter-item correlations which were not in line with the general pattern observed for the subscale. This is evident as item 20 and 23 ranged from 0.068 to 0.438 and 0.137 to 0.438 (correlating highest with one another), while the remaining items ranged from 0.291 to 0.497. The two negatively keyed items were therefore clearly out of range with the rest of the items in the subscale. Interestingly, items 20 and 23 did not fall far out of the range of squared multiple correlations, obtaining 0.203 and 0.242 respectively, while the remaining values varied between 0.171 (item 24) and 0.388 (item 21). In terms of reliability, the subscale alpha value would increase only slightly from 0.692 to 0.699 if item 20 was deleted and would drop marginally from 0.692 to 0.663 if item 23 was removed. Although these are still potentially problematic items, removing them will not have the desired effect of a substantial improvement in subscale reliability. The effect of the negatively keyed items is potentially less of an issue in the White sample, as it is well known that many White respondents would have indicated Afrikaans or English as their first language. The White respondents may therefore possibly be more bilingual than the respondents from the other ethnic groups, improving their comprehension of the negatively keyed items.

Table 3.9

The means, standard deviations and reliability statistics for the PCQ24 subscales – White group

PCQ 24 Subscales	No of items	M	SD	α
Self-efficacy	6	29.19	4.962	0.878
Hope	6	27.78	4.911	0.849
Resilience	6	26.68	3.955	0.710
Optimism	6	26.60	4.490	0.692

Item Analysis: Coloured group

Inspection of the results of the item analysis for the Coloured group (Table 3.10) revealed that the reliability of the Self-efficacy ($\alpha = 0.876$) and Hope ($\alpha = 0.832$) subscales exceeded 0.70 (Nunnally, 1987); a finding which is in line previous research (i.e. Dawkins et al., 2013; Luthans et al., 2007). The inter-item correlations matrix for the Self Efficacy subscale ranged from 0.333 (item 1) to 0.672 (item 2) with squared multiple correlations ranging from 0.395 (item 5) to 0.631 (item 4). Therefore, no items appeared out of sync with the others. The inter-item correlation matrix for the Hope subscale ranged from 0.269 (item 12) to 0.656 (item 10). The squared multiple correlations for the subscale revealed that item 7 was slightly out the range observed amongst the other items; obtaining 0.273 while the remaining values fluctuated between 0.357 to 0.576. Despite this, the subscale reliability would not increase significantly if any items were removed.

Similar to the results obtained in the Black and White samples, a slightly lower Cronbach's alpha value was obtained for the Resilience ($\alpha = 0.714$) subscale, although this still met the criteria for satisfactory reliability (Nunnally, 1987). It was evident that item 13 ("When I have a setback at work, I have trouble recovering from it, moving on") again fell out of sync with the other items on the subscale, obtaining the lowest inter-item correlations in the subscale ranging from 0.045 to 0.205. Item 13 also obtained the lowest squared multiple correlation of the subscale at 0.061. The remaining items obtained inter-item correlations ranging from 0.278 to 0.559 and squared multiple correlations ranging from 0.258 to 0.491. Furthermore, the Cronbach's alpha would increase rather significantly from 0.714 to 0.787 if deleted, indicating that item 13 was clearly a problematic item in the subscale. A similar trend was observed for the Black and White sample groups with item 13 obtaining the lowest squared multiple correlations of the subscale. Therefore, it is evident that this negatively keyed item impacted the internal consistency reliability of this subscale, regardless of the sample group in question.

Finally, it is evident from the results that the Optimism subscale ($\alpha = 0.663$) slightly missed the 0.70 cut off value (Nunnally, 1987). Therefore, the trends in literature that the Self-Efficacy and Hope subscales consistently obtain higher internal consistency reliability than the Resilience and Optimism subscales was also observed in the present study, across all three sample groups (Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013). Again, items 20 and 23 did not fall in line with the range of inter-item correlations observed for the subscale. Item 20 produced correlations ranging from -0.062 to 0.462 and item 23 from -0.007 to 0.462 (correlating highest with one another). The remaining items in the subscale ranged from 0.297 to 0.631, evidently much higher than the two negatively keyed items. Interestingly, the items did not fall out of sync with the pattern of squared multiple correlations observed, returning values of 0.266 (item 20) and 0.309 (item 23) while the remaining values ranged

from 0.210 to 0.510. The Cronbach's alpha of the subscale would increase marginally from 0.663 to 0.682 if item 20 was deleted and decrease slightly from 0.663 to 0.638 if item 23 was deleted. Removal of these two items would consequently not improve the reliability of the subscale to a large degree.

Table 3.10

The means, standard deviations and reliability statistics for the PCQ24 subscales – Coloured group

PCQ 24 Subscales	No of items	M	SD	α
Self-efficacy	6	29.11	4.726	0.876
Hope	6	28.23	4.456	0.832
Resilience	6	28.62	3.889	0.714
Optimism	6	26.20	4.277	0.663

Item Analysis: Combined sample

Lastly, the results of the item analyses for the combined dataset are provided in Table 3.11 below. Consistent with the previously reported results, the results for the combined dataset revealed that the Self-efficacy and Hope subscales obtained very good reliability results (Nunnally, 1987). The Self-Efficacy inter-item correlations matrix revealed no items that were out of sync with the rest, ranging from 0.332 to 0.667. Moreover, the squared multiple correlations ranged from 0.267 (item 1) to 0.579 (item 2). Secondly, the Hope subscale produced inter-item correlations ranging from 0.280 to 0.648 with a similar pattern of squared multiple correlations (0.301 to 0.585). The results further indicated that no items, if deleted, would result in an increase in subscale reliability.

The Resilience subscale obtained a Cronbach's alpha value of 0.647, narrowly missing Nunnally's (1987) critical value of 0.70 for satisfactory reliability. When analysing the inter-item correlation matrix, it was evident that item number 13 was again flagged as potentially problematic. Its correlations ranged from -0.035 to 0.098 while the remaining items ranged from 0.303 to 0.480. Furthermore, its squared multiple correlation was far out of sync with the other items, producing a value of 0.028, while the other items ranged from 0.213 to 0.351. Consequently, the subscale alpha value would increase significantly from 0.647 to 0.744 if deleted. The Optimism subscale obtained a rather low Cronbach's alpha value of 0.596, again confirming previous research in this regard. Inspection of the inter-item correlation matrix revealed, as expected, that items 20 and 23 were not in alignment with the rest of the items. Items 20 and 23 produced inter-item correlations ranging from -0.088 0.455 and -0.042 - 0.455 respectively (again, correlating highest with each other), while the remaining items obtained larger correlations which fluctuated from 0.294 to 0.540. Despite this, items 20 and 23 still appeared to be aligned with the others in terms of their squared multiple correlations, with values of 0.227 and 0.242 produced while the other items ranged from 0.193 to 0.355. Furthermore, removal of item 20 would result in a rather marginal increase in subscale reliability, from 0.596 to 0.626, while removal

of item 23 would result in a slight decrease in reliability, from 0.596 to 0.579; therefore deleting of items would have no significant impact on subscale reliability.

Table 3.11

The means, standard deviations and reliability statistics for the PCQ24 subscales – Combined sample

PCQ 24 Subscales	No of items	M	SD	α
Self-efficacy	6	29.49	4.628	0.858
Hope	6	28.67	4.567	0.839
Resilience	6	28.69	3.752	0.647
Optimism	6	26.54	4.132	0.596

It is evident that the negatively keyed items present in the PCQ-24, namely items 13, 20 and 23, negatively impacted the reliability of the relevant subscales (Resilience and Optimism). Negatively keyed items have traditionally been utilised in psychological measures with the intention of reducing response styles. Allen (2017, p. 1475), explained that response styles are distinctive ways of responding to questionnaire surveys that are unrelated to the content of the actual survey items. Considering a measure in which respondents must indicate their level of agreement or disagreement with survey items, examples of response styles include; cases where respondents may agree with all items in a measure, even when items have meanings which contradict each other. This form of response style is commonly referred to as acquiescence (Allen, 2017; Suárez-Alvarez et al., 2018; van Sonderen et al., 2013). On the other hand, participants may favour more extreme responses, either strongly agreeing or disagreeing with most scale items. Lastly, participants may select mostly moderate responses to statements. Hence, response styles can also differ in terms of response extremity or moderation (Allen, 2017; Suárez-Alvarez et al., 2018; van Sonderen et al., 2013).

Therefore, although the use of negatively-keyed items has been encouraged (i.e. Weijters et al., 2013), many authors have demonstrated that this practice can have significant negative effects on the psychometric properties of a measure (Dalal & Carter, 2015; DiStefano & Motl 2006; Jackson Barnette, 2000; Suárez-Alvarez et al., 2018), which was evident in the present study. Furthermore, Suárez-Alvarez et al. (2018) discouraged the practice as they argued that the comprehension of negatively keyed items requires greater linguistic skills and hence will be biased toward respondents with lower levels of verbal ability. Considering that a substantial number of respondents in the present study did not indicate English as their first language, the impact of language proficiency needs to be considered when interpreting the research findings.

3.6.2 Evaluation of the PsyCap Measurement Model

The following section will highlight the variable type of the observed variables in this study, elaborate on how model fit will be evaluated, as well as how measurement invariance and equivalence will be evaluated.

3.6.2.1 Variable type

An important consideration in fitting the measurement model is whether the four PsyCap dimensions should be represented by individual items or item parcels. Item parcelling solves a variety of data problems, including unstable parameter estimates, small sample sizes and non-normality (Bandalos et al., 2001; Bandalos, 2002; Bonnacio & Reeve, 2006; Little et al., 2002; Little et al., 2013). Various arguments against the use of item parcelling have been noted in literature, however. For example, when inappropriate techniques are used to create the parcels, it can result in multidimensional constructs being obscured. This may result in the construct appearing unidimensional while the parcelling technique has in fact resulted in its incorrect representation (Little et al, 2002; Little et al, 2013). Furthermore, it is argued that model misspecifications such as cross-loadings or residual correlations may be masked by item parcels, resulting in improper estimation of other model parameters (Little et al., 2002; Little et al, 2013).

Little et al. (2002) argue that the decision regarding the use of individual items versus item parcels should be carefully considered and based on the substantive goal of the study, i.e., whether the researcher aims to understand the nature of a set of constructs or to examine the structure of a set of items. The aim of the current study is to evaluate the MI and ME of the PCQ-24. Moreover, if the measure is found to lack invariance in any level of the analyses, the study aims to examine precisely where these differences lie. As a result, fitting the measurement models with individual items is better fit for this purpose, as the researcher will be able to identify the sources of a lack of invariance more precisely, than if item parcels were used. Furthermore, as the instrument in the present study consists of a relatively small number of items (24) and has only 4 sub-dimensions, evaluation of the measure at the item-level was deemed feasible.

The PCQ-24 uses a six-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree), which produces ordinal data. Despite this, Muthén and Kaplan (1985) indicate that Likert scales which produce ordinal data with five or more scale points can be specified as continuous data when conducting confirmatory factor analysis. The authors conducted a Monte Carlo study to analyse the results generated from four different estimation techniques, namely Generalised Least-Squares, Maximum Likelihood (ML), Categorical variable methodology and Asymptotically Distribution Free,

where non-normal categorical (ordinal) variables were treated as non-normal interval (continuous) variables. Their results indicated that in cases where variables were moderately kurtotic and skewed, the chi-square, standard error and parameter estimates were not impacted significantly under ML estimation. Consequently, they concluded that the normal theory estimators (Generalised Least-Squares and ML) performed well under these circumstances, i.e. with non-normal ordinal (categorical) variables (Muthén & Kaplan, 1985).

These findings were further supported in later studies, where authors classified ordinal variables with five or more scale points as continuous, without impacting the subsequent analyses (Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993). Based on these findings the individual items will be regarded as approximating continuous variables even though strictly speaking, these are ordinal variables.

3.6.2.2 Testing Model Fit

Measurement model fit refers to the success with which the proposed model accounts for the observed covariance matrix. If the proposed model is able to reproduce the observed covariance matrix closely, then the fit of the model is considered satisfactory (Byrne, 1998; Kelloway, 1998). Obtaining satisfactory fit however, cannot be considered evidence that the hypothesised model provides an exact account of the psychological processes underpinning the behaviour, but rather that the model can be considered a valid account of the observed covariance matrix (Dunbar et al., 2011).

For the purposes of examining the single-group measurement model fit, a range of GOF indices provided by LISREL (Diamantopoulos & Siguaaw, 2000) were used, which was discussed in section 3.3.1. Furthermore, the fit of the multi-group measurement models were assessed by testing the close fit null hypotheses H_{0j} ; $j=4,5,6,7$. LISREL 8.8 was used to examine the fit of the single and multi-group measurement models, in order to evaluate the MI and ME of the PCQ-24.

3.6.2.3 Testing Measurement Invariance and Measurement Equivalence

This section will outline the series of tests specified by Dunbar et al. (2011) which was used to determine whether MI and ME were demonstrated for the respective measurement models. The fully constrained multi-group model (configural invariance model) was used as the baseline model. A series of multigroup measurement models with varying degrees of constraints placed on the models, were then compared to the baseline model to assess whether measurement equivalence is evident. Conversely, if equivalence is found to be lacking, the analyses would indicate exactly where the differences lie (Dunbar et al., 2011). An overview of the steps that were conducted in assessing the MI and ME of the PCQ-24 are depicted below.

Table 3.12
Steps in conducting analyses of MI and ME

Levels of MI and ME analyses						
Tests of MI	Step 1: Assess whether the single-group measurement model displays acceptable fit when fitted to each group sample independently.	Step 2: Assess whether the multi-group Configural Invariance model (MI1) demonstrates acceptable fit.	Step 3a: Assess whether the multi-group Weak Invariance model (MI2) demonstrates acceptable fit.	Step 4a: Assess whether the multi-group Strong Invariance model (MI3) displays acceptable fit.	Step 5a: Assess whether the multi-group Strict Invariance model (MI4) displays acceptable fit.	
Tests of ME			Step 3b: Establish whether the MI2 model fits the multi-group data poorer than the MI1 model, i.e. testing Metric Equivalence.	Step 4b: Establish whether MI3 model, fits the multi-group data poorer than the MI1 model; i.e. testing Scalar Equivalence.	Step 5b: Establish whether the MI4 model fits the multi-group data poorer than the MI1 model; i.e. testing Conditional Probability Equivalence.	

The GOF statistics that were considered when judging the fit of the measurement models have been discussed in section 3.3.1, however a summary of these and the respective cut-off values that were used to evaluate MI are provided in Table 3.13 below.

Table 3.13
GOF Statistics to judge model fit in MI analyses

GOF Statistics	Respective cut-off values
χ^2	P value > .05
RMSEA	$\leq .05$ = Close fit; $\leq .08$ = Reasonable fit
P value for RMSEA	P value > 0.05
SRMR	< .08 = Good fit; < .1 = Acceptable fit
NNFI	$\geq .95$ = Good fit; $\geq .97$ = Very good fit
CFI	$\geq .95$ = Good fit; $\geq .97$ = Very good fit

Prior to conducting the multi-group measurement model analyses, Dunbar et al (2011) explain that acceptable fit needs to be obtained on all three samples, i.e. the measurement model should display good fit in all groups independently. This practice was recommended by Meade et al. (2008). In their study of Alternative GOF indices, they found that poor model fit in configural invariance model testing could be attributed to two reasons, namely; "(a) good model fit for one group and poor model fit for the other (which could be indicative of a lack of configural invariance), or (b) equally poor model fit

for both groups, in which case configural invariance may indeed exist, but the specified model does not accurately represent the data structure” (Meade et al., 2008, p. 590). Consequently, Meade et al. (2008) proposed that single-group measurement models should be examined in isolation prior to analysing the fit of the multi-group data, as these may yield valuable information before the MI / ME analyses are conducted. Step 1 of the series of analyses, were therefore conducted in LISREL 8.8 by performing CFA on the single-group measurement models in independent analyses and judging the fit of each individual measurement model (Meade et al., 2008). If the null hypothesis of close fit would be rejected ($H_{02i} \leq .05$; $i = 1_{\text{Black}}, 2_{\text{White}}, 3_{\text{Coloured}}$), it would illustrate that the measurement model failed to display acceptable fit for the Black, White and Coloured groups, when tested independently. In this case, further analyses would be questioned (Dunbar et al., 2011).

Step 2 assesses configural invariance, which focuses on the theoretical structure of the PCQ-24. A finding of configural invariance would indicate that the same underlying construct is being measured in all three groups. Consequently, a lack of configural invariance would indicate the presence of construct bias. According to Meiring et al. (2005, p.2), “Construct bias occurs when the construct measured is not identical across cultures or when behaviours that characterise the construct are not identical across cultures”. Therefore, in this case if construct bias was present it would indicate that the PsyCap constructs measured may not be the same across Black, White and Coloured groups. Configural invariance needs to be demonstrated in order for the subsequent MI and ME tests to be conducted, as comparisons will only be relevant if the same construct exists and is measured similarly across groups (Dunbar et al., 2011; Meade et al., 2008; Vandenberg & Lance, 2000).

Once configural invariance has been shown, the multi-group weak invariance model is fitted to the data, which tests whether the factor loadings of the items are invariant across groups. Stated otherwise, testing the weak invariance null hypothesis of close fit ($H_{04}: RMSEA \leq 0.05$) assesses whether the slope of the regression of the items on the latent variables differ across the samples (Dunbar et al., 2011). Rejecting the weak invariance null hypothesis would indicate that the groups are interpreting the content of some of the items differently (Byrne & Watkins, 2003).

It should be noted that it is possible that some items may differ in the slope of the regression amongst the three sample groups. As a result, if $H_{04}: RMSEA \leq 0.05$ were to be rejected because of particular factor loadings with significant differences, the inspection of partial weak invariance would be warranted. Partial MI and ME involves applying less stringent conditions in the respective analyses to enable cross-group comparisons (Byrne et al., 1989; Vandenberg & Lance, 2000). Inspection of partial MI and ME allows the researcher to identify the source of variance which is causing the misfit between the multi-group measurement model and the data. This is achieved by freeing invariant parameter

estimates until close fit is obtained and the difference in fit between the MI1 (baseline model) and the respective partially constrained model is no longer (statistically or practically) significant (Cheung & Rensvold, 1999). Items suffering from non-uniform bias, uniform bias and error variance bias were identified utilised the technique described by Cheung and Rensvold (1999), whereby the researcher calculated, and rank ordered the configural invariance model's absolute differences in the common metric completely standardised factor loadings (or intercepts/ error variances). This process enabled the identification of items which obtained the largest absolute differences in their factor loadings, intercepts or error variances, which were freed to be estimated iteratively. It should be noted however that freeing model constraints should be done with caution. According to Vandenberg and Lance (2000, p. 56), "Partial invariance should be viewed as a critical decision point when the researcher must carefully consider whether to engage in the practice in the first place". They propose that model constraints should only be lessened under certain conditions. These include, that the decision is supported by a thorough theoretical argument; only a few items have been identified as problematic; and when the feasibility of lessening constraints is supported by cross-validation evidence (Vandenberg & Lance, 2000). These recommendations were taken into consideration when analysing the results in the present study and will be discussed in the following chapter.

In the subsequent MI and ME analyses, the evaluation of partial MI or ME would enable the identification of different forms of item bias. Item bias, or differential item functioning, is evident when the regression of a particular item on the latent variable differs across groups, when the respondents have the same standing on the latent variable (Byrne & Watkins, 2003; Van de Vijver & Rothmann, 2004). Hence, a differential response is elicited as the meaning of the items is not the same across the respective groups (Meiring et al., 2005). The forms of items bias inspected in the analyses differ depending on the parameter in question, where differences in the slope of the regression of the items on the indicator variables would indicate non-uniform bias; differences in the regression intercepts would indicate uniform bias, while differences in the error variance terms would suggest some items suffering from error variance bias (Van de Vijver & Poortinga, 1997; Vandenberg & Lance, 2000).

The decision to evaluate partial weak invariance would therefore enable the identification of items suffering from non-uniform bias. "Non-uniform bias indicates cross-cultural differences in item difficulty or endorsement rate which is not the same across the similar ability or attitude range" (Van de Vijver & Poortinga, 1997, p. 30). This suggests that the item's discriminatory power differs across the cultural groups, which is causing the biased result. For example, if a mathematics test was conducted across two schools, but the one school had not completed the curriculum yet, if a question

was asked relating to content which was not familiar to the students, it would result in a steep curve in scores for the school which had covered the content, while the other school's responses may not increase on the particular item. The item may therefore be appropriate to the one group, but not the other, resulting in an inconsistent item response (Van de Vijver & Poortinga, 1997; Byrne & Watkins, 2003). Demonstrating partial weak invariance would therefore indicate evidence of specific items containing non-uniform bias (Wu et al., 2007).

Once weak or partial weak invariance has been shown, the following step may commence, i.e. testing for metric equivalence. Metric equivalence is assessed by judging the practical significance of the difference in fit between the MI2 and MI1 multi-group measurement models (Cheung & Rensvold, 2002). The indices that will be considered to judge the practical significance of the difference in fit were provided in Table 2. Demonstrating metric equivalence would provide further evidence of a lack of non-uniform bias (Dunbar et al., 2011). Here, it is likely to observe differences (albeit subtle) in the slope of the regression of some items on the latent variables between the Black, White and Coloured groups. If significant differences in the practical significance indices are evident between the MI1 and MI2 models, resulting in the Cheung and Rensvold (2002) criteria not being met, partial metric equivalence would be inspected. Demonstrating partial metric equivalence would provide further evidence that specific items may suffer from non-uniform bias (Dunbar et al., 2011; Vandenberg & Lance; 2000; Wu et al., 2007).

Demonstrating metric (or partial metric) equivalence, would warrant further tests of MI and ME, namely testing strong invariance (MI3) and scalar equivalence. Consequently, Step 4a assesses whether the MI3 model displays acceptable fit. If the null hypothesis of close fit ($H_{05}: RMSEA \leq 0.05$) is not rejected, this would provide evidence of a lack of uniform bias (Wu et al., 2007). If, however $H_{05}: RMSEA \leq 0.05$ was rejected due to some of the intercept terms differing significantly, this would warrant the investigation of partial strong invariance. This process would enable the identification of items suffering from uniform bias. According to Van de Vijver and Poortinga (1997, p. 30), "Uniform bias occurs when the item score for examinees with a certain test score is lower in one of the populations across the entire range of test scores." For example, if we conducted a geography test in a school, where one question related to the Capital of Japan, a Japanese student would be more likely to answer correctly than a South African student, irrespective of their overall performance on the assessment. Group membership therefore has a consistent effect on individuals' performance on an item (Van de Vijver & Poortinga, 1997; Byrne & Watkins, 2003).

Once strong or partial strong invariance has been shown, the next step in the analyses would be to assess scalar equivalence. If it is found that the difference between the MI1 and the MI3 models meet

the cut-off values depicted in Table 3.2, scalar equivalence would be demonstrated (Cheung & Rensvold, 2002). At this stage, demonstrating scalar equivalence would provide further evidence of a lack of uniform bias (Van de Vijver & Poortinga, 1997). If the practical significance of the difference in fit exceeds the critical values however, it would indicate that a significant variance exists in the intercept of the regression of particular items on the latent variables between the Black, White and Coloured groups. Such a finding would warrant the inspection of partial scalar equivalence (Dunbar et al., 2011). Obtaining partial scalar equivalence would suggest that specific items may suffer from uniform bias (Byrne & Watkins, 2003; Van de Vijver & Poortinga, 1997; Wu et al., 2007).

The final level of MI and ME analyses examines whether the strict invariance multi-group model (MI4) demonstrates reasonable fit. This is assessed by testing the null hypothesis of close fit (H_{06} : $RMSEA \leq 0.05$), namely whether the regression slopes, error variances and intercepts vary between the Black, White and Coloured groups. A finding of strict invariance would therefore indicate that the different groups responded in such a manner that no significant differences were evident in the error terms relating to the indicator variables (Dunbar et al., 2011). If H_{06} : $RMSEA \leq 0.05$ would be rejected due to significant differences being observed in some of the error variance terms, partial strict invariance would be evaluated. This process would allow the items affected by error variance bias to be identified. While demonstrating strict invariance would provide evidence for a lack of error variance bias, obtaining partial strict invariance would indicate specific items suffering from error variance (residual) bias (Van der Bank, 2019; Wu et al., 2007).

Demonstrating strict invariance (or partial strict) invariance will warrant testing for conditional probability equivalence. Testing conditional probability equivalence involves evaluating the variance between the MI4 and MI1 measurement models. If the differences are in line with the values stipulated in Table 3.2 (Cheung & Rensvold, 2002), conditional probability equivalence would be demonstrated providing evidence of a lack of further error variance bias (Van der Bank, 2019). If the difference between the MI4 and MI1 models was greater than the critical values, it would indicate that differences exist in the error variance of some items, across the Black, White and Coloured groups. In this case, partial conditional probability equivalence would be inspected using the method discussed previously. Demonstrating partial conditional probability equivalence would serve as evidence that specific items may suffer from error variance bias (Van der Bank, 2019).

3.7 Ethical Considerations

Research done in the field of Industrial Psychology requires important considerations apart from matters pertaining to the epistemic ideal of science. Specifically, research involving individuals requires certain ethical considerations and risk analyses and weighing the benefits of the research

against these. The following section will elaborate on ethical principles to uphold during research, as well as potential ethical risks to consider.

3.7.1 Risk-Benefit Analysis

Prior to considering potential ethical issues related to the present study, it is important to ascertain whether the research is worthwhile conducting. While every precaution may be taken to protect participants, their involvement could still compromise their safety, rights, well-being or dignity, and thus it is necessary to evaluate whether the ends justify the means (Stellenbosch University, 2013). Consequently, the question arises whether the benefits sought from the research are balanced with the costs incurred by participants. There are factors involved in research which could impact the participant, and care should be taken to ascertain that these do not result in unnecessary risk. For example, the researcher needs to ensure that participants' anonymity and confidentiality are maintained. The precautionary measures that were taken to uphold ethical standards in the present study will be discussed.

The current research aimed to add value by contributing to the validation of the PCQ-24 in South Africa, by establishing whether the measure may be biased against Black, White or Coloured groups. It is envisioned that the current study will contribute to the body of knowledge regarding the measure and inform its responsible use in a diverse cultural / linguistic environment, such as South Africa.

3.7.2 Ethical Risks, Guiding Principles and Legislation

While the benefit of the study is understood, sound ethical principles need to be stressed as they protect the participants. This is necessary due to various ethical pitfalls which can arise during research, due to the interaction between researchers and participants in the process (Babbie & Mouton, 2001).

Annexure 12 of the Health Professions Act [Act no. 56 of 1974] relates to Ethical Rules of Conduct for Practitioners. Here it is stated that psychologists who perform research need to gain consent from their participants through an agreement which stipulates the nature of the research project and the rights and responsibilities of both parties. Consequently, the agreement would need to abide by the following provisions as stated in Annexure 12 (Republic of South Africa (RSA), 2006. p, 42):

- (1) A psychologist shall use language that is reasonably understandable to the research participant concerned in obtaining his or her informed consent; (2) Informed consent referred to in sub-rule (1) shall be appropriately documented, and in obtaining such consent the psychologist shall – (a) inform the participant of the nature of the research; (b) inform the participant that he or she is free to participate or decline to participate in or to withdraw from the research; (c) explain the foreseeable consequences of declining or withdrawing; (d) inform the participant of significant

factors that may be expected to influence his or her willingness to participate (such as risks, discomfort, adverse effects or exceptions to the requirement of confidentiality); (e) explain any other matters about which the participant enquires; (f) when conducting research with a research participant such as a student or subordinate, take special care to protect such participant from the adverse consequences of declining or withdrawing from participation; (g) when research participation is a course requirement or opportunity for extra credit, give a participant the choice of equitable alternative activities; and (h) in the case of a person who is legally incapable of giving informed consent, nevertheless – (i) provide an appropriate explanation; (ii) obtain the participants assent; and (iii) obtain appropriate permission from a person legally authorized to give such permission.

It is evident that informed consent requires careful consideration in behavioural research. Informed consent will be demonstrated if the participants are informed of: (i) the research objective and purpose; (ii) what the research will involve; (iii) how the results will be distributed and used; (iv) the researcher's identity and affiliations; (v) where participants can further inquire if they wish; and (vi) the participants' rights relating to the research and information relating thereto, and willingly decide to participate (Stellenbosch University, 2013). The researcher is also responsible for ensuring that participants are provided with information in a manner which is clear to them. Since secondary data was utilised for this study, the researcher ensured to check that the data were collected with the necessary informed consent procedures, which was demonstrated under paragraph 3.5. Evidence of this was also submitted together with the ethics application to the Departmental Ethics Screening Committee (DESC) of the Department of Industrial Psychology and the Research Ethics Committee (REC) of Stellenbosch University.

Annexure 12 of the Ethical Rules of Conduct for Practitioners Registered under the Health professions Act [Act no. 56 of 1974] (Republic of South Africa (RSA), 2006, p. 41) further mentions that the researcher needs to obtain permission from the organisation or institution from which the participants (or in this case, data) will be obtained:

A psychologist shall – (1) obtain written approval from the host institution or organisation concerned prior to conducting research; (2) provide the host institution or organisation with accurate information about his or her research proposals; and (3) conduct the research in accordance with the research protocol approved by the institution or organisation concerned.

Written permission was obtained from the researchers that collected the primary data, in order to use the data as secondary data in this research study. Proof thereof was submitted along with the ethics application to the DESC and REC.

It is also the researcher's responsibility to ensure that participants' data is kept confidential. The datasets obtained from the fellow researchers were anonymous and the composite dataset was securely stored on a password-protected computer in a password-protected file. The primary researcher and her supervisor were the only individuals who have access to the combined dataset.

3.8 Summary

Chapter 3 dissected the methodology used in testing the hypotheses of the present study. This included a discussion of the research design and the sample that was utilised in the study. Furthermore, the chapter elaborated on the statistical analysis techniques which were employed in evaluating the MI and ME of the PCQ-24. The chapter also reflected on ethical considerations relating to the present study and how these were handled. The following chapter will discuss the results of the study as well as inferences based on these.

CHAPTER 4: RESULTS

4.1 Introduction

Psychological Capital is a well-researched positive psychological construct which has been shown to be highly valuable to individuals and organisations due to its relevance to workplace outcomes (Avey et al., 2010a; Avey et al., 2010b; Avey et al., 2009; Dawkins et al., 2013), its malleable character (Luthans et al., 2007) and demonstrated performance impact (Luthans et al., 2010). Comparisons over groups in terms of their PsyCap are routinely reported in literature in terms of gender (Barmola, 2013; Rani & Chaturvendula, 2018), nationality and status (Avcı & Erdem, 2017), as well as age, language and ethnicity (Du Plessis & Barkhuizen, 2012). In the absence of invariance evidence however, it is uncertain whether the findings are as a result of true construct variance or measurement bias (Steenkamp & Baumgartner, 1998). This study aimed to contribute to the body of knowledge regarding cross-cultural assessment in South Africa by investigating the MI and ME of the PCQ-24. The purpose of this chapter is to present the research findings of the study.

4.2 Assessing the PCQ-24 Measurement Model

Various estimation techniques can be employed to estimate model parameters in LISREL, such as Maximum Likelihood (ML), Robust Maximum Likelihood (RML), Generally Weighted Least Squares (WLS) and Diagonally Weighted Least Squares Estimation (DWLS) (Diamantopoulos & Sigua, 2000; Jöreskog & Sörbom, 1996). The decision regarding the most appropriate estimation technique will depend on the nature of the variables in question and the manner in which the data is distributed.

It has been argued in this study that Likert scales with five or more data points, such as is utilised in the PCQ-24, could be regarded as approximating continuous data, based on research conducted by Muthén and Kaplan (1985) and further supported in later studies (Norman, 2010; Sullivan & Artino, 2013; Zumbo & Zimmerman, 1993). Therefore, the data in the present study was specified as continuous. Furthermore, the data was fitted on item level, as opposed to using item parcels. While there are many benefits to using item parcels, e.g., addressing small sample sizes, non-normality and unstable parameter estimates (Bandalos et al., 2001; Bandalos, 2002; Bonnacio & Reeve, 2006; Little et al, 2002; Little et al, 2013), the goal of the study was not conducive to using parcels. The current study aimed to investigate the MI and ME of the PCQ-24 at the item level. As the use of item parcels would not produce parameter estimates for the individual items of the PCQ-24, this method would not be suitable. Furthermore, it is possible that the effects of bias could be misinterpreted when using item parcels, resulting in considerable differences between groups being masked by parcels (Meade & Kroustalis, 2005).

When selecting the appropriate estimation technique, the distributional properties of the data were also considered in light of the assumptions underlying the estimation techniques. According to Diamantopoulos and Siguaw (2000, p 3), ML estimation is the default estimation technique in LISREL 8 and is more widely used in practice. This estimation technique assumes that the variables follow a multivariate normal distribution, when fitted to continuous data (Diamantopoulos & Siguaw, 2000; Muthén & Kaplan, 1985; Mels, 2003). In cases where the assumption is not met however, RLM is recommended as the preferred technique (Tabachnick & Fidell, 2013). This is due to the fact that this method minimises estimation problems associated with a lack of multivariable normality, such as the underestimation of fit indices (Mels, 2003). The assumption of multivariate normality was therefore inspected via PRELIS in LISREL 8.8 in consideration of the appropriate estimation technique to utilise in the present study. The findings for the data of the three ethnic groups are shown in Table 4.1 below.

Table 4.1

Test of multivariate normality for continuous variables

Population Group	Skewness			Kurtosis			Skewness and Kurtosis	
	Value	Z-Score	P-Value	Value	Z-Score	P-Value	χ^2	P-Value
Black	207.390	46.060	0.000	837.231	17.704	0.000	2434.934	0.000
White	176.532	49.186	0.000	849.472	20.176	0.000	2826.372	0.000
Coloured	17.586	42.276	0.000	836.363	18.363	0.000	2138.270	0.000

It is evident from the results that the evaluation of multivariate normality of the indicator variables resulted in the null hypothesis being rejected for all three groups ($p < 0.05$). This indicated that the data was not normally distributed and hence, RML estimation was employed as the appropriate estimation technique.

4.3 Assessing the PCQ-24 Single-Group Measurement Models via CFA

The PCQ-24 single-group measurement models are required to show good fit prior to assessing MI and ME over the different groups. The PCQ-24 measurement model was fitted using SEM. CFA was performed, and the measurement model was specified consisting of six observed variables (χ 's) per each latent PsyCap dimension (ξ i.e. Self-efficacy, Hope, Optimism, Resilience). The proposed regression of the observed variables onto the latent dimensions were indicated with single-headed arrows from the ξ to the χ 's in the measurement model. The following fit indices were consequently derived from CFA via LISREL.

4.3.1 Measurement Model Fit Indices

The null hypotheses of exact fit (H01i: RMSEA = 0; $i = 1_{\text{Black}} 2_{\text{White}} 3_{\text{Coloured}}$) and close fit (H02i: RMSEA \leq 0.05; $i = 1_{\text{Black}} 2_{\text{White}} 3_{\text{Coloured}}$) for the three single-group measurement models were tested. The results for the three groups will be discussed separately.

4.3.1.1 Black sample

Table 4.2 demonstrates that for the Black group, the exact fit null hypothesis (H01i: RMSEA = 0; $i = 1_{\text{Black}}$) was rejected in favour of the alternative hypothesis (H02i: RMSEA \leq 0.05; $i = 1_{\text{Black}}$). The close fit null hypothesis however was not rejected, and therefore the results indicated that close fit was obtained by the Black single-group measurement model. As H02i: RMSEA \leq 0.05; $i = 1_{\text{Black}}$ could not be rejected for the Black group, the position that the PCQ-24 measurement model fitted the Black group closely in the parameter was permissible. Moreover, the RMSEA value of 0.055 further underscored this conclusion (Hair et al., 2010). The NNFI and CFI are comparative fit indices of which scores of \geq 0.95 and above are considered to indicate good fit, while values of \geq 0.97 are considered very good fit (Schermelleh-Engel et al., 2003; Tabachnick & Fidell, 2013). From the table below, it is evident that the NNFI (0.943) and CFI (0.949) values marginally missed the 0.95 cut off, indicating good model fit. The SRMR provides an indication of the average residual matrix value. In line with the recommendation by Tabachnick and Fidell (2013) values of 0.08 and less are desirable, while smaller values indicate good fit. Values which exceed 0.10 therefore will be indicative of poor fit (Hair et al., 2010). Lastly, The SRMR value of 0.078 provided further evidence in support of good fit. Overall, therefore, it could be concluded that the GOF results for the Black single-group measurement model provided evidence of good fit.

Table 4.2

Summary of GOF indices: Black sample

Hypotheses Tested	DF	RMSEA	90 % (RMSEA)	P (Close)	χ^2	P (Exact)	F0	NNFI	CFI	SRMR
H01 & H02	246	0.0551	0.0455; 0.0644	0.185	575.584	0.000	0.747	0.943	0.949	0.0783

4.3.1.2 White sample

The results of the measurement model analysis for the White sample (Table 4.3) indicated that the exact fit null hypothesis (H01i: RMSEA = 0; $i = 2_{\text{White}}$) was rejected in favour of the alternative hypothesis (H02i: RMSEA \leq 0.05; $i = 2_{\text{White}}$). The close fit null hypothesis was not rejected however, as $p > 0.05$ (0.153) which indicated that close fit was obtained for the White single-group measurement model. As H02i: RMSEA \leq 0.05; $i = 2_{\text{White}}$ could not be rejected for the White group, it can be argued

that the PCQ-24 measurement model fitted the White group closely in the parameter. This was supported by the RMSEA value (0.055), indicating good model fit. Furthermore, consideration of the additional GOF indices indicated that the White group obtained NNFI (0.977) and CFI (0.979) values which far exceeded 0.95. As both indices were above 0.97, they provide further evidence in support of very good model fit (Schermelleh-Engel et al., 2003). Lastly, in terms of the SRMR, the White group value of 0.066 indicates that good fit had been achieved (Hair et al., 2010). Consequently, it can be concluded that the overall GOF results for the White single-group measurement model indicated good fit.

Table 4.3

Summary of GOF Indices: White sample

Hypotheses Tested	DF	RMSEA	90 % (RMSEA)	P (Close)	X2	P (Exact)	FO	NNFI	CFI	SRMR
H ₀₁ & H ₀₂	246	0.055	0.046; 0.063	0.153	664.643	0	0.744	0.977	0.979	0.066

4.3.1.3 Coloured sample

Inspection of the GOF results for the Coloured sample (Table 4.4) revealed that the exact fit null hypothesis (H_{01i}: RMSEA = 0; $i = 3_{\text{Coloured}}$) was rejected in favour of the alternative hypothesis (H_{02i}: RMSEA \leq 0.05; $i = 3_{\text{Coloured}}$). As was the trend with the results from the other two groups, the close fit null hypothesis however was not rejected, as $p = 0.521$ ($p > 0.05$) and hence close fit could be concluded for the Coloured single-group measurement model. Moreover, the RMSEA for the Coloured measurement model provided further evidence in support of model fit, as the value of 0.049 indicated good model fit (RMSEA \leq 0.05). In terms of the NNFI (0.977) and CFI (0.982), it is evident that both indices far exceeded the 0.95 criterion, as well as the 0.97 criterion, further supporting very good model fit. In addition, a satisfactory result was also obtained on the SRMR (0.065) indicating good fit. Taken together, as the Coloured measurement model obtained close fit, a favourable RMSEA value, as well as additional GOF indices (NNFI, CFI and SRMR) in favour of the model's fit, it can be concluded that the Coloured single-group measurement model obtained very good fit overall.

Table 4.4

Summary of GOF Indices: Coloured sample

Hypotheses Tested	DF	RMSEA	90 % (RMSEA)	P (Close)	X2	P (Exact)	FO	NNFI	CFI	SRMR
H ₀₁ & H ₀₂	246	0.049	0.040; 0.058	0.521	553.469	0	0.604	0.977	0.982	0.065

4.3.2 Measurement Model Residuals

The variation between the observed and fitted covariance matrices in terms of their equivalent indicator variables are termed measurement model residuals (Diamantopoulos & Siguaw, 2000). Standardised residuals are obtained by dividing the residuals by their estimated standard errors (Jöreskog & Sörbom, 1993). The measurement model residuals and standardised residuals provide valuable insights to understanding the extent of a model's misfit (Kelloway, 1998). Although the ideal is that measurement model residuals should be dispersed evenly around zero, misalignment will be evident in cases where the negative or positive residuals have absolute values which are large (i.e. surpassing the $|2.58|$ cut-off) and higher than zero (Diamantopoulos & Siguaw, 2000). A high volume of residuals observed on either side of zero would indicate that the variance and covariance terms were systematically over- or underestimated. Here, large negative residuals will indicate overestimation and hence a need for the paths connecting to the related indicator variables to be removed, while large positive residuals will indicate underestimation, demonstrating that additional paths between the indicator variables and latent variables are required (Diamantopoulos & Siguaw, 2000). The standardised residuals will be discussed for each single-group measurement model.

4.3.2.1 Standardised residuals: Black measurement model

The summary of standardised residuals for the Black measurement model (Table 4.5) indicates that eight large positive residuals and nine large negative residuals were observed. The large residuals therefore constituted 5.66%¹⁴ of the total unique variance and covariance terms in the variance-covariance matrix. Consequently, approximately 6% of the total unique variance and covariance terms were inaccurately estimated from the measurement model parameter estimates. While this is not ideal, the percentage is still relatively small and considered acceptable.

¹⁴ The percentage was calculated as the total large residuals out of the total terms in the variance-covariance matrix (300). The variance-covariance matrix was calculated as $24 \times 25 / 2$.

Table 4.5*Summary: Standardised residuals for Black measurement model*

Description	Value
Smallest Standardized Residual	-6.301
Median Standardized Residual	0.000
Largest Standardized Residual	5.140
Largest Positive standardised residuals	
Residual for PCQ9 and PCQ1	2.585
Residual for PCQ20 and PCQ4	2.636
Residual for PCQ14 and PCQ9	2.686
Residual for PCQ18 and PCQ9	3.504
Residual for PCQ20 and PCQ13	5.140
Residual for PCQ23 and PCQ13	3.843
Residual for PCQ19 and PCQ18	2.592
Residual for PCQ23 and PCQ20	4.837
Largest Negative standardised residuals	
Residual for PCQ5 and PCQ3	-2.723
Residual for PCQ19 and PCQ6	-3.218
Residual for PCQ12 and PCQ9	-6.301
Residual for PCQ23 and PCQ9	-3.531
Residual for PCQ17 and PCQ10	-3.208
Residual for PCQ15 and PCQ11	-3.463
Residual for PCQ15 and PCQ13	-3.042
Residual for PCQ20 and PCQ9	-2.592
Residual for PCQ24 and PCQ23	-2.958

Furthermore, from Figure 4.1 below, the stem-and-leaf plot for the Black measurement model indicates that a slightly positively skewed distribution was obtained, as few large negative residuals were observed that were not observed on the positive side. The slightly positively skewed distribution indicates that the Black single-group measurement model was inclined to underestimate the covariance terms slightly more than it overestimated them (Diamantopoulos & Sigauw, 2000).

Figure 4.1*Stem-and-leaf plot of standardised residuals for the Black measurement model*

Stemleaf Plot

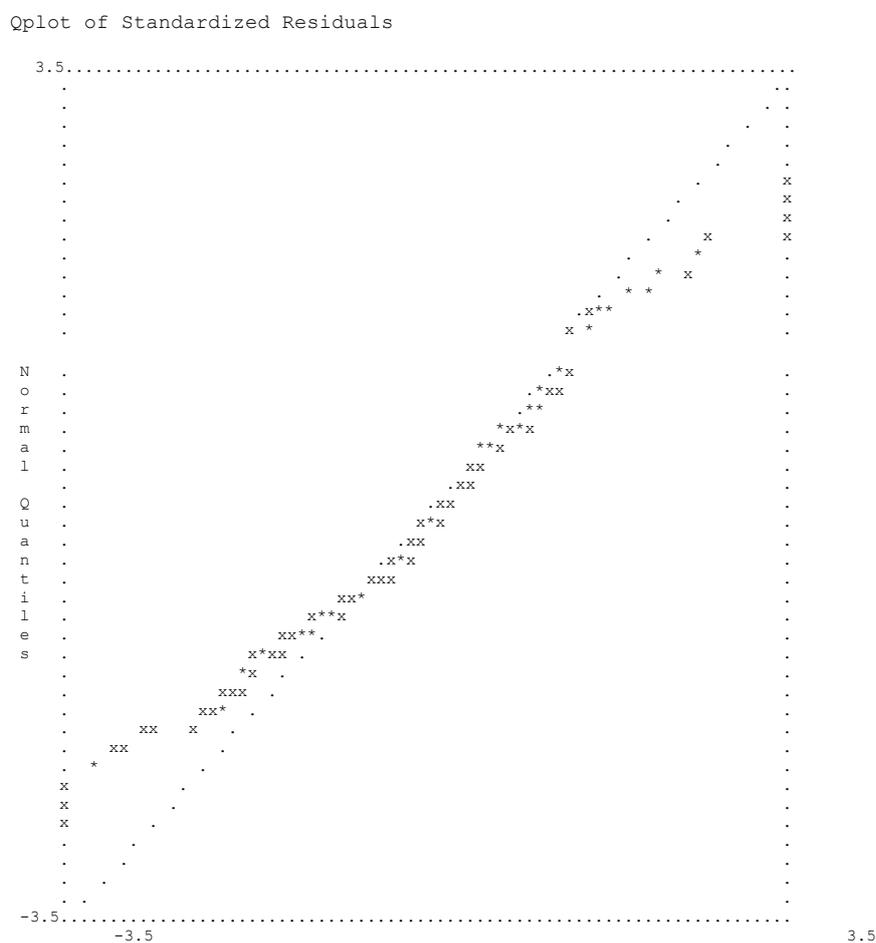
```

- 6|3
- 5|
- 4|
- 3|552200
- 2|762210
- 1|9999888877776665555444433222211111000000
- 0|9999888877777666666555544444333333222222222211111111111110000000+27
0|111111111111222222222333333333344444444445555555555555556666666667+12
1|0000000111111112222222333333333334456666667788
2|0022236667
3|58
4|8
5|1

```

Another manner of displaying the standardised residuals graphically is using a Q-plot which illustrates them against the quantiles of a normal distribution (Diamantopoulos & Sigua, 2000). The model will demonstrate good fit if the data points follow the 45-degree reference line closely (Jöreskog & Sörbom, 1993). The Q-plot of the Black measurement model is depicted in Figure 4.2. Upon observation, it is apparent that the data points deviated only slightly from the upper and lower regions of the X-axis, signifying that the measurement model showed reasonable to good fit. These findings were in line with the results depicted in the fit statistics, which also indicated reasonable to good model fit.

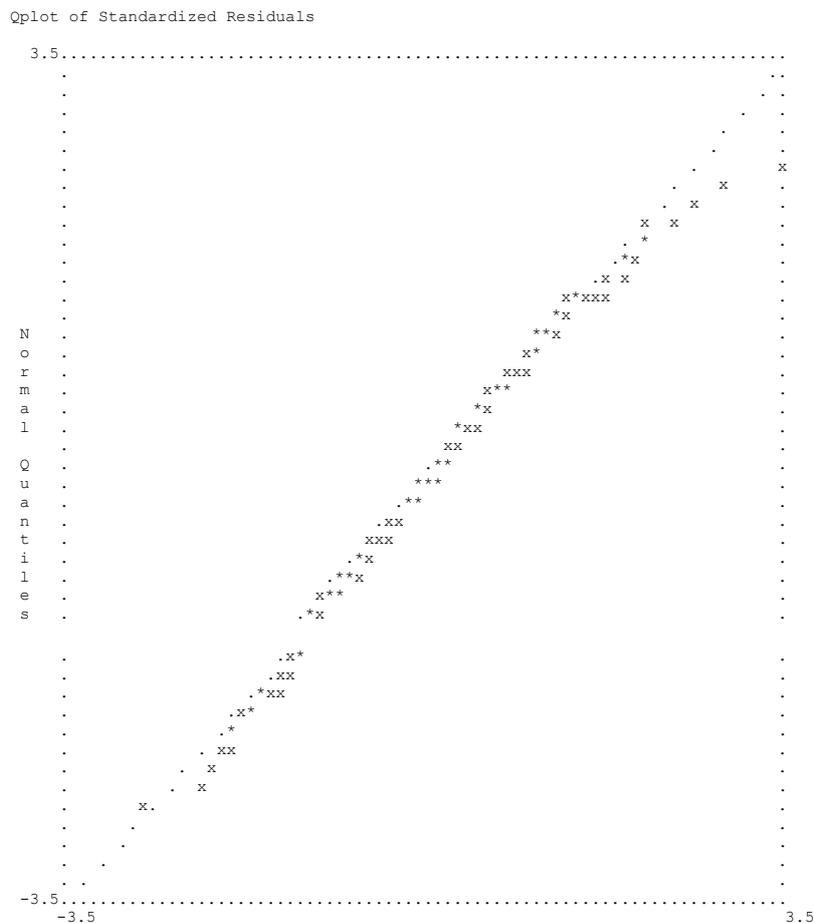
Figure 4.2
Q-plot of standardised residuals for the Black measurement model



4.3.2.2 Standardised residuals: White measurement model

The results of the standardised residuals derived for the White single-group measurement in Table 4.6 revealed that only three large positive residuals and one large negative residual was observed. This equated to 1.33% or approximately 1% of the total unique variance and covariance terms, which were

Figure 4.4
Q-plot of standardised residuals for the White measurement model



4.3.2.3 Standardised residuals: Coloured measurement model

Lastly, the summary of standardised residuals for the measurement model fit on the Coloured data indicated that six large positive residuals and four large negative residuals were obtained. The large standardised residuals equated to 3.33% of the total unique variance and covariance terms in the observed variance-covariance matrix. Evidently, roughly only 3% of the total variance and covariance terms were imprecisely estimated from the measurement model parameter estimates, which is a relatively small and acceptable amount.

Table 4.7*Summary: Standardised residuals for Coloured measurement model*

Description	Value
Smallest Standardized Residual	-4.263
Median Standardized Residual	0.000
Largest Standardized Residual	6.201
Largest Positive standardised residual	
Residual for PCQ7 and PCQ2	2.618
Residual for PCQ7 and PCQ6	2.653
Residual for PCQ9 and PCQ7	2.733
Residual for PCQ20 and PCQ13	5.025
Residual for PCQ23 and PCQ13	3.347
Residual for PCQ23 and PCQ20	6.201
Largest Negative standardised residual	
Residual for PCQ11 and PCQ5	-2.592
Residual for PCQ10 and PCQ9	-4.263
Residual for PCQ24 and PCQ13	-2.618
Residual for PCQ24 and PCQ23	-2.778

Figure 4.5 shows the stem-and-leaf plot for the Coloured measurement model. Here, a slightly positively skewed distribution was again observed, with a few large negative residuals observed that were not observed on the positive side. The slightly positively skewed distribution indicates that the Coloured single-group measurement model was inclined to underestimate the covariance terms slightly more than it overestimated them (Diamantopoulos & Siguaw, 2000), suggesting reasonable to good model fit.

Figure 4.5*Stem-and-leaf plot of standardised residuals for the Coloured measurement model*

Stemleaf Plot

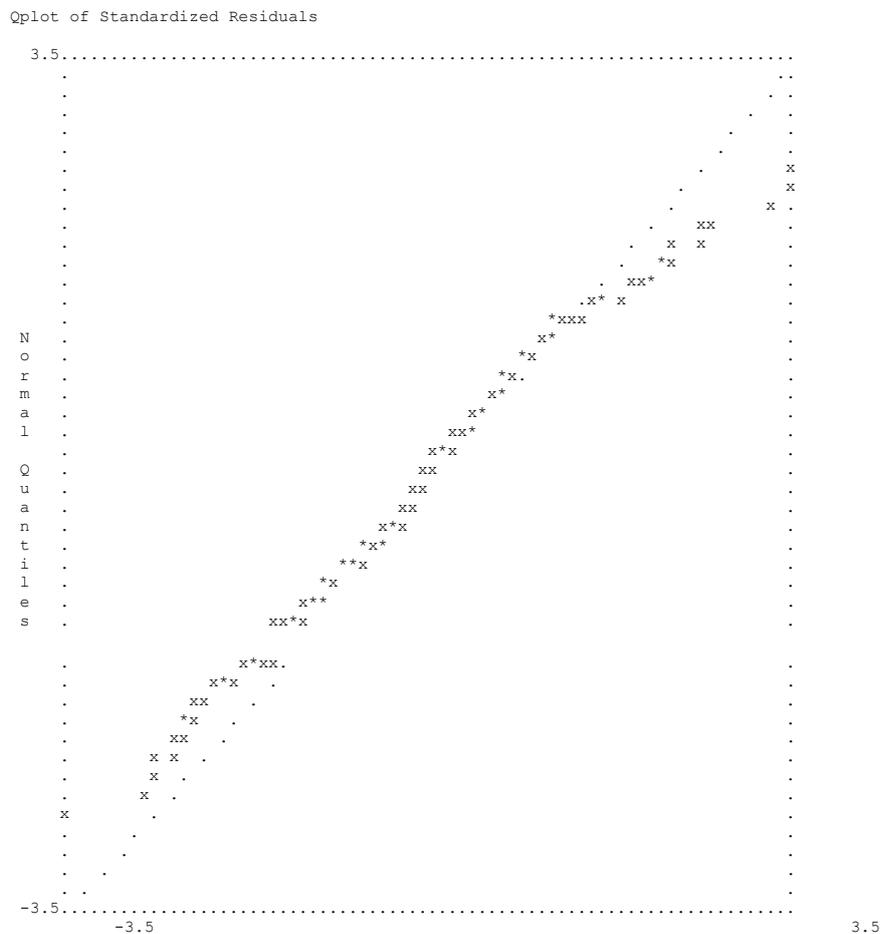
```

- 4|3
- 3|
- 2|866544443221100
- 1|9998887776554333322211111110000
- 0|99999999999887777777666666665555554444444433333333333322222222+51
  0|111111111111222222222233333333444444444455555566666666667777778+15
  1|00000111122222233345556778
  2|00122234677
  3|3
  4|
  5|0
  6|2

```

The Q-plot of the Coloured measurement model is depicted in Figure 4.6 below. Upon observation, it is evident that the data points deviated only slightly from the upper and lower regions of the X-axis, signifying that the measurement model showed good fit. These findings were in line with the conclusions drawn from the fit statistics, which also suggested good model fit.

Figure 4.6
Q-plot of standardised residuals for the Coloured measurement model



4.3.3 Parameter Estimates for the Single-group Measurement Models

When evaluating the measurement model, the researcher inspects the relationship between the indicators and the latent variables. Inspection of the magnitude and the significance of the paths between these variables provides the researcher with an indication of the validity of a measure (Diamantopoul & Siguaw, 2000). The parameter estimates for the three single-group measurement models are reflected in the tables below. Firstly, the unstandardized lambda-X matrix will be inspected, where factor loadings with t-values that exceed the critical value of 1.6649 will be considered significant ($p < .05$) (Diamantopoul & Siguaw, 2000). Diamantopoul and Siguaw (2000, p 9) advise however:

“One problem with relying on unstandardized loadings and associated t-values is that it may be difficult to compare the validity of different indicators measuring a particular construct. This problem arises because indicators of the same construct may be measured on very different scales; if this is the case, then direct comparisons of the magnitudes of the loadings are clearly inappropriate. In addition, bearing in mind that each latent variable has to be assigned a scale by

fixing the loading of one of its indicators to unity, the loadings of the other indicators for that latent variable are only interpretable relative to the unit of the reference indicator. Clearly, if a different indicator is used as the reference variable, the magnitudes of the loadings will change. For these reasons, it is recommended that the magnitudes of the *standardized* loadings are also inspected.”

Therefore, the results for the unstandardised lambda-X matrix as well as the completely standardised lambda-X matrix will be discussed for each ethnic group. When analysing the completely standardised lambda-X matrix, some authors (e.g. Hair et al., 2010) argue that factor loadings should be considered satisfactory if the estimates exceed 0.70. However, this criterion value is considered quite stringent, and therefore others suggest that a factor loading which exceeds 0.40 can be considered acceptable (Kline, 2013). According to Cattell (1978) however, factor loadings which exceed 0.15 could be considered potentially satisfactory. This moderate criterion is considered in cases where analyses are conducted on the item level (which is the case in the present study), which obtain lower reliability statistics in comparison to subscale level analyses (Gignac, 2005). The unstandardized and completely standardised parameter estimates for the Black, White and Coloured single-group measurement models are discussed below.

4.3.3.1 Black sample

Table 4.8 reveals that for the Self-efficacy and Hope subscales, all items exceeded the $|1.6649|$ critical value and were statistically significant. For the Resilience subscale, item 13 obtained a non-significant factor loading (0.494). Item 13 has been shown to be a problematic item throughout the reliability analyses. Further inspection of the Resilience subscale revealed that the remaining items on this subscale were all significant (exceeding $|1.6649|$), however they all returned significant negative factor loadings. This is a noteworthy finding which has not previously been reported in PsyCap research. Obtaining significant negative factor loadings for items 14 (-0.561), 15 (-0.482), 16 (-0.420), 17 (-0.533) and 18 (-0.550), indicates that the latent variable scale may be “flipped” for the Black sample group. This would imply that that higher positions on the latent trait result in individuals choosing the lower end of the response scale, i.e. disagreeing with the item statement. It is unclear whether there is a cultural convention which would move Black South Africans to answer in this manner, however the results suggest that Resilience manifests in Black South Africans differently. The table below provides the item content of the implicated items.

Table 4.8*Item content of Resilience subscale: Significant negative factor loadings*

Items	Item content
Item 14	"I usually manage difficulties one way or another at work."
Item 15	"I can be "on my own", so to speak at work if I have to."
Item 16	"I usually take stressful things at work in stride."
Item 17	"I can get through difficult times at work because I have experienced difficulty before."
Item 18	"I feel that I can handle many things at a time at this job."

Considering the initial sample which the Black data was drawn from, i.e. a sample of Black middle managers (Roux, 2014), the current results are rather perplexing. It may be reasonable to argue that the results may not be due to the respondents having a lower level of education; resulting in them not understanding the scale items. This is since the Black sample consisted of relatively well-educated individuals (i.e. Most Black respondents held a Diploma [33%], followed by a Bachelors Degree [22%] and Grade 12/Equivalent [20%]. A further 10% held a Higher certificate, 9% held a Honours Degree and the remaining 7% had either obtained a Masters or Doctoral Degree; Roux, 2014). Another possible explanation may be that these findings are due to cultural factors which are impacting the measurement of the Resilience construct, which will be discussed in Chapter 5. The fact that significant negative factor loadings were obtained will have implications in the subsequent analyses, which will be highlighted in the following section.

Lastly, for the Optimism subscale, most items obtained significant factor loadings, except items 20 and 23; the two negatively keyed items in the subscale. It is noted in this case that all the negatively keyed items did not load significantly onto the factors they were intended to reflect. This finding correlates with previous research on the PCQ-24 which indicates that the negatively keyed items perform consistently poorer than the other items (Dawkins, et al., 2013; Gørgens-Ekermans & Herbert, 2013).

Table 4.9*Unstandardized lambda-X matrix for the Black measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.341 (0.064)			
	5.340			
PCQ2	0.665 (0.065)			
	10.271			
PCQ3	0.744 (0.071)			
	10.411			
PCQ4	0.696 (0.068)			
	10.179			

PCQ5	0.598 (0.092) 6.504		
PCQ6	0.495 (0.074) 6.678		
PCQ7		0.430 (0.067) 6.438	
PCQ8		0.498 (0.072) 6.929	
PCQ9		0.314 (0.049) 6.452	
PCQ10		0.818 (0.091) 9.000	
PCQ11		0.683 (0.070) 9.780	
PCQ12		0.673 (0.072) 9.315	
PCQ13			0.063 (0.127)
PCQ14			0.494 -0.444 (0.066) 6.726
PCQ15			-0.500 (0.102) 4.893
PCQ16			-0.524 (0.112) 4.665
PCQ17			-0.505 (0.100) 5.058
PCQ18			-0.414 (0.069) 6.033
PCQ19			0.548 (0.084) 6.533
PCQ20			-0.157 (0.122) -1.286
PCQ21			0.574 (0.065) 8.840
PCQ22			0.744 (0.097) 7.705
PCQ23			0.039 (0.114) 0.341

PCQ24	0.492 (0.080) 6.109
-------	----------------------------------

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc). Bold values indicate statistically significant factor loadings.

Table 4.9 demonstrates that for the Self-efficacy subscale, all items obtained completely standardised loadings above 0.40, besides item 1, which fell marginally below at 0.386. Furthermore, three of the Self-efficacy items exceeded the stringent 0.70 value (items 2 – 4). In terms of the Hope subscale, three items either exceeded or approximated the 0.70 critical value, ranging from 0.687 (item 12) to 0.817 (item 11). The remaining three items all exceeded 0.40, ranging from 0.45 (item 9) to 0.539 (item 8). This is a positive finding for the Self-efficacy and Hope subscales which indicates that the respective items provided a relatively valid reflection of the factors they were intended to measure. For the Resilience subscale, item 13 obtained an insignificant factor loading, missing the 0.40 value as well as the 0.15 critical value (Kline, 2013; Cattell, 1978). The remaining items on the subscale all exceeded 0.40 in their magnitude, however negative values were evident for all these items, a phenomenon that does not coincide with the PsyCap literature and the rest of the results for the other groups in this study. For the Optimism subscale, four of the items obtained satisfactory factor loadings ranging from 0.439 (item 24) to 0.714 (item 21). The remaining two items obtained insignificant loadings, at 0.027 (item 23) and -0.121 (item 20) respectively. Again, these were the negatively keyed items.

Table 4.10*Completely standardised lambda-X matrix for the Black measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.386			
PCQ2	0.758			
PCQ3	0.714			
PCQ4	0.736			
PCQ5	0.551			
PCQ6	0.627			
PCQ7		0.523		
PCQ8		0.539		
PCQ9		0.450		
PCQ10		0.760		
PCQ11		0.817		
PCQ12		0.687		
PCQ13			0.042	
PCQ14			-0.561	
PCQ15			-0.482	
PCQ16			-0.420	
PCQ17			-0.533	
PCQ18			-0.550	
PCQ19				0.498
PCQ20				-0.121
PCQ21				0.714
PCQ22				0.673
PCQ23				0.027
PCQ24				0.439

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc).

4.3.3.2 White sample

Table 4.10 reveals that for the White measurement model, items for the Self-efficacy and Hope subscales all loaded significantly onto the respective factors (Self-efficacy: PCQ1-6; Hope: PCQ7 – 12). Item 13 in the Resilience subscale, in contrast to the results on the Black sample, obtained a significant loading. Furthermore, the remaining items on the Resilience subscale (PCQ14 – 18) all obtained significant, positive factor loadings. In terms of the Optimism subscale, evidently all items emerged with statistically significant values. Again, in contrast with the Black sample, the negatively keyed items in this subscale (items 20 and 23) obtained significant loadings.

Table 4.11*Unstandardized lambda-X matrix for the White measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.487 (0.073) 6.688			
PCQ2	0.813 (0.072) 11.362			
PCQ3	0.915 (0.069) 13.304			
PCQ4	0.868 (0.075) 11.649			
PCQ5	0.775 (0.085) 9.114			
PCQ6	0.827 (0.078) 10.635			
PCQ7		0.626 (0.078) 8.013		
PCQ8		0.865 (0.069) 12.507		
PCQ9		0.501 (0.083) 6.063		
PCQ10		0.807 (0.078) 10.291		
PCQ11		0.854 (0.069) 12.371		
PCQ12		0.921 (0.071) 12.943		
PCQ13			0.247 (0.117) 2.120	
PCQ14			0.555 (0.070) 7.932	
PCQ15			0.532 (0.080) 6.652	
PCQ16			0.703 (0.073) 9.628	
PCQ17			0.665 (0.075) 8.844	
PCQ18			0.696	

	(0.064)	
	10.926	
PCQ19		0.813 (0.069) 11.863
PCQ20		0.310 (0.111) 2.797
PCQ21		0.724 (0.077) 9.356
PCQ22		0.793 (0.090) 8.776
PCQ23		0.477 (0.096) 4.961
PCQ24		0.593 (0.096) 6.203

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc). Bold values indicate statistically significant factor loadings.

Inspection of the completely standardised factor loadings revealed that five out of the six items for the Self-efficacy subscale either exceeded or approximated the 0.70 cut-off value (PCQ2-6). Item 1 (PCQ1) fell slightly below this criterion, however the factor loading was still above 0.50 (.576) which is considered satisfactory. The Hope subscale also produced a positive result, as four of the items exceeded the stringent 0.70 critical value (PCQ8; PCQ10 - 12), while the remaining two items were above 0.50 (PCQ7; PCQ9). In terms of the Resilience subscale, five out of the six items obtained a satisfactory result with factor loadings above 0.50, ranging from 0.555 (item 15) to 0.729 (item 18). However, item 13 emerged with low loading of 0.184. This is clearly out of synch with the results of the results for this subscale, and possibly points towards some effect of the negatively keyed nature of this item – although not nearly as pronounced in this group as was evident from the results of the Black group. Finally, for the Optimism subscale, although no items exceeded the stringent 0.70 cut-off value of Hair et al. (2010); four items exceeded or approximated 0.50 (PCQ19; PCQ21 – 22; PCQ24), indicating a satisfactory result. The remaining two items (PCQ20; PCQ23 – the negatively keyed items in the subscale) obtained values slightly below 0.40, at 0.243 and 0.380 respectively. Although the results could still be considered acceptable, the finding for the White group is similar to the previous group, illustrating the detrimental effect of the negatively keyed items across both samples.

Table 4.12*Completely standardised lambda-X matrix for the White measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.576			
PCQ2	0.810			
PCQ3	0.796			
PCQ4	0.849			
PCQ5	0.705			
PCQ6	0.721			
PCQ7		0.631		
PCQ8		0.764		
PCQ9		0.514		
PCQ10		0.725		
PCQ11		0.800		
PCQ12		0.764		
PCQ13			0.184	
PCQ14			0.654	
PCQ15			0.555	
PCQ16			0.699	
PCQ17			0.656	
PCQ18			0.729	
PCQ19				0.682
PCQ20				0.243
PCQ21				0.705
PCQ22				0.693
PCQ23				0.380
PCQ24				0.475

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc).

4.3.3.3 Coloured sample

Examination of the results for the Coloured sample (Table 4.12), showed that all items for the Self-efficacy and Hope subscales loaded significantly onto the factors they intended to reflect, exceeding the critical value $|1.6649|$ substantially (PCQ1 - PCQ12). Likewise, all items in the Resilience subscale obtained statistically significant factor loadings. Interestingly, in this case item 13 obtained a statistically significant factor loading, as was the case in the White group. Therefore, the Black group was the only subsample in which item 13 was statistically insignificant. Regarding the Optimism subscale, five items obtained statistically significant factor loadings (PCQ19; PCQ21 – 24). Item 20 (one of the negatively keyed items in the subscale) marginally missed the $|1.6649|$ cut off (1.627). This was an interesting result, as only one of the negatively keyed items in the subscale obtained a non-significant loading. In contrast the Black sample obtained non-significant loadings for all three negatively keyed items (PCQ13; PCQ20; PCQ23), while for the White sample, the loadings for these items were significant, although lower in comparison to the other items. For the coloured sample, only item 20 was non-significant, although (similar to the White sample) the factor loadings for all three negatively keyed items were lower in comparison to the rest of the items.

Table 4.13*Unstandardized lambda-X matrix for the Coloured measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.605 (0.069) 8.706			
PCQ2	0.840 (0.073) 11.545			
PCQ3	0.845 (0.078) 10.782			
PCQ4	0.808 (0.073) 11.044			
PCQ5	0.581 (0.060) 9.618			
PCQ6	0.756 (0.070) 10.775			
PCQ7		0.521 (0.066) 7.874		
PCQ8		0.718 (0.074) 9.679		
PCQ9		0.504 (0.068) 7.442		
PCQ10		0.848 (0.069) 12.214		
PCQ11		0.737 (0.056) 13.261		
PCQ12		0.766 (0.076) 10.117		
PCQ13			0.293 (0.103) 2.860	
PCQ14			0.586 (0.057) 10.351	
PCQ15			0.450 (0.081) 5.564	
PCQ16			0.606 (0.081) 7.521	
PCQ17			0.701 (0.076) 9.219	
PCQ18			0.637	

	(0.084)	
	7.609	
PCQ19		0.669 (0.078) 8.588
PCQ20		0.167 (0.103) 1.627
PCQ21		0.771 (0.063) 12.163
PCQ22		0.835 (0.081) 10.266
PCQ23		0.471 (0.098) 4.798
PCQ24		0.528 (0.079) 6.689

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc). Bold values indicate statistically significant factor loadings.

Review of the completely standardised lambda-X matrix for the Coloured sample (Table 4.13) showed a positive finding regarding the Self-efficacy subscale, as five out of the six items either exceeded or approximated the stringent 0.70 cut-off value. The remaining item (PCQ5) was above 0.50 (.599), which is still considered a satisfactory outcome. For the Hope subscale, it is evident that four out of the six items exceeded or approximated the 0.70 value in this case, ranging from 0.659 (PCQ8) to 0.817 (PCQ11). The remaining items were still above 0.50, indicating satisfactory factor loadings. In terms of the Resilience subscale, a satisfactory result was obtained for five out of the six items in the subscale, where three items ranged around the 0.70 cut-off value (PCQ14; PCQ17 – 18), and two other items obtained factor loadings above 0.50 (PCQ15 – 16). The remaining item (PCQ13), obtained a lower factor loading of 0.211, which is in line with the results obtained in the other groups. Lastly, the factor loading results for the Optimism subscale ranged from 0.128 to 0.812. Two of the items obtained satisfactory results exceeding the 0.70 criterion (PCQ21; PCQ22), while items 19 and 24 either exceeded or approximated 0.50. However, items 20 and 23 again obtained loadings out of synch with the rest of the subscale, at 0.128 and 0.332 respectively, which was also in line with the results of the other groups.

Table 4.14*Completely standardised lambda-X matrix for the Coloured measurement model*

	SELF	HOPE	RES	OPT
PCQ1	0.699			
PCQ2	0.817			
PCQ3	0.729			
PCQ4	0.859			
PCQ5	0.599			
PCQ6	0.735			
PCQ7		0.543		
PCQ8		0.659		
PCQ9		0.570		
PCQ10		0.786		
PCQ11		0.817		
PCQ12		0.693		
PCQ13			0.211	
PCQ14			0.695	
PCQ15			0.500	
PCQ16			0.587	
PCQ17			0.787	
PCQ18			0.710	
PCQ19				0.628
PCQ20				0.128
PCQ21				0.812
PCQ22				0.758
PCQ23				0.332
PCQ24				0.474

Note: PCQ1= Psychological Capital Questionnaire 1 (i.e. Item 1, etc).

While the overall results of the item analyses for the three groups was reasonably positive, a trend was evident as items 13, 20 and 23 consistently obtained lower factor loadings in comparison to the rest of the items, obtaining non-significant loadings in many cases. This trend coincides with the PsyCap literature (e.g. Avey et al., 2010a; Dawkins et al., 2013) which has suggested that these negatively keyed items negatively impact the performance of the subscales.

4.4 Evaluating the PCQ-24 Multigroup MI and ME

Generally, inconsistency exists in the MI / ME literature regarding procedures for evaluating MI and ME, as authors differ in terms of the sequence or order of tests (Bagozzi & Edwards, 1998; Marsh, 1994), as well as in terminology used to describe the various levels of MI, which makes comparison of approaches challenging (Vandenberg & Lance, 2000). In response to this, Vandenberg and Lance (2000) conducted a review of MI literature to consolidate the various views and present an integrated process for testing MI. According to the authors, there was general consensus in literature that the omnibus test of the equality of covariance matrices is the necessary first step in MI testing (Alwin & Jackson, 1981; Bagozzi & Edwards, 1998; Byrne et al., 1989; Cole & Maxwell, 1985; Horn & McArdle, 1992; Jöreskog, 1971; Rock et al., 1978; Schaie & Hertzog, 1985; Steenkamp & Baumgartner, 1998).

Dunbar et al. (2011) however, questioned the usefulness of the omnibus test, explaining that it has been shown to indicate full equivalence in cases where further MI tests have indicated a lack of equivalence at the metric, scalar and conditional probability levels. Consequently, confidence in the findings of the omnibus test is weakened (Byrne, 1998; Dunbar et al., 2011; Meade and Lautenschlager, 2004). Dunbar et al. (2011, p.9) concluded by stating, “If the verdict of the omnibus test cannot be trusted and subsequent tests of specific hypotheses regarding equivalence in the factor loadings, intercepts and error variances are required irrespective of the nature of the verdict, there is little point in performing the test as an initial screening to determine whether further analyses are required”. In line with the argument presented by Dunbar et al. (2011), the omnibus test of the equality of covariance matrices was not performed in the present study. The results of the sequential series of MI and ME tests based on the taxonomy of Dunbar et al. (2011) discussed in section 2.3.6 is presented next. Table 4.15 provides a summary of the findings¹⁵.

Table 4.15

Summary of the multigroup measurement models’ goodness of fit statistics

	Configural invariance	Weak invariance	Strong invariance	Partial strong invariance	Strict invariance
Hypotheses Tested	H _a	H ₀₁	H ₀₂	H ₀₂₁₃ ¹⁶	H ₀₃
Degrees of Freedom	738	786	834	821	840
RMSEA	0.053	0.051	0.063	0.054	0.053
90 Percent Confidence Interval for RMSEA	0.048; 0.058	0.046; 0.056	0.058; 0.063	0.049; 0.059	0.048; 0.058
P-Value for Test of Close Fit (RMSEA < 0.05)	0.137	0.267	0.000	0.055	0.111
Normal Theory Weighted Least Squares Chi-Square	1793.69	1881.32	2264.01	1998.86	2141.33
P-Value for Test of Exact Fit (RMSEA = 0)	0.0	0.0	0.0	0.0	0.0
Population Discrepancy Function Value (F0)	0.700	0.704	1.107	0.816	0.802
Non-Normed Fit Index (NNFI)	0.972	0.974	0.961	0.971	0.9725
Comparative Fit Index (CFI)	0.975	0.975	0.961	0.971	0.9721
Standardised Root Mean Square Residual (SRMR)	0.066	0.083	0.083	0.083	0.0812

¹⁵ The complete fit statistics for all levels of the MI analyses are provided in Appendices A – E.

¹⁶ The numbering of the hypotheses for the partial invariance and partial equivalence analyses is in accordance with the iterative process which was followed in testing the respective level of MI and ME. A new measurement model was tested each time a parameter was freed to be estimated, therefore in this case the partial strong invariance model refers to H0213, i.e. The model was based on model H02 (the strong invariance model) and 13 intercepts needed to be freed to be estimated before close fit was achieved; H0213.

4.4.1 Configural Invariance

The first analysis in the sequence of MI tests, is the least stringent analysis. It examines whether a model with its factor structure constrained to be equal across the groups (while its factor loadings, intercepts and measurement error variances are freed to be estimated) fits the data to a satisfactory level (Dunbar et al., 2011). This is termed the configural invariance model, also known as the test of weak factorial invariance (Vandenberg & Lance, 2000). Obtaining at least close fit for configural invariance would indicate that the PCQ lacks construct bias, supporting the notion that the same construct is being measured successfully by the PCQ across the three ethnic groups (Dunbar, et al., 2011; Vandenberg & Lance, 2000; Meade et al., 2008). If the null hypothesis was rejected and support for configural invariance was not found, further analyses of MI and ME would not be necessary or justified. Assessing whether the factor loadings are invariant across groups for example, would be meaningless if it had been demonstrated that the groups differ in terms of the underlying construct being measured (Vandenberg & Lance, 2000). A finding of configural invariance therefore enables one to continue with the subsequent levels of MI and ME analyses.

When evaluating the fit of the invariance models, the ideal would be to obtain perfect fit, i.e. an RMSEA value of 0.0. In reality however, this is not likely, and a more tenable position is that of obtaining at least close fit. This is evaluated by inspecting the p value for the close fit hypothesis, which should be ≥ 0.05 , together with the RMSEA, which should be ≤ 0.05 to conclude good fit, while values of < 0.06 to < 0.08 are still considered reasonable. Furthermore, the NNFI, CFI and SRMR will also be considered to find support for model fit. NNFI and CFI values of ≥ 0.95 are interpreted as good while values of ≥ 0.97 are considered indicative of very good fit, while SRMR values ≤ 0.08 are considered indicative of good fit and values < 0.1 are considered acceptable. These GOF statistics will be used as the benchmark throughout the MI analyses to assess the fit of the MI measurement models.

The results for the configural invariance model (see Table 4.14), revealed that close fit was evident (p-value for the test of close fit obtained was > 0.05 ; 0.137). This was supported by the RMSEA (0.053) which met the ≤ 0.05 cut-off value for good fit. Considering the other GOF indices, the NNFI (0.972) and CFI (0.975) exceeded the 0.97 cut-off for very good fit. Lastly a satisfactory result was obtained on the SRMR (0.066), also indicating good model fit. Taken together, these results indicated that the configural invariance (baseline) model obtained good fit. Obtaining sufficient evidence to support configural invariance, as was the case here, may be interpreted to indicate that the PCQ sub-dimensions are measured in the same way over the Black, White and Coloured ethnic groups and hence the position that the PCQ-24 lacks construct bias was permissible (Dunbar et al., 2011; Vandenberg & Lance, 2000; Meade et al., 2008).

According to the Dunbar et al. (2011) methodology at this stage, it was concluded that configural invariance was achieved and hence the PCQ-24 was free of construct bias across the groups. It was noted however in the single-group measurement model analyses, that the Black measurement model results returned significant, but negative factor loadings in the Resilience subscale (PCQ14 – 18). This indicated that Resilience may manifest differently for the Black group, i.e. the descriptors of Resilience may not be regarded as examples of Resilience for the Black sample. It is interesting therefore that the configural invariance analysis presented a positive result in favour of a lack of construct bias, when construct bias may in fact be present.

Inspection of the completely standardised factor loadings for the configural invariance model¹⁷ confirmed this assumption, as negative factor loadings were observed across the Resilience subscale for the Black group (PCQ14 = -.054; PCQ15 = -0.52; PCQ16 = -0.48; PCQ17 = -0.53; PCQ18 = -0.47). Furthermore, the negatively keyed items in the subscale all produced non-significant factor loadings (PCQ13 = 0.04 – Resilience subscale; PCQ20 = -0.12 and PCQ23 = 0.03 – Optimism subscale). In the case of the configural invariance model, when interpreting the standardised solution, negative signs are not interpreted; instead, this indicates that the items are statistically insignificant. Such a finding would indicate that the validity of the Resilience scale is highly questionable for Black participants. While the negatively keyed items are well known for being problematic, the finding regarding the Resilience subscale is noteworthy as it has not been shown in previous PsyCap research.

This finding is important as it would suggest that considering the fit of the multigroup measurement model as evidence of achieving configural invariance in isolation is insufficient to confirm the absence of construct bias across the groups. This is identified as a limitation of the Dunbar et al. (2011) methodology, which will be discussed further in Chapter 5 along with recommendations for future research. Due to this finding, it would also be expected that the non-significant factor loadings (for the Resilience subscale as well as the negatively keyed items in general) will result in further bias in the subsequent analyses, which will be commented on in the following section. Nonetheless, achieving configural invariance according to the Dunbar et al. (2011) methodology allowed for examination of the next level in the sequential analyses, namely weak invariance.

¹⁷ The path diagrams for the Black, White and Coloured group Configural invariance models are provided in Appendices F – H.

4.4.2 Weak Invariance

Weak invariance, also known as metric invariance or strong factorial invariance (Vandenberg & Lance, 2000), examines whether the factor loadings of the items on the PCQ latent variables lack non-uniform bias (Dunbar et al., 2011; Wu et al., 2007). Furthermore, Wu et al. (2007, p.8) explain,

“Weak invariance postulates that, for all items, one unit change in the item score is scaled to an equal unit change in the factor score across groups. Often, a substantive researcher’s interest is to compare or explain the variation of a construct due to group membership. For such cross-group study to be meaningful, the scale (unit of measurement) of the latent variable should be identical across groups so that the variances derived are on the same metric regardless of group membership. Variance obtained from different units of measurement is not explainable or comparable. Lack of weak invariance is problematic because the test items are calibrated to the factor scores with different units of measurement across groups. If one-unit change in the item score does not result in equal unit change in the factor score across groups, the regression lines are not identical because the slopes are unequal; hence the regression lines are not identical for the groups.”

Weak invariance is investigated by constraining the slope of the regression of the indicators on the latent variables (as well as the model structure) to be equal across the Black, White and Coloured ethnic groups, while all the other variables are freed to be estimated (Cheung & Rensvold, 2002; Dunbar et al., 2011; Vandenberg & Lance, 2000; Wu et al., 2007). Achieving close fit in the sample would demonstrate that the factor loadings of the PCQ are invariant, stated otherwise, the Black, White and Coloured groups perceived and responded to the items in a similar way (Byrne & Watkins, 2003).

The multigroup weak invariance measurement model was fitted on the sample, assessing the weak invariance null hypothesis H_{01} : $RMSEA \leq 0.05$. The fit statistics are provided in Table 4.14. The results revealed that the weak MI model obtained close fit ($p > 0.05$) as $p = 0.267$. This was supported by the $RMSEA = 0.051$, which indicated good model fit (< 0.06 ; Browne & Cudeck, 1993; Schermelleh-Engel et al., 2003; Loehlin & Beaujean, 2017). Both the NNFI (0.974) and CFI (0.975) exceeded the 0.97 cut off value for good model fit (Schermelleh-Engel et al., 2003). Lastly, the SRMR obtained for the model was 0.083. In line with the recommendation by Tabachnick and Fidell (2013), values of ≤ 0.08 are desired, hence the result in this case is acceptable. Taken together, these results indicate that the multigroup weak MI model demonstrated good fit. Consequently, the assumption that the PCQ measurement model lacks non-uniform bias was permissible in line with the Dunbar et al. (2011) methodology. This indicated that the three groups perceived and interpreted the items in a similar way (Byrne & Watkins, 2003).

This result is perplexing however, considering the negative factor loadings which were observed in the single-group measurement model analyses for the Black sample, as well as the insignificant factor loadings identified in the configural invariance analysis for the Resilience subscale. In the weak invariance analysis however, statistically significant, positive factor loadings were obtained. One would have expected the negative factor loadings to show up in the weak invariance analysis as non-uniform bias according to the Dunbar et al. (2011) methodology. This however was not the case; in fact, no factor loadings were flagged as differing significantly across the groups in the Resilience subscale; providing support of a lack of non-uniform bias in the PCQ across the groups. Despite this, in accordance with the MI / ME taxonomy, it is argued that the position that the PCQ measurement model lacks non-uniform bias would be strengthened if empirical evidence of metric equivalence was shown. That is, that the multi-group measurement model with its model structure and factor loadings constrained to be equal across the three groups, did not fit the data significantly poorer than the baseline model (Dunbar et al., 2011; Wu et al., 2007). Consequently, metric equivalence was evaluated.

4.4.3 Metric Equivalence

Equivalence will be evaluated by assessing the statistical and practical significance of the difference in fit between the respective MI model and the baseline model (Cheung & Rensvold, 2002; French & Finch, 2006; Vandenberg & Lance, 2000). The statistical significance of the difference in fit can be established by using the χ^2 difference test. However, due to sensitivity questions about this test, the Cheung and Rensvold (2002) recommendation, to base the decision on whether ME has been achieved on the practical significance of the difference in fit between the two nested models, was applied. Therefore, although the results for the statistical significance test between nested models will be presented, the criteria for practical significance as stated in Table 3.2 will be used throughout the analyses to judge whether ME has been achieved. Consequently, if the change in the Mc, Gamma Hat and CFI are greater than the critical values, it will be concluded that the respective ME model fits practically significantly poorer than the baseline model (Cheung & Rensvold, 2002). Table 4.16 below displays the indices of statistical and practical significance for the metric equivalence analyses.

Table 4.16
Statistical and Practical significance – Analysis of Metric Equivalence

Hypothesis	Statistical significance			Practical significance			
	Satorra-Bentler chi square	Normal theory chi-square	Df	Prob normal theory chi- square diff	CFI	Gamma Hat	Mc
Weak invariance model	1302.216	1881.325	786		0.975	0.980	0.705
Configural invariance model	1251.344	1793.696	738		0.975	0.980	0.703
Diff(H_{01} - H_a) Metric equivalence	50.872	87.628	48	0.000	-0.000	-0.000	-0.001

In this case, the probability of observing the normal theory chi square difference was 0.000, and hence metric equivalence was not achieved using statistical significance as the benchmark. From a practical significance perspective however, it is evident that the cut-off values for equivalence were met for the CFI (-0.000), Gamma Hat (-0.000) and Mc (-0.001). Therefore, evidence was present in favour of metric equivalence, as the shift from the MI1 model to the weak MI model did not result in a substantial decline in model fit (Cheung & Rensvold, 2002). In this level of the analyses, it is investigated whether the factor loadings are identical across groups by assessing the equality in item-factor scaling, in addition to the configural model constraints (Wu et al., 2007, p.8). A lack of invariance at this level would indicate that test items are calibrated to factor scores with different units of measurement between the three ethnic groups. One unit of change in the item would therefore not equate to the same unit of change in the factor, resulting in the slopes of the regression across the three groups being unequal (Wu et al., 2007). According to the Dunbar et al. (2011) methodology, obtaining metric equivalence therefore provided strong evidence that the PCQ measurement model lacked non-uniform bias, even using the more stringent equivalence test (Dunbar et al., 2011; Wu et al., 2007). As a result, it was concluded that the assumption that the Black, White and Coloured groups interpreted the item content in a similar way was still permissible (Wu et al, 2007).

In reflecting on these results up to this point; it was shown that the factor loadings of the weak invariance model (H_{01}) were statistically significant, and positive. Furthermore, the factor loadings (for the Resilience subscale) in the configural invariance model were non-significant. The difference then, between the two models (i.e. the metric equivalence analysis) was non-significant, indicating that the difference in fit was not practically significant (Cheung & Rensvold, 2002). This provided strong evidence that the measure was free from non-uniform bias across the Black, White and Coloured

groups. This result is worrying however, considering the previous findings regarding nonsignificant factor loadings for the Resilience subscale in the Black group, indicating that the validity of this subscale is highly questionable for Black participants. It would have been expected at this stage, that the weak invariance and metric equivalence tests identified bias at this level, however non-uniform bias was not shown at either level of the analyses. However, an identified limitation to this procedure is that the procedure does not test the significance of the difference in factor loadings directly; instead it tests the significance of the fit of the two nested models. Statistically speaking there is a difference in fit, but the difference may not be practically significant in terms of the Cheung and Rensvold (2002) criteria. This point will be discussed further in the limitations section of Chapter 5.

4.4.4 Strong Invariance

Finding support for metric equivalence according to the Dunbar et al. (2011) taxonomy allowed for the analysis of the multigroup strong invariance model, which examines whether the intercepts of the regression of the items, over the different groups, on the latent variables, lack invariance (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000; Wu et al., 2007). Strong invariance is examined by constraining the intercepts of the regression of the indicators on the latent variables (as well as the model structure and the slopes) to be equal across the groups, while the other parameters are freed to be estimated. Obtaining close fit in the sample would show that the that the intercepts of the regression of the indicators on the latent variable were invariant across the three groups. This would indicate that no items of the PCQ suffer from uniform bias, providing further support that the three groups interpreted the item content in a similar way (Wu et al., 2007).

The multigroup strong invariance measurement model was fitted to the sample to test the strong invariance null hypothesis H02: $RMSEA \leq 0.05$. From the results in Table 4.14 it was evident that the strong invariance measurement model did not obtain close fit, as the p-value obtained for the test of close fit was not > 0.05 (0.0). However, after inspection of the basket of fit indices, it was concluded that although no evidence of statistical close fit was obtained, the overall conclusion about the rest of the GOFs pointed towards an evaluation of good – reasonable fit. That is, the RMSEA value of 0.0631 supported this conclusion. In addition, both the NNFI (0.961) and CFI (0.961) exceeded the 0.95 criterion of good fit, while slightly missing the 0.97 criterion for very good fit. A similar inference was provided by the SRMR, as the value of 0.083 indicated good model fit (≤ 0.08) (Hu and Bentler, 1995). Therefore, although the strong invariance measurement model did not obtain close fit from a statistical significance perspective, all the other GOF indices discussed above point towards reasonable to good model fit. A clear discrepancy is therefore evident between the strong invariance

model and the previous multigroup measurement models, with the strong invariance model fitting poorer than the previous MI models.

This result suggested that some of the intercepts of the regression of the indicators on the latent variables lack invariance, hence particular items of the PCQ may suffer from uniform bias (Dunbar et al., 2011; Wu et al., 2007). These findings cast some doubt on the assumption that the respondents from the Black, White and Coloured groups perceived the items in a similar manner, which necessitated the identification of items with the largest differences. Hence, the process to test for partial measurement invariance was initiated. Partial MI involves placing less stringent conditions on a measurement model in order to enable meaningful comparisons across the groups (Byrne et al., 1989; Vandenberg & Lance, 2000).

4.4.5 Partial Strong Invariance

Evaluation of partial MI and ME allows the researcher to identify the source of variance which is causing the misfit between the multigroup measurement model and the data (Vandenberg & Lance, 2000). This process involves freeing invariant parameter estimates until close fit is obtained and the difference in fit between the MI1 (baseline model) and the respective partially constrained model is no longer (statistically or practically) significant (Cheung & Rensvold, 1999). The process¹⁸ is iterative in nature. That is, model parameters are estimated each time when one parameter estimate is freed per analysis, and significance tests are repeated for each iterative model, until close fit was obtained.

Table 4.17 below outlines the series of analyses that were conducted in assessing partial strong invariance. Thirteen intercepts needed to be freed in subsequent iterative analyses until the model obtained close fit. Eleven out of the thirteen intercepts that were freely estimated were from the Black group, while the remaining two were from the Coloured group (which also happened to be two of the negatively keyed items in the scale). This result revealed that eleven out of twenty-four, or approximately 50% of the items in the Black differed significantly from the White and Coloured groups, indicating a clear lack of invariance at this level of analysis. From the thirteen intercepts that needed to be freed, the items were mainly from two subscales, namely the Hope (PCQ7 – PCQ9; PCQ11 – 12)

¹⁸ To evaluate the partial strong invariance measurement model, the unstandardized tau estimates for each group were entered into an Excel macro from the configural invariance output. The macro then calculated the differences between the group tau estimates, and these absolute differences were copied into a second sheet. The second sheet of the Excel macro then rank ordered the absolute differences based on the largest difference. The tau/ intercept with the highest rank order (and hence the highest lack of invariance) was freed in the first analysis, and parameter estimates calculated again. This process was continued, i.e. freeing the second highest rank order intercept, if the results of the first model did not provide support for invariance. This process continued until evidence of invariance was evident.

and Optimism (PCQ19 – 21; PCQ23 – 24) subscales. The remaining item intercept which was freely estimated was from the Self-efficacy subscale (PCQ6).

Close fit was demonstrated for the partial strong invariance model as the p-value for the close fit null hypothesis was > 0.05 (0.05590). The RMSEA value further supported this finding, indicating good model fit (≤ 0.05 ; 0.054). Inspection of the additional GOF indices, namely the NNFI, CFI and SRMR also provided results in favour of the fit of the model. The NNFI (0.971), and CFI (0.971) indices exceeded the 0.97 threshold, indicating very good model fit, while the SRMR (0.838) was ≤ 0.08 , indicating the model fitted the data to an acceptable degree. Taken together, the basket of fit indices point toward reasonable to good model fit, overall. These results illustrate that the PCQ measurement model obtained partial strong invariance, indicating that some of the intercepts of the regression of the indicators on the latent variables lack invariance. Evidently, nearly 50% that had to be freely estimated were from the Black group; therefore it is clear that certain items of the PCQ suffer from uniform bias (Dunbar et al., 2011; Wu et al., 2007).

The strong invariance analysis hypothesises that the intercepts (as well as the model structure and factor loadings) across the groups are equal. If strong invariance is not achieved, it indicates that the intercepts for the respective groups are not equal in terms of their location (Wu et al., 2007). Unequal calibration of this nature results in uniform bias against one group and in this case, it is evident that nearly 50% of the items in the Black group showed uniform bias (Wu et al., 2007; Van de Vijver & Poortinga, 1997; Byrne & Watkins, 2003). Furthermore, it is noted that items 20 “If something can go wrong for me workwise, it will” and 23 “In this job, things never work out the way I want them to” showed uniform bias for both the Black and Coloured groups. As discussed earlier, these are the negatively keyed items in the scale which were shown to have lower (or insignificant) factor loadings in the item analyses, as well as lower reliability coefficients compared to the other items. Despite this undesirable outcome, obtaining partial strong invariance allowed the analyses to continue, by evaluating partial scalar equivalence.

Table 4.17*Summary of Measurement Model GOF Indices - Partial Strong Invariance Analyses*

Model tested	Group	Df	RMSEA	P (Close)	X2	P (Exact)	NNFI	CFI	SRMR
Fully Constrained	-	834	0.063	0.000	2264.013	0.0	0.961	0.961	0.083
PC Free tau – item 20	Black	833	0.060	0.000	2176.494	0.0	0.965	0.964	0.083
PC Free tau – item 23	Black	832	0.059	0.001	2123.273	0.0	0.967	0.966	0.081
PC Free tau – item 8	Black	831	0.058	0.001	2117.434	0.0	0.967	0.967	0.081
PC Free tau – item 21	Black	830	0.057	0.005	2084.317	0.0	0.968	0.968	0.081
PC Free tau – item 20	Coloured	829	0.057	0.006	2078.571	0.0	0.968	0.968	0.081
PC Free tau – item 23	Coloured	828	0.057	0.008	2069.229	0.0	0.968	0.968	0.081
PC Free tau – item 24	Black	827	0.056	0.014	2051.252	0.0	0.969	0.969	0.081
PC Free tau – item 11	Black	826	0.056	0.022	2035.101	0.0	0.970	0.970	0.081
PC Free tau – item 6	Black	825	0.055	0.035	2018.116	0.0	0.970	0.970	0.082
PC Free tau – item 12	Black	824	0.055	0.040	2013.388	0.0	0.970	0.971	0.082
PC Free tau – item 7	Black	823	0.055	0.044	2009.129	0.0	0.971	0.971	0.082
PC Free tau – item 9	Black	822	0.055	0.048	2005.081	0.0	0.971	0.971	0.083
PC Free tau – item 19 (Final)	Black	821	0.055	0.056	1998.864	0.0	0.971	0.971	0.083

Note: PC= Partially Constrained

4.4.6 Partial Scalar Equivalence

Testing the partial scalar equivalence measurement model examines whether the multigroup partial strong MI model fits the data significantly poorer than the baseline MI model (Dunbar et al., 2011). From the table below, it is evident that the chi square difference value obtained was ≤ 0.05 , and hence the null hypothesis of no difference in fit was rejected. Consequently, partial scalar equivalence was not obtained from a statistical significance perspective, which indicated that some difference in fit was apparent between the two multigroup measurement models in the parameter.

Table 4.18*Statistical significance of the scaled chi-square difference statistic: A test of partial scalar equivalence*

Hypothesis	Satorra-bentler chi square	Normal theory chi-square	Df	Prob normal theory chi-square diff
H ₀₂ Strong invariance model	1645.453	2264.013	834	
H _a Configural invariance model	1251.344	1793.696	738	
Diff(H ₀₁ -H _a) Metric equivalence	50.872	87.628	48	0.000
Diff(H ₀₂ -H _a) Scalar equivalence	394.108	470.316	96	1.02231E-50

To assess the practical significance of the difference in fit, a further iterative process was followed in which it transpired that an additional 29 intercepts had to be freely estimated in subsequent analyses. In order to establish which intercepts needed to be freed, the differences in tau (intercepts) were calculated between the Black, White and Coloured groups. These differences were then rank ordered, and the tau associated with the largest group difference was freed first. The process was continued until the partially constrained measurement model did not fit the multi-group data practically significantly worse than the baseline model. The CFI, Gamma Hat and Mc indices for the respective analyses are shown in Table 4.18 below. Partial scalar equivalence was demonstrated in the final model (H₀₂₄₂) where the cut-off values for equivalence were met; that is CFI (-0.001) Gamma Hat (-0.001) and Mc (-0.011).

It should be noted that although partial scalar equivalence was shown from a practical significance perspective; this was not a satisfactory finding as 42 out of the 48 total constrained intercepts needed to be freed to obtain this outcome. Inspection of the results for the partial scalar equivalence analysis showed that of the additional 29 intercepts which were freely estimated, 18 were from the Coloured group and 11 from the Black group. The effected intercepts were relatively evenly distributed across the four subscales of the PCQ, affecting the Self-efficacy (PCQ1- 3; PCQ5- 6), Hope (PCQ8; PCQ10- 12), Resilience (PCQ13; PCQ15- 18) and Optimism subscales (PCQ19; PCQ21- 22; PCQ24) for the Coloured group. In terms of the Black group, the Self-efficacy (PCQ1- 3; PCQ5), Hope (PCQ10), Resilience (PCQ14- 18) and Optimism subscales (PCQ22) were also affected.

Considering the previous intercepts freed in the partial strong invariance analysis (Hope: PCQ7 – PCQ9; PCQ11 – 12, Optimism: PCQ19 – 21; PCQ23 – 24 and Self-efficacy: PCQ6; for the Black group and Optimism: PCQ20; PCQ23 for the Coloured group), it is evident that in total, 21 out of 24 intercepts were freed for the Coloured group and 22 out of 24 from the Black group. In other words, only 3 and 2 items respectively from the two groups did *not* show uniform bias, which is a rather concerning finding. Therefore, uniform bias was prevalent in almost all items of the PCQ-24 in both the Black and

Coloured groups (Dunbar et al., 2011; Van de Vijver & Poortinga, 1997; Wu et al., 2007). This indicates strong evidence that a lack of invariance is prevalent at this level of MI across the groups. As uniform bias was prevalent, it demonstrated that group membership accounted for significant variance in item responses, not accounted for by the latent variable, i.e. the intercepts of the regression of the indicators on the latent variable differed across groups (Fontaine, 2008). As a result, the stance that the three groups perceived the item content of the PCQ in a similar way was not permissible. Despite these findings, demonstrating partial strong invariance and partial scalar equivalence (albeit poorly) allowed for the evaluation of strict invariance.

Table 4.19

Practical significance of the CFI, Gamma Hat and McDonald difference statistics: a test of partial scalar equivalence

Model	Group	F0	# X	P	CFI	G1	Mc
Configural invariance H _a model		0.700	24	72	0.975	0.980	0.704
Partial strong invariance H ₀₂₁₃ model		0.732	24	72	0.974	0.980	0.693
Partial scalar equivalence [H ₀₂₁₃ - H _a]					-0.004	-0.003	-0.039
[H ₀₂₁₄ -H _a] PC: Free tau – Item 24	Coloured				-0.004	-0.003	-0.040
[H ₀₂₁₅ -H _a] PC: Free tau – Item 21	Coloured				-0.004	-0.003	-0.039
[H ₀₂₁₆ -H _a] PC: Free tau – Item 16	Black				-0.003	-0.002	-0.038
[H ₀₂₁₇ -H _a] PC: Free tau – Item 17	Black				-0.003	-0.002	-0.034
[H ₀₂₁₈ -H _a] PC: Free tau – Item 2	Black				-0.003	-0.002	-0.034
[H ₀₂₁₉ -H _a] PC: Free tau – Item 22	Black				-0.003	-0.002	-0.034
[H ₀₂₂₀ -H _a] PC: Free tau – Item 12	Coloured				-0.003	-0.002	-0.032
[H ₀₂₂₁ -H _a] PC: Free tau – Item 15	Black				-0.003	-0.002	-0.030
[H ₀₂₂₂ -H _a] PC: Free tau – Item 11	Coloured				-0.003	-0.002	-0.030
[H ₀₂₂₃ -H _a] PC: Free tau – Item 5	Coloured				-0.002	-0.001	-0.025
[H ₀₂₂₄ -H _a] PC: Free tau – Item 3	Coloured				-0.002	-0.001	-0.025
[H ₀₂₂₅ -H _a] PC: Free tau – Item 19	Coloured				-0.002	-0.001	-0.025
[H ₀₂₂₆ -H _a] PC: Free tau – Item 10	Black				-0.002	-0.001	-0.025
[H ₀₂₂₇ -H _a] PC: Free tau – Item 14	Black				-0.002	-0.001	-0.025

[H ₀₂₂₈ -H _a] PC: Free tau – Item 15	Coloured	-0.002	-0.001	-0.024
[H ₀₂₂₉ -H _a] PC: Free tau – Item 17	Coloured	-0.002	-0.001	-0.024
[H ₀₂₃₀ -H _a] PC: Free tau – Item 5	Black	-0.002	-0.001	-0.024
[H ₀₂₃₁ -H _a] PC: Free tau – Item 18	Black	-0.002	-0.001	-0.024
[H ₀₂₃₂ -H _a] PC: Free tau – Item 1	Coloured	-0.002	-0.001	-0.023
[H ₀₂₃₃ -H _a] PC: Free tau – Item 8	Coloured	-0.002	-0.001	-0.021
[H ₀₂₃₄ -H _a] PC: Free tau – Item 6	Coloured	-0.001	-0.001	-0.018
[H ₀₂₃₅ -H _a] PC: Free tau – Item 13	Coloured	-0.001	-0.001	-0.017
[H ₀₂₃₆ -H _a] PC: Free tau – Item 3	Black	-0.001	-0.001	-0.017
[H ₀₂₃₇ -H _a] PC: Free tau – Item 10	Coloured	-0.001	-0.001	-0.017
[H ₀₂₃₈ -H _a] PC: Free tau – Item 1	Black	-0.001	-0.001	-0.018
[H ₀₂₃₉ -H _a] PC: Free tau – Item 16	Coloured	-0.001	-0.001	-0.013
[H ₀₂₄₀ -H _a] PC: Free tau – Item 18	Coloured	-0.001	-0.001	-0.013
[H ₀₂₄₁ -H _a] PC: Free tau – Item 22	Coloured	-0.00140	-0.001	-0.013
[H ₀₂₄₂ -H _a] PC: Free tau – Item 2 (Final)	Coloured	-0.00120	-0.000	-0.011

Note: PC – Partially Constrained

4.4.7 Strict Invariance

According to Wu et al. (2007) strict invariance analyses have often been viewed as an unnecessary rigorous step in the MI/ME literature. These authors argue, however, that assessing strict MI is a necessary step in the process, enabling the identification of systematic error in measurement, as opposed to random error relating to instrument reliability, as is often thought in literature (Wu et al., 2007). Therefore, even though the analysis up until this point was not as favourable for the PCQ, in line with the reasoning of Wu et al. (2007) the analysis of strict invariance was conducted to assess whether the error variances associated with the indicator variables lack invariance. Obtaining close fit in the sample would indicate that no items of the PCQ suffer from error variance bias, providing support that the items operate in a similar way, regardless of the ethnic group.

To assess the strict invariance null hypothesis H_{03} $RMSEA \leq 0.05$, the multigroup strict invariance measurement model was fitted to the sample. This model was, however, based on the partial scalar equivalence model in which 42 intercepts in total were freed. The results (Table 4.16) for the strict MI model obtained close fit, as the p value was > 0.05 (0.111). The RMSEA value of 0.053 supported this finding, as the finding was ≤ 0.05 , suggesting good model fit. Favourable results were also observed

for the NNFI, CFI and SRMR. This was evident as the NNFI and CFI both exceeded the 0.97 criterion of very good fit at 0.972 and 0.972 respectively. Furthermore, the SRMR value indicated acceptable fit at 0.081 (≤ 0.08). Although the results taken together indicate that the strict invariance model obtained reasonable to good fit, it needs to be acknowledged that this is the measurement model in which 42 intercepts were freed and therefore estimated separately for the different groups, a fact which must be highlighted when interpreting the findings. That having been said, these results indicate that the strict invariance measurement model (based on the partial strong invariance model) obtained close fit, which can be interpreted as weak evidence that the error variance associated with the indicator variables are invariant (Dunbar et al., 2011; Wu et al., 2007). These findings provide some support for the assumption that no significant differences were evident in the error terms relating to the indicator variables of the PCQ-24, across the Black, White and Coloured groups. The position that the PCQ measurement model lacks error variance bias would be strengthened, however, if it was shown that the multigroup strict MI model, did not fit the data significantly poorer than the MI1 baseline model. Consequently, conditional probability equivalence was evaluated.

4.4.8 Conditional Probability Equivalence

The indices used to determine the statistical and practical significance of the difference in fit between the two nested MI models are shown in Table 4.19. The results showed a highly statistically significant difference obtained in terms of the χ^2 difference test (0.000), therefore no equivalence was shown when the difference in fit is evaluated in terms of the criterion of statistical significance. Furthermore, conditional probability equivalence was not demonstrated from a practical significance perspective as the criteria for the CFI, Gamma Hat and Mc indices were not met (Cheung & Rensvold, 2002). Finding a lack of conditional probability equivalence suggested that the argument that the PCQ lacks error variance was no longer valid, as the error variances of some of the items on the respective latent dimensions obtained practically significant differences across the three ethnic groups (Wu et al., 2007). This indicated that the items showed significant residual variance, not attributed to the underlying factor, i.e. some of the variance in the items could be attributed to 'unmodelled systematic influences' (Wu et al., 2007). As a result, the stance that the Black, White and Coloured groups perceived, interpreted and responded to the item content of the PCQ in a similar manner was not permissible (Dunbar et al., 2011; Wu et al., 2007). This necessitated the identification of the error variances which displayed the largest differences and consequently, partial conditional probability equivalence was evaluated.

Table 4.20*Statistical and Practical significance – Analysis of Conditional Probability Equivalence*

Hypothesis	Statistical significance				Practical significance		
	Satorra-Bentler Chi-square	Normal theory Chi-square	DF	Prob scaled s-b chi-square diff	CFI	Gamma Hat	Mc
H ₀₃ Strict invariance model	1428.064	2141.338	840		0.975	0.980	0.704
H _a Configural invariance model	1251.344	1793.696	738		0.972	0.978	0.669
Diff (H ₀₃ -H _a) Cond prob equivalence	176.719	347.641	102	7.82E-06	-0.003	-0.002	-0.035

4.4.9 Partial Conditional Probability Equivalence

Testing the partial conditional probability equivalence measurement model examines whether the partial strong MI model fits the data significantly poorer than the baseline model. An iterative process was followed to assess partial conditional probability equivalence. That is, error variances which showed the highest values were freed in successive analyses. The items associated with the measurement error variances which were freed in order to obtain partial conditional probability equivalence are depicted in Table 4.20. Partial conditional probability equivalence was obtained from a practical significance perspective after 9 measurement error variances were freed. This was evident as the criteria for the CFI (-0.001), Gamma Hat (-0.000) and Mc (-0.012) were met (Cheung & Rensvold, 2002). A review of the affected items indicated that 6 items were in the Black group, while 3 were from the Coloured group. Furthermore, the items were relatively evenly distributed across the four subscales. For the Black sample, the affected items were PCQ1; PCQ6 (Self-efficacy), PCQ9 (Hope), PCQ16 (Resilience) and PCQ21; PCQ23 (Optimism). For the Coloured sample, the three affected items were in the Resilience (PCQ17) and Optimism (PCQ21; PCQ23) subscales. Interestingly, only one of the negatively keyed items was found to display error variance bias across the two groups (PCQ23). Therefore, the variation which was evident in the items identified in Table 4.21, can be attributed to ‘unmodelled sources of systematic effects that influence people’s item scores’ (Wu et al., 2007, p.16). As a result, the stance that the Black, White and Coloured groups perceived, interpreted and responded to the item content of the PCQ in a similar manner was not permissible.

Table 4.21*Practical significance indices: Analysis of Partial conditional probability equivalence*

Model	Group	F0	# X	P	CFI	G1	Mc
Configural invariance H_a model		0.700	24	72	0.975	0.980	0.704
Strict invariance H_{03} model		0.802	24	72	0.972	0.978	0.669
Conditional probability equivalence [$H_{03}-H_a$]					-0.003	-0.002	-0.035
[$H_{031}-H_a$] PC: Free MEV – Item 16	Black				-0.003	-0.002	-0.029
[$H_{032}-H_a$] PC: Free MEV – Item 9	Black				-0.002	-0.001	-0.025
[$H_{033}-H_a$] PC: Free MEV – Item 23	Black				-0.002	-0.001	-0.020
[$H_{034}-H_a$] PC: Free MEV – Item 17	Coloured				-0.001	-0.001	-0.016
[$H_{035}-H_a$] PC: Free MEV – Item 21	Coloured				-0.001	-0.001	-0.016
[$H_{036}-H_a$] PC: Free MEV – Item 1	Black				-0.001	-0.001	-0.018
[$H_{037}-H_a$] PC: Free MEV – Item 6	Black				-0.001	-0.001	-0.015
[$H_{038}-H_a$] PC: Free MEV – Item 21	Black				-0.001	-0.001	-0.013
[$H_{039}-H_a$] PC: Free MEV – Item 23 (Final)	Coloured				-0.001	-0.000	-0.012

Note: PC = Partially Constrained; MEV = Measurement Error Variance

In summary, although strict invariance and partial conditional probability equivalence were shown from a practical significance perspective, obtaining these results required freeing 42 intercepts and 9 measurement error variances, mainly across both the Black and Coloured groups. That is, 22 of the 24 PCQ items demonstrated uniform bias, while 7 items displayed error variance bias across the groups. The fact that such a large percentage of the items displayed uniform bias indicates a clear lack of invariance at that the intercept level, and therefore a large part of the items did not function in the same way for the Black or Coloured groups. Otherwise stated, there is vast evidence of group membership main effect; i.e. group membership explained significant variance in item responses not explained by the latent variable (Van de Vijver & Poortinga, 1997; Wu et al., 2007). The implications thereof and recommendations for future research will be discussed in the following chapter. An important but often neglected issue in measurement invariance testing however, is the impact of biased items on the dimension level. Consequently, in line with recommendations by Van der Bank (2019), the present study attempted to fit the PCQ-24 measurement model with dimension scores, as opposed to the item level.

4.3 Assessing the PCQ-24 Measurement Model at the Dimension Level

Construct-referenced inferences on individuals' standing on the latent PsyCap dimension are derived from the dimension scores that they obtain, as opposed to the item scores directly. On the contrary, measurement bias studies utilising multigroup CFA, however, are typically performed on the item level only. Item scores from the items comprising each subscale (some of which are biased) are typically summed in accordance with a scoring key to achieve dimension scores. That leaves the question how uniform, non-uniform and error variance bias in subscale items combine to affect bias in the dimension score (Theron, 2007).

Ample literature exists in terms of how to use multigroup CFA to detect measurement bias on the item level, which was also described and illustrated in the current study. Considering the foregoing line of reasoning, this begs the question whether the same multigroup CFA methodology cannot also be applied to detect bias on the dimension score level (Van der Bank, 2019). A potential problem with such an approach, however, is that when a first-order multigroup measurement model would be fitted, each latent first-order PsyCap dimension would be represented by a single observed dimension score. Consequently, in this case the resultant multigroup measurement model would be under-identified with negative degrees of freedom (where the number of freed parameters in the model exceed the number of unique pieces of information from which to derive these parameter estimates) (Diamantopoulos & Siguaaw, 2000). Table 4.22 below depicts the number of free parameters that would have to be estimated in the case of the configural invariance multigroup first-order PsyCap measurement model. In this model, the latent first-order PsyCap dimensions are operationalised with the subscale dimension scores as indicators and the four latent first-order PsyCap dimensions are allowed to correlate. Evidently, there are only 30 unique observed variance and covariance terms and 12 unique observed means (a total of 42 unique pieces of information from which to derive estimates of the freed model parameters). Estimates for 54 unknown model parameters cannot be obtained from 42 unique pieces of known information. Consequently, the multigroup configural invariance measurement model in which all model parameters are freely estimated is under-identified¹⁹ (Diamantopoulos & Siguaaw, 2000).

¹⁹ Van der Bank (2019) also came to a similar conclusion when investigating the possibility of fitting a multigroup configural invariance model in which the 20 first-order personality dimensions were operationalised by an observed dimension score. It is noted however, that she miscalculated the number of available unique pieces of information by not taking the observed dimension score means into account. In this case, her supervisor acknowledged prime responsibility for this omission.

Table 4.22

Calculation of the number of freed measurement model parameters in the configural invariance PsyCap first-order multigroup measurement model

Parameter	Number of freed parameters
τ	12
λ	12
θ_{δ}	12
ϕ	18
Total	54

According to Van der Bank (2019), the identification problem can be solved by rather fitting a second-order measurement model, as this would reduce the number of parameters to be estimated. To do so, the multigroup PsyCap measurement model would be fitted in which the four latent first-order PsyCap dimensions are represented by the subscale dimension scores and the four latent first-order PsyCap dimensions load on a single second-order PsyCap factor. As shown in table 4.23 however, if Van der Bank's (2019) suggestion is followed (i.e. the number of freed parameters are calculated for the configural invariance multigroup second-order PsyCap measurement model), the model is still under-identified.

Table 4.23

Calculation of the number of freed measurement model parameters in the configural invariance PsyCap second-order multigroup measurement model

Parameter	Number of freed parameters
τ	12
λ	12
θ_{δ}	12
γ	12
ψ	12
Total	60

Evidently, there are still only 30 unique observed variance and covariance terms and 12 unique observed means (a total of 42 unique pieces of information from which to derive estimates of the freed model parameters). The under-identification problem may therefore have been aggravated by adding a second-order factor. Van der Bank's (2019) conclusion in this case thus can be attributed to the fact that she unfortunately failed to account for the twenty residual error variances that were introduced to the model by the introduction of the five second-order factors²⁰.

In order to allow the examination of bias on the dimension level through multigroup SEM, an over-identified configural invariance multigroup measurement model with positive degrees of freedom needs to be achieved (Diamantopoulos & Siguaw, 2000). Here, imposing equality constraints on the

²⁰ Again, her supervisor acknowledged prime responsibility for this mistake.

single and multigroup measurement models may offer a possible solution. Equality constraints on λ_{yij} , θ_{dii} and t_i can, however, only be imposed within groups, as the potential differences in these model parameters are the focus of the measurement bias study. Nonetheless, equality constraints on g_{jk} and y_{jj} in the second-order model could be imposed both within and across groups, in principle.

The attempt to achieve an over-identified model first focussed on the first-order configural invariance multigroup PsyCap measurement model. This was due to the fact that the magnitude of its negative degrees of freedom was smaller than that of the second-order model measurement model. The first attempt to achieve an over-identified multigroup configural invariance measurement model was by transforming the model from a congeneric model to an essentially tau-equivalent measurement model. This means constraining the factor loadings to be equal within each of the three groups (Graham, 2006) while still allowing them to differ across groups; which reflected the assumption that the items measure the same latent variable on the same scale. This attempt, however, was not successful as it only reduced the number of freed parameters by nine to 45. Hence, the multigroup configural invariance measurement model was still under-identified with -15 negative degrees of freedom.

Consequently, a second approach was attempted to achieve an over-identified multigroup configural invariance measurement model by transforming the model from an essentially tau-equivalent measurement model to a tau-equivalent measurement model. This process involved also constraining the intercepts of the regression of the dimension scores on the latent PsyCap dimensions to be equal within each of the three groups (Graham, 2006), while still allowing them to differ across groups. This approach (seemingly) was not possible in the Simplis command language and therefore, the Simplis (.spj) configural invariance syntax file was translated to a Lisrel (.lspj) syntax file. This further reduced the number of freed parameters by 9 to 36, thereby still leaving the multigroup configural invariance measurement model under-identified with -6 negative degrees of freedom.

The third attempt to achieve an over-identified multigroup configural invariance measurement model involved transforming the model from a tau-equivalent measurement model to a classically parallel measurement model. This approach required also constraining the residual variance of the regression of the dimension scores on the latent PsyCap dimensions to be equal within each of the three groups (Graham, 2006) while still allowing them to differ across groups. This further reduced the number of freed parameters by 9 to 27 (see Table 4.24) thereby resulting in an over-identified multigroup configural invariance measurement model with 15 positive degrees of freedom.

Table 4.24

Calculation of the number of freed measurement model parameters in the configural invariance PsyCap first-order multigroup measurement model defined as a classically parallel measurement model

Parameter	Number of freed parameters
τ	3
λ	3
θ_{δ}	3
ϕ	18
Total	27

The configural invariance syntax file was run in LISREL 8.8. Unfortunately, after 30 iterations, LISREL presented an inadmissible solution and the following warning²¹ was issued:

W_A_R_N_I_N_G: TD 1_1 may not be identified. Standard errors, t-values, modification indices, and standardized residuals cannot be computed.

A parameter is unidentified if there are more unknowns that need to be estimated than there are known pieces of information from which the estimates have to be derived (reference). Consequently, when a parameter is not identified it is not possible to obtain an estimate. In an effort to correct this problem, an attempt was made to fix the measurement error variance associated with the dimension score of the latent Self-efficacy dimension to 0.02 (using the VA command) in the Black, White and Coloured groups. In addition, the commands to freely estimate TD (1,1) and to constrain TD (1,1) to be equal to the remaining elements in the diagonal of Θ_{δ} were deleted. Unfortunately, this did not resolve the problem but resulted in a mere shift by having TD (2,2) now flagged as possibly not identified. It is acknowledged that it is not clear whether this problem is a function of the specific dataset or whether it is a necessary consequence of the nature of the classically parallel multigroup configural invariance measurement model being fitted.

Furthermore, the multigroup configural invariance measurement model returned a seemingly poor fit with both the exact and close fit null hypotheses rejected ($p < .05$). A summary of the fit statistics are provided in Table 4.25²². According to Heene et al. (2011) and McNeish and Hancock (2017), however, the conventional Hu and Bentler (1999) criteria for the evaluation of fit statistics should be used with caution when measurement quality is either extremely low (i.e. the factor loadings are low (0.04) and the error variances are high) or extremely high (i.e. the factor loadings are high (0.90) and the error variances are low).

²¹ The warning was raised with regards to the error variance parameter in all three groups.

²² It is acknowledged that due to the inadmissible solution the fit statistics should not be interpreted to render any verdict on the fit of the model. The fit statistics were presented and discussed here merely to raise awareness regarding the interpretation of fit statistics in cases of extremely high or low measurement quality.

Table 4.25

Summary of fit statistics for the classically parallel multigroup configural invariance measurement model

Fit Statistics
Degrees of Freedom = 15
Normal Theory Weighted Least Squares Chi-Square = 343.2421 ($p = 0.0$)
Satorra-Bentler Scaled Chi-Square = 324.2414 ($p = 0.0$)
Population Discrepancy Function Value (F0) = 0.4219
Root Mean Square Error of Approximation (RMSEA) = 0.2905
90 Percent Confidence Interval for RMSEA = (0.2634; 0.3184)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0000
Non-Normed Fit Index (NNFI) = 0.7154
Comparative Fit Index (CFI) = 0.7628
Standardized RMR = 0.06496

While measurement model fit may be expected to improve with enhanced measurement quality, seemingly this may not be the case. McNeish & Hancock (2017, p. 45) summarise the effect of measurement quality of a number of frequently used fit statistics as follows:

For a given set of misspecifications in a latent variable model, holding all else equal, models with poor measurement quality appear to fit much better than models with excellent measurement quality. This phenomenon was first noted in a study investigating properties of RMSEA by Saris and Satorra (1992) but has been developed further over the last few years. Hancock and Mueller (2011) coined the phrase reliability paradox to describe this relationship. The paradoxical nature of the phenomenon is evoked by the fact that researchers often strive for the highest measurement quality possible for their latent variables, but once obtained, AFIs will be far worse than if measurement quality were much poorer. Using a population study, Hancock and Mueller systematically showed how, with one hypothetical model, evaluations of data–model fit slowly deteriorate as a function of measurement quality, even when all other model and design factors are held constant. In their study, they kept the degree of misspecification, sample size, and the model identical and only changed the magnitude of the standardized factor loadings from 0.40 to 0.95. For example, in their hypothetical model, the RMSEA value with standardized loadings of 0.40 was 0.00 and the RMSEA value with standardized loadings of 0.95 was 0.10. Hancock and Mueller further showed that standard error estimates of structural parameters are much larger with poorer measurement quality and Lagrange multiplier test statistics (more commonly known as modification indexes) similarly lose their effectiveness to denote the path that should be introduced into the model to improve the fit of the model when of poorer measurement quality. Hancock and Mueller concluded that the nature of the AFI cut-offs is in direct contrast to best data analytic practice: Poor measurement quality is rewarded and good measurement quality is punished.

The fitted classically parallel multigroup configural invariance measurement model (see Table 4.25) returned a RMSEA of 0.290, a CFI of 0.762 and a SRMR of 0.064. McNeish and Hancock (2017) and Heene et al. (2011) however, cautioned against a hasty conclusion of a lack of configural invariance and construct bias on the dimension score level. Therefore, inspection of the standardised residuals would have been valuable in further exploring both McNeish and Hancock (2017) and Heene et al.

(2011) recommendations in this regard. Unfortunately, the unidentified measurement error variance parameter prevented this analysis from continuing.

The classically parallel multigroup weak, strong and strict invariance measurement models in this case are all over-identified. Therefore, (provided that the problem of the (possibly) unidentified error variance parameters is not a necessary consequence of the measurement model), the fitting of these models is in principle, possible. As a result, it should be possible to examine non-uniform, uniform and error variance bias on the dimension level. The inadmissible solution that was obtained for the classically parallel multigroup configural invariance measurement model, however, unfortunately prevented the current study from practically pursuing these possibilities. Despite this, the line of reasoning presented seems to offer the possibility of empirically examining measurement bias on the dimension score level via multigroup CFA, enabling the practitioner to evaluate the impact of item level bias on the dimension level. The implications thereof will be discussed further in Chapter 5.

4.9 Conclusion

This chapter has discussed the results of the PCQ-24 MI and ME tests in detail. The findings indicated a lack of construct bias as well as non-uniform bias, according to the Dunbar et al. (2011) methodology. It was evident however, that 22 of the PCQ items displayed uniform bias and a further 7 suffered error variance bias, which was prevalent for both the Black and Coloured groups. These findings demonstrated that although the PCQ measured the same underlying construct across the Black, White and Coloured groups, the evidence to suggest that respondents from different groups perceived and responded to the content of some of the items, differently. Consequently, the results of the PCQ MI and ME analyses displayed metric – partial scalar – partial conditional probability equivalence.

CHAPTER 5: DISCUSSION

5.1 Introduction

In South Africa, practitioners utilising psychometric assessments are required to comply with strict legal parameters, which stipulate that “psychological testing and other similar assessments are prohibited unless the test or assessment being used (a) has been scientifically shown to be valid and reliable; (b) can be applied fairly to all employees; and (c) is not biased against any employee or group” (Republic of South Africa, 1998, p. 7). Despite the increased awareness and research interest regarding the appropriate usage of psychometric assessments in a culturally diverse context such as SA since 1994 (Laher & Cockroft, 2013), the inspection of potential bias in measurement instruments is seldom reported in literature (e.g. Avci & Erdem, 2017; Barmola, 2013; Du Plessis & Barkhuizen, 2012; Rani & Chaturvendula, 2018). The consequence thereof, is that practitioners conducting comparisons between groups in terms of a particular measured construct, may report their findings as evidence of true group differences, while in fact these may be the result of bias in the measure. This could provide misleading findings which may result in unintended negative consequences on individuals and organisations (Steenkamp & Baumgartner, 1998).

The present study introduced MI testing as a rigorous manner of testing bias in measurement and attempting to meet the strict requirements of the Employment Equity Act. In particular, the study aimed to evaluate the MI of the PCQ-24, a measure which assesses the well-researched positive psychological construct, Psychological Capital (Avey et al., 2010a; Avey et al., 2009; Dawkins et al., 2013; Luthans et al., 2007b; Luthans et al., 2010). PsyCap has been shown to be positively related to various workplace outcomes. More specifically, studies have shown that PsyCap moderates the relationship between stress and work-related burnout (Görgens-Ekermans & Herbert, 2013), mediates the relationship between a supportive organisational environment and employee performance (Luthans, et al., 2008), predicted better performance and job satisfaction (Luthans et al., 2007a), and was a significant predictor of desirable and undesirable work attitudes and behaviour (Avey et al., 2010a). In the latter study, PsyCap was shown to be negatively related to organisational cynicism, intention to quit and counterproductive work behaviour, and positively related to organisational citizenship behaviours (Avey et al., 2010a). Collectively, the findings indicate that focusing on PsyCap in workplace interventions can add value beyond other widely recognised variables, including conscientiousness, core self-evaluations, person-organisation and person-job fit (Avey et al., 2010a). In addition, due to PsyCap’s malleable character, it can be developed in individuals to capitalise on its positive impact on employee outcomes (Dawkins et al., 2013; Nolzen, 2018; Wan & Hu, 2017).

Furthermore, various authors have investigated the reliability and validity of the PCQ-24 in the South African context (e.g. Bernstein & Volpe, 2016; Du Plessis & Barkhuizen, 2012; Görgens-Ekermans & Herbert, 2013; Van Wyk, 2016). Interestingly, in a study by Du Plessis and Barkhuizen (2012), the EFA results implied a three-factor structure underlying the measure and authors suggested merging the subdimensions of Hope and Self-efficacy (Confidence) to create Hopeful-Confidence (Du Plessis & Barkhuizen, 2012). A few issues regarding their study should be noted however, as many items did not load on the intended dimensions and multiple cross-loading items were also evident. In addition, the study had a small sample size (131) of predominantly white males (Du Plessis & Barkhuizen, 2012). As a result, the generalisability of the results is questionable. Different findings were obtained by Görgens-Ekermans and Herbert (2013), indicating a four-factor structure underlying the PsyCap measure. Their results were in line with previous PsyCap research (e.g. Avey et al., 2010a; Luthans et al., 2007a) and supported the construct validity of the measure on a SA sample.

Furthermore, regarding reliability, South African studies have shown that the Resilience and Optimism subscales of the PCQ-24 perform slightly poorer than the Self-efficacy and Hope subscales, which could be attributed to the negatively keyed items in the subscales (Görgens-Ekermans & Herbert, 2013; Langenhoven, 2015; Roemer & Harris, 2018). These findings were not unique to SA however, as international studies have consistently shown that the Resilience and Optimism subscales obtain lower internal consistency reliabilities than the Self-efficacy and Hope subscales (e.g. Avey et al., 2006; Luthans et al., 2008; Dawkins et al., 2013; Roberts et al., 2011). Overall, though, the studies provided preliminary evidence in support of the validity and reliability of PsyCap within the South African context.

Whether the measure is potentially biased against one or more ethnic groups in SA however, and hence could be applied fairly in decision making, is currently unknown. Consequently, the purpose of the study was to assess the MI of the PCQ-24 across Black, White and Coloured groups in SA using the taxonomy of MI and ME developed by Dunbar et al. (2011). This involved testing the significance of the difference in fit between subsequent measurement models with increasing constraints placed on the models, which enabled the identification of potential construct, uniform, non-uniform and error variance bias (Dunbar et al., 2011; Wu et al., 2007). This study therefore intended to contribute to the body of knowledge regarding the use of the PCQ-24 in the South African context in order to inform more responsible usage of the instrument in a diverse cultural and linguistic environment such as SA. This chapter will provide a summary of the results of the MI/ME testing, discuss the practical implications for practitioners and recommendations for future studies, as well as the current study's limitations.

5.2 Summary of Key Findings

The results of the CFA for the single-group measurement models showed that the null hypothesis of close fit was not rejected for all three groups ($p > 0.05$). This was supported by the RMSEA which was ≤ 0.05 for the three groups (Black = 0.055; White = 0.055; Coloured = 0.049), indicating model good fit. The standardised residuals provided further evidence in support of reasonable to good model fit, although the Black sample did obtain a higher number of large residuals in comparison to the White and Coloured samples. While this was not ideal, the number of large residuals were still acceptable overall, supporting reasonable to good model fit. Lastly, inspection of the parameter estimates for the three groups revealed a concerning finding for the Black sample. For the Resilience subscale, five out of the six items obtained significant negative factor loadings. Furthermore, the remaining item in the subscale (Item 13 – the negatively keyed item), obtained a non-significant factor loading. Whilst the three negatively keyed items in the scale obtained lower factor loadings in comparison to the other items for the White and Coloured samples, the observation regarding significant negative factor loadings in the Resilience subscale was not found in the White and Coloured samples in the present study. This finding was also not in line with previous PsyCap literature. Previous studies, both internationally and locally, have consistently shown that the negatively keyed items in the PsyCap subscale do not perform as well as the other items in the measure (e.g. Dawkins et al., 2013; Roberts et al., 2011). Therefore, the three negatively keyed items were expected to be problematic in the present study (i.e. displaying non-significant factor loadings). While the results indicated that these items were problematic throughout the three subgroups, there was potentially a more salient problem here, particularly in the item loadings in the Black sample group. That is, for the Resilience subscale, the remaining items (five out of six) all obtained significant *negative* loadings.

To understand these results it is necessary to revisit the definition of Resilience, which Luthans (2002a, p.702) defined as “the capacity to rebound or bounce back from adversity, conflict, failure or even positive events, progress and increased responsibility”. Furthermore, Masten and Reed (2002, p. 75) refer to Resilience as “a phenomenon characterised by patterns of positive adaption in the context of significant adversity or risk”. Considering the definition of Resilience, the results suggest that the items are negatively related to the underlying factor and therefore one could possibly deduce that latent variable subscale for Resilience was “flipped” for the Black sample (DiStefano et al., 2009). In other words, individuals would tend to choose the lower end of the response scale when in fact, they had a higher standing on the latent trait.

Similar findings were not noted in previous South African studies involving PsyCap (e.g. Görgens-Ekermans & Herbert, 2013). In addition, research regarding the Resilience Scale (RS) by Wagnild and

Young (1993; the seminal scale which the PsyCap Resilience Scale was developed from) did not reveal similar results in non-western samples. Studies involving the RS have been conducted on Nigerian (Abiola & Udofia, 2011; Oladipo & Idemudia, 2017), Polish (Surzykiewicz et al., 2019), Mexican²³ (Heilemann et al., 2003), Portuguese²⁴ (Pinheiro et al., 2015), Dutch²⁵ (Portzky et al., 2010), Finnish²⁶ (Losoi et al., 2013), Swedish²⁷ (Lundman et al., 2007) and Italian²⁸ (Callegari et al., 2016) samples. In each case, the authors supported the use of the scale as a valid and reliable measure of the Resilience construct.

A review of literature regarding Resilience across cultures, however, provides some insight and possible explanations for the observations in the present study. According to Van Rensburg et al. (2015), Resilience studies have been plagued as a result of inconsistent conceptualisations of the Resilience construct, as well as problematic measurement of resilience. Firstly, inconsistencies in conceptualisation are evident as Resilience was traditionally characterised as a person-centred construct (Anthony & Cohler, 1987). This implied that Resilience was an inherent strength that could be developed within the individual (Masten, 2011) – the approach also applied within the PCQ-24 conceptualisation of Resilience. Other authors (e.g. Lerner, 2006; Luthar et al., 2000), however, criticised this view of Resilience, arguing that Resilience should be conceptualised in terms of a person – context transaction, i.e. “a dynamic transaction between the environment and the individual that supports access to, and use of, resilience-promoting resources” (Van Rensburg et al., 2015, p. 2). Van Rensburg (2015) explains that the issue with person-centred conceptualisations of Resilience is that they imply an individual skill and ignore the contributions of pro-active, supportive socioecological factors toward individual’s Resilience.

Secondly, as the differing conceptualisations of Resilience inform the development of Resilience scales, inconsistencies emerged in the construction of the scales (Van Rensburg et al., 2015). It has been argued that these inconsistencies can cause potential construct, sampling and item biases (He & Van de Vijver, 2015; Ungar, 2013; Windle et al., 2011), as was evident in the present study. Therefore, despite growing awareness that Resilience most probably involves a transaction between the individual and socioecological factors, the majority of instruments that measure Resilience do not share this conceptualisation of the construct (Klika & Herrenkohl, 2013; Walsh et al., 2010; Windle et

²³ The instrument used was a Spanish translation of the original 25 item RS.

²⁴ The study was based on a Portuguese translation of the original 25 item RS.

²⁵ A Dutch translation of the 25 item RS was administered in the study.

²⁶ The study investigated the psychometric properties of the Finnish versions of the 25 item and shortened 14 item RS.

²⁷ The study investigated the underlying structure of the Swedish version of the 25 item RS.

²⁸ An Italian translation of the shortened 14-item RS (RS-14) was utilised in the study.

al., 2011). The scope of Resilience measurement is thus limited as measures focus solely on characteristics of the individual and do not address the dynamic socioecological influences (Gartland et al., 2011; Tol et al., 2013).

Considering Luthans' (2002) preceding definition of resilience as an individual's 'capacity' to bounce back from adversity, it could therefore be argued that the results observed in the present study could potentially be due to Resilience being conceptualised and measured in a more person-centred manner, rather than a person – context transaction (Van Rensburg et al., 2015). Considering the cultural differences between the groups, i.e. as the culture of Black South Africans is more collectivistic in nature in comparison to White and Coloured South Africans, their interpretation of Resilience may be more related in terms of their evaluation of socioecological influences from the group around them, which may not be the case for the other two groups. This is evident considering Theron et al.'s (2013, p. 66) explanation of African culture,

“Traditionally, African people, including the Basotho, are considered to be deeply spiritual peoples who respect traditional African culture (Byrnes, 1996; Prozesky, 2009; Renner et al., 2003). While we acknowledge that the concept of African culture or traditional African beliefs is shallow in its assumption of a one-size-fits-all culture, there are multiple scholarly publications that suggest core facets of African culture (Mkhize, 2006; Mokwena, 2007; Murove, 2009; Prozesky, 2009). These include strong spiritual (religious and ancestral), kinship, and collective beliefs and practices. Essentially, traditional African culture implies that an African way of being is an embedded or “anthropocentric” (Bujo, 2009, p. 115) way of being: Individuals are integrally part of a larger community, and it is the community that facilitates individual self-realization. African people call this collective way of being Ubuntu (Mokwena, 2007; Prozesky, 2009).”

Consequently, these cultural values could influence an individual's conceptualisation of Resilience, potentially resulting in differing item responses across the groups. As a result of this finding, it was expected that bias would be identified in the succeeding analyses in the Resilience subscale, which will be highlighted in the discussion of the MI/ME analyses.

Furthermore, across all three groups, the negatively keyed items in the scale obtained lower factor loadings in comparison to the other items in the subscale, in some cases obtaining insignificant loadings. These findings are in line with previous PsyCap literature which has consistently shown that the negatively keyed items perform poorly, impacting the Resilience and Optimism subscales (e.g. Avey et al., 2010a; Luthans et al., 2007a). These results were magnified in cases where respondents were assessed in a language other than their mother tongue (Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013). The negatively keyed items, in conjunction with the negative factor loadings

obtained in the Black sample, must be considered when interpreting the following MI analyses. Nevertheless, as the close fit null hypothesis was not rejected for the individual measurement models in all three groups, the MI testing proceeded, starting with the multigroup configural invariance analysis.

The results indicated that the close fit null hypothesis for the multigroup configural invariance model was not rejected ($p > 0.05$), hence the model showed close fit. This was supported by the RMSEA of 0.053 which demonstrated good fit in the sample. According to the Dunbar et al. (2011) methodology at this stage, demonstrating configural invariance indicated that the PCQ-24 was measuring the same underlying constructs across the Black, White and Coloured groups and hence the position that the PCQ-24 lacks construct bias was permissible (Dunbar et al., 2011; Meade et al., 2008; Vandenberg & Lance, 2000). While the individual CFAs obtained significant, negative factor loadings in the Black sample for the Resilience subscale, inspection of the factor loadings for the configural invariance model revealed statistically insignificant factor loadings for the same subscale. This is a concerning finding as it suggested that the validity of the subscale is highly questionable for Black participants.

In terms of the Dunbar et al. (2011) MI testing procedure, the absence of construct bias is confirmed by considering the significance of fit of the multigroup measurement model (in which only the factor structure is constrained to be equal across groups). The methodology, as documented in Dunbar et al. (2011), however, does not include a prerequisite to concluding configural invariance by specifically checking the significance of factor loadings as criteria to confirm the absence of construct bias at this stage of the analysis. Doing so would have raised the issue of the problematic Resilience subscale in this case – i.e., the non-significant loadings for all the items in this scale for the Black group.

A review of seminal MI literature indicates that this criterion is not incorporated as a prerequisite in testing MI (Cheung & Rensvold, 1999; Hu & Bentler, 1998; Vandenberg & Lance, 2000). Vandenberg and Lance (2000) in particular, reviewed existing literature in the aim of recommending best practices for assessing model 'goodness-of-fit' in MI testing. In explaining the role of overall model fit indices, they state:

"Overall model fit refers to evaluating the ability of the a priori model to (at least approximately) reproduce the observed covariance matrix. Overall model fit has its largest role in appraising the test for configural invariance. Recall that a good overall fit results in deciding that the factor structure (i.e., configuration) underlying a set of measures is equivalent from one group to the next. In contrast, a poor fit means that the measures were not anchored to the same configuration of latent variables in each of the comparison groups. In the former "good-fit" case, it is permissible

to continue with further invariance tests, but in the latter “poor-fit” case, further testing is not appropriate” (Vandenberg & Lance, 2000, p. 43).

The authors continue by discussing various alternative GOF indices which should be considered in conjunction with the overall model fit to conclude a level of MI has been achieved (Vandenberg & Lance, 2000). Similarly, Cheung and Rensvold (1999, p. 6) state, “If the overall fit is not adequate with respect to appropriate statistics (χ^2 , CFI, TLI, etc.), then an adequate baseline model does not exist. The results indicate that either some items load on different factors across groups, or different groups produce different numbers of factors, or both. If this situation exists, then further tests are not performed”. In the present study however, the issue was that even though good fit was achieved, inspection of the parameter estimates revealed insignificant factor loadings for an entire subscale in the Black sample. Evidently, none of the preceding criteria specify the significance of factor loadings to confirm configural invariance has been achieved specifically. It is therefore recommended that future studies include the significance of factor loadings as an additional criterion when assessing the absence of construct bias across groups. Nevertheless, obtaining close fit for the configural (baseline) model in terms of the Dunbar et al. (2011) taxonomy warranted further tests of MI and ME.

The second level of MI testing involved the evaluation of weak invariance. The results showed that the null hypothesis of close fit was not rejected for the weak invariance measurement model ($p > 0.05$). This model also obtained an RMSEA of 0.051, indicating good model fit ($RMSEA \leq 0.05$). Given the questionable results pertaining to the Black single group MM (i.e. negative loadings) and the non-significant loadings for the Black group in the configural invariance model, it was expected that the Resilience items would possibly exhibit non-uniform bias. However, it was clear from the results that no evidence of non-uniform bias emerged when analysing the results of the weak invariance test. Moreover, as the weak invariance model did not fit the multigroup data significantly poorer in comparison to the baseline model, metric equivalence was demonstrated. This provided strong evidence that the PCQ-24 lacked non-uniform bias, which showed that there was no evidence of a group membership x latent variable interaction effect (Dunbar et al., 2011; Theron, 2007; Wu et al., 2007). Otherwise stated, group was not shown to moderate the effect of the item responses on the latent variable, i.e. the slopes of the regression between the Black, White and Coloured groups did not differ significantly (Theron, 2007; Wu et al., 2007). This supported the position that the Black, White and Coloured groups perceived the item content in a similar manner.

Demonstrating metric equivalence enabled the assessment of the more stringent test, namely strong invariance. The results indicated that the null hypothesis of close fit for the strong invariance measurement model was rejected ($p < 0.05$) as the p value obtained was 0.0. Furthermore, the RMSEA

obtained was > 0.05 at 0.063, indicating only reasonable model fit (Browne & Cudeck, 1993; Loehlin & Beaujean, 2017; Schermelleh-Engel et al., 2003). As the close fit null hypothesis was rejected however, the PCQ-24 did not demonstrate strong invariance which indicated that some of the intercepts of the regression of the indicators on the latent dimensions lacked invariance. This finding casted some doubt on the assumption that the respondents from the Black, White and Coloured groups perceived the items in a similar manner, which necessitated the identification of items with the largest differences. Therefore, partial strong invariance was inspected.

A series of analyses were conducted in which thirteen intercepts were freed through an iterative process until close fit was obtained and the model showed good fit ($p > 0.05 = 0.054$; RMSEA = 0.054). The results revealed that thirteen items exhibited uniform bias. Of these thirteen item intercepts that had to be freed, 11 were in the Black group, and 2 from the Coloured group. Closer inspection of the results revealed that the affected items were mainly from two subscales, namely 5 items from the Hope subscale and 5 items from the Optimism subscales. The remaining item was from the Self-efficacy subscale. The affected items for the Black group included all the above-mentioned items, i.e. 5 from the Hope subscale, 5 from the Optimism subscale and 1 from the Self-efficacy subscale. Only two items from the Optimism subscale were affected for the Coloured group. Furthermore, it was also evident of the 5 items affected in the Optimism subscale, items 20 and 23 (two of the negatively keyed items) showed uniform bias in both the Black and Coloured groups. These findings indicated a significant group membership main effect, indicating that group membership accounted for significant variance in the 13 items not accounted for by the latent variable (Barendse et al., 2015; Fontaine, 2008). Interestingly, at this stage it is evident that of the 13 intercepts that needed to be freed, none of the items were from the Resilience subscale in the Black group. This subscale was identified in both the single-group measurement model analyses and the configural invariance analyses as problematic for the Black sample, and therefore it was perplexing that none of these items were shown to display uniform bias in this stage of the analysis.

Obtaining partial strong invariance allowed the analyses to continue, by evaluating partial scalar equivalence. Partial scalar equivalence was investigated by assessing whether the multigroup partial MI model fitted the data significantly poorer than the baseline model (Dunbar et al., 2011). Once more, an iterative process was followed. The results required a further 29 intercepts to be estimated freely in subsequent analyses; 11 of which were from the Black group and 18 from the Coloured group. Inspection of the results illustrated that the affected items were relatively evenly distributed across the four subscales. For the Coloured group, 5 items were affected from the Self-efficacy subscale (83% of items in subscale), 3 from the Hope subscale (50% of the items in subscale), 5 from the Resilience

subscale (83% of subscale items) and 4 from the Optimism subscale (considering the 2 item intercepts freed in the partial strong analysis, 100% of the Optimism items were affected in total). Furthermore, for the Black group, 4 items were affected for the Self-efficacy subscale (83% of total items affected), 1 from the Hope subscale (bringing the total to 100% of items in the subscale affected, considering the 5 intercepts freed in partial strong analysis), 5 from the Resilience subscale (83% of items in subscale affected) and 1 from the Optimism subscale (again, bringing the total affected items to 100%).

Taking the additional thirteen intercepts into account (freed during the partial strong invariance analysis), it is evident that in total, 21 out of 24 intercepts had to be freely estimated for the Coloured group, and 22 out of 24 from the Black group; indicating a clear lack of invariance at this level of MI/ME. That is, only 3 items from the Coloured and 2 items from the Black group, did *not* differ in terms of the intercepts of the regression of the indicators on the latent dimensions across the Black, White and Coloured groups. As discussed in Chapter 2, differing opinions exist in literature regarding the number of invariant items which can be freed to be estimated while claiming partial MI or ME has been achieved. For example, Steenkamp and Baumgartner (1998) claim that parameter constraints can be freed until only the reference indicator and one other item per factor remain. Vandenberg and Lance (2000) however, argue that this may result in inappropriate evaluation of mean group differences being conducted on non-comparable measures. Evidently in the present study, partial strong invariance was only obtained when 21 and 22 out of the constrained intercepts respectively, were freed in the Coloured and Black groups. Furthermore, in some cases, the intercepts of all items in the subscale needed to be freed before partial MI was achieved.

In accordance with the recommendations by Steenkamp and Baumgartner (1998), partial strong invariance would be considered achieved as in both groups, two or more invariant items remained. The concern raised by Vandenberg and Lance (2000) however needs to be considered, as when so many items are affected, the question becomes whether valid cross-group comparisons can be made regarding the construct of interest. Uniform bias was therefore very prevalent and indicated a significant group membership main effect, i.e. group membership accounted for significant variance in item responses, not accounted for by the latent variable (Barendse et al., 2015; Fontaine, 2008). Hence the stance that the three groups perceived and responded to the item content of the PCQ-24 in a similar manner was no longer permissible. Despite these concerning findings, as partial strong invariance was demonstrated according to the Dunbar et al. (2011) methodology, strict invariance could be assessed.

The strict invariance measurement model (based on the partial strong – partial scalar equivalence model) was fitted to the multigroup sample. The results showed that the null hypothesis of close fit was not rejected for the multigroup strict invariance model ($p > 0.05$; 0.111). This was also supported by the RMSEA value (0.053), indicating good model fit (Schermelleh-Engel et al., 2003). Although strict invariance was obtained, it must be noted that the measurement model was based on the preceding measurement model, in which 42 out of 48 intercepts were freed to be estimated in order to obtain close fit. Obtaining strict invariance constituted weak evidence that the error variance associated with the indicator variables were invariant. The position that the PCQ-24 measurement model lacked error variance bias would be strengthened however if it was shown that the strict MI model, did not fit the multigroup data significantly poorer in comparison to the baseline model. Thus, conditional probability equivalence (Dunbar et al., 2011) was evaluated.

Inspection of the results revealed that conditional probability equivalence was not demonstrated from a practical significance perspective, as the criteria for the CFI, Gamma Hat and Mc indices were not met (Cheung & Rensvold, 2002). Consequently, the position that the PCQ-24 lacked error variance bias was no longer valid, which necessitated the identification of the error variances which displayed the largest differences; hence, partial conditional probability equivalence was evaluated. To assess partial conditional probability equivalence, 9 measurement error variances were freed to be estimated in successive analyses until the criteria for practical significance were met (Cheung & Rensvold, 2002). A review of the affected items revealed that 6 items from the Black group and 3 items from the Coloured group showed error variance bias. In addition, it was evident that the items were relatively evenly distributed across the four subscales. In terms of the Black sample, 2 items in the Self-efficacy subscale were affected, 1 from the Hope subscale, 1 from the Resilience subscale and 2 from the Optimism subscale. For the Coloured sample however, the three affected items were in the Resilience and Optimism subscales, with 1 and 2 items affected respectively. Interestingly, in this case only one of the negatively keyed items was found to display error variance bias in both groups (PCQ23). This demonstrated that 7 items of the PCQ-24 in total, showed significant residual variance, not attributed to the underlying factor which was measured (Dunbar et al., 2011). The variation in the items could therefore be attributed to “unmodelled sources of systematic effects that influence people’s item scores” (Wu et al., 2007, p.16).

According to the Dunbar et al. (2011) methodology, overall, these findings indicated that while the PCQ-24 measured the same underlying construct across the Black, White and Coloured groups, they appeared to perceive and respond to the content of some of the items differently. Consequently, the results of the analyses showed metric – partial scalar – partial conditional probability equivalence.

5.3 Implications and Recommendations for Future Research

The present study followed the procedure advocated by Dunbar et al. (2011) and according to the application of the 'rules' of this procedure, configural invariance was evident. Inspection of the factor loadings in the configural invariance analysis, however, revealed nonsignificant loadings, indicating that the Resilience construct manifests differently for the Black sample. This highlights a potential gap in the proposed methodology in the Dunbar et al. (2011) process. In addition, previous MI literature which recommends processes for confirming MI across groups also does not include the significance of factor loadings as a prerequisite in the configural invariance analysis (Cheung & Rensvold, 1999; Vandenberg & Lance, 2000). Consequently, it is recommended that future studies include the significance of factor loadings in conjunction with obtaining good fit in the sample, to confirm the absence of construct bias across groups.

Furthermore, the significant negative factor loadings observed for the Resilience subscale in the Black sample indicated that Resilience may manifest in Black South Africans differently. Therefore, future studies utilising the PCQ-24 on Black respondents should take extra care in ensuring the validity and reliability of the Resilience subscale. In addition, future studies could repeat the current study on a bigger sample of Black participants, to ascertain whether the results are replicable. Should this be the case, this may warrant the reconceptualisation of Resilience in the PCQ-24. Doing so would ensure that the items account for the person-centred, as well as socioecological influences which play a role in individual's Resilience (Van Rensburg et al., 2015).

In terms of the weak and metric invariance/equivalence tests, what could be described as a limitation is that the procedure does not test the significance of difference in the factor loadings directly. Rather, the methodology tests the significance of the difference between two nested multigroup measurement models (Dunbar et al., 2011; Vandenberg & Lance, 2000). Therefore, while significant differences may be evident between the factor loadings of the configural invariance model and the weak invariance model, the difference in fit between the two measurement models may not be practically significant. As a result, the presence of non-uniform bias could go unnoticed due to the criteria for practical significance of fit being achieved (Cheung & Rensvold, 2002). Inclusion of the significance of factor loadings in the configural invariance analysis however, would have flagged the nonsignificant factor loadings in the Resilience scale for the Black sample, which would have warranted further inspection of the subscale in subsequent analyses.

What does the presence of non-uniform, uniform and error variance bias mean, however? Evidence of non-uniform bias indicates a group - latent variable interaction effect, i.e. group membership moderates the effect of the latent dimension on the item responses (Dunbar et al., 2011; Theron,

2007; Wu et al., 2007). Otherwise stated, the slopes of the regression of the indicators on the latent dimension are not the same for different groups. The present study utilised the methodology proposed by Dunbar et al. (2011). This procedure showed no evidence of a group x latent variable interaction effect across the Black, White and Coloured groups; hence non-uniform bias was not identified (Wu et al., 2007).

There was vast evidence however, of group main effect, i.e. group membership accounted for significant variance in item responses, not accounted for by the latent variable (indicating significant uniform bias) (Byrne & Watkins, 2003; Theron, 2007; Van de Vijver & Poortinga, 1997; Wu et al., 2007). This suggests the presence of potential response bias for the Black and Coloured samples (Bollen, 1989; Vandenberg & Lance, 2000). Some potential sources of response biases could include differential social desirability and differential response styles (Byrne & Watkins, 2003; Van de Vijver & Tanzer, 2004). According to Hofstede (1980), differences in social desirability across cultures can result in a culturally specific response set, which can lead to a shift in average test scores (Van de Vijver and Tanzer, 2004). Odendaal (2015) studied differences in social desirability scores across ethnic groups and found that Black participants (Sotho- and Nguni- speaking) displayed higher social desirability in comparison to the English and Afrikaans speaking participants. The study indicated however, that social desirability was negatively related to general reasoning and in addition, the relationship was moderated by ethnicity (Odendaal, 2015). In the present study, while the level of general reasoning ability of the sample is unknown, differences in social desirability as a result of culture is still a possibility (Middleton & Jones, 2000).

Furthermore, another potential source of uniform bias could be differential response styles, such as acquiescence or extreme response styles (Allen, 2017; Suárez-Alvarez et al., 2018; van Sonderen et al., 2013). Acquiescence could therefore potentially be a reason for the response set observed. Johnson et al. (2011) explain that there is evidence which suggests that non-White or ethnic minority respondents are more inclined toward acquiescent responding than White respondents. The authors suggest that this form of responding has been associated with a collectivistic cultural background, as well as dimensions of culture including uncertainty avoidance (Johnson et al., 2011). An extreme response style is a further potential explanation of the results, as a few cultural processes have been suggested to contribute to an extreme response style. These include, “differential cultural emphases on sincerity, moderation, modesty, willingness to be judgmental, clarity, assertiveness and decisiveness during interpersonal communications, as well as familiarity with survey instruments developed within Western scientific traditions” (Johnson et al., 2011). Any of the above-mentioned cultural factors could therefore potentially have influenced the respondents from the Black and

Coloured groups in the present study, resulting in uniform bias. Moreover, demonstrating error variance bias showed that 7 items of the PCQ-24 showed significant residual variance, not attributed to the underlying factor which was measured. Therefore, the variation which was evident in the items can be attributed to ‘unmodelled sources of systematic effects that influence people’s item scores’ (Wu et al., 2007, p.16).

The aim of the present study was to investigate the bias in the PCQ-24 as it is part of the professional obligations of I/O Psychologists to ensure that measures utilised are proactively shown to be valid, reliable, unbiased and can be fairly applied across groups (Republic of South Africa, 1998, p. 7). In the context of Employment Equity legislation, the results of the present study could be interpreted by some as concerning, as significant uniform bias was evident, as well as evidence of error variance bias. This demonstrated that group membership explained significant variance in item responses for the Black and Coloured samples, not explained by the latent variable (Dunbar et al., 2011; Wu et al., 2007).

According to Theron (2007) however, the presence of bias does not always translate to unfair discrimination. If potential bias in a measure is not known or understood by the practitioner, it could potentially (although not necessarily) disadvantage the members of a particular group, resulting in unfair discrimination (Theron, 2007). Consequently, Theron (2007) has argued that the responsibility rests on the test user, to ensure that the inferences derived from a measure are fair and do not discriminate *unfairly*, even if it is shown that the scale may be biased against one or more groups. This is an important point as psychometric assessments are utilised in selection to enable practitioners to discriminate amongst applicants based on inherent job requirements. This form of discrimination is considered fair (Theron, 2007). Unfair discrimination would therefore be evident if the measure used, results in differentiation between groups on factors other than the target variable. This would constitute unfair discrimination in terms of the act, as the EEA (Republic of South Africa, 1998, p.14) states:

“No person may unfairly discriminate, directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language and birth.”

When measures are influenced by respondents’ ethnic or social origin for example (i.e. either as a group main effect [uniform bias] or group membership x latent variable interaction effect [non-uniform bias]), it could potentially result in unfair discrimination against a group as the inferences derived from the measure could be biased.

In practice, inferences regarding an individual's standing on the latent variable or criterion, are derived from their dimension (scale) scores, not individual items. Theron (2007) has therefore argued that the practitioner therefore needs to consider how the differential item functioning (DIF) translates at the scale level. To understand this phenomenon, it has been suggested by Theron (2007) that it may be necessary to revisit the definition of item bias. Item bias, or DIF, exists when group membership accounts for unique variance in observed item scores, not accounted for by the latent dimension (as either an interaction effect, i.e. non-uniform bias; or as a main effect, i.e. uniform bias) (Millsap & Everson, 1993; Theron, 2007). Items are combined to produce an observed predictor scale or dimension score. Scale bias is thus evident when group membership accounts for unique variance in the dimension score (as an interaction effect or main effect) which is not accounted for by the latent dimension (Drasgow & Hulin, 1990; Millsap & Everson, 1993). Bias at the scale level therefore results in the regression of the predictor score on the latent dimension to differ across groups, in terms of slope and/or intercept (Theron, 2007, p. 108).

This begs the question, however, how the biased items in the PCQ-24 combined across the Black and Coloured items at the dimension/ scale level, i.e. whether the biased items accumulate is in the same direction or cancel each other out. If the biased items over-estimate (or underestimate) individuals' standing on the latent variable equally, then the bias prevalent in the measure would not be problematic in practice. If, however, the bias in the items to some degree accumulates at the dimension/ scale level, could it be practically significant and impact practical decision making? In practice, practitioners can derive two forms of inferences from an individual's scale scores, namely construct-referenced inferences and criterion-referenced inferences (Kaplan & Saccuzzo, 2009). Construct-referenced inferences relate to what the test measures, i.e. deriving inferences regarding the test taker's standing on the latent variable. This dimension score is transformed to a norm score (such as a z-score, percentile rank, McCall T-score, stanine and sten), which describe the comparative position of the respondent's dimension score in the normative distribution (Kaplan & Saccuzzo, 2009). The concern is therefore whether the biased items impact an individual's dimension score to the extent that an individual's norm score would be affected? In the case of a stanine or sten score for example, if the inclusion of the biased items would result in a shift in a sten score, it would be significant and could impact practical decision making. If the effect of the biased items was minor, however, and would not result in a shift in a norm score; then the inclusion of these items would not be as problematic as they would not result in any unfair decision making (Theron, 2007).

On the other hand, criterion-referenced inferences relate to predictions regarding a test-takers' performance on a criterion, as a result of their score on the latent variable, i.e. the prediction of

subsequent performance or outcomes derived from scale scores (Frey, 2018). Practitioners deriving criterion inferences will convert an individual's dimension score to a norm score which describes the individual's relative expected position in the criterion distribution (i.e. their expected performance/criterion score). The concern then, is whether the accumulated item bias could result in a shift of an individual's score in the criterion/ predictor distribution, either in slope or intercept (Theron, 2007). Such a shift could result in group-related predictive bias, underestimating the criterion performance of the affected group, which could create indirect unfair discrimination against the disadvantaged group (Theron, 2007).

If the bias at the dimension score level is so minor that it would not result in a significant shift in an individual's score, then the bias should not be alarming to the practitioner as it will not have a significant practical implication. Moreover, should the scale bias result in a significant shift in an individual's predictor score (causing an over- or under estimation of their performance on a criterion or shifting their score on a selection rank order), Theron (2007) has argued that this can be corrected in the selection decision rule. To this end, Theron (2007, p. 109) stated, "When criterion inferences are derived clinically from predictor scale scores containing measurement bias, unfair discrimination most likely would occur. The unfair discrimination should, however, ultimately not be blamed on the scale bias existing in the predictor but rather on the inappropriate manner in which criterion inferences are derived from the predictor scale scores". Consequently, if a group membership main effect or group membership x latent variable interaction effect are taken into account in the regression equation which dictates the selection decision rule, biased measures can still be used responsibly in practice, not resulting in unfair discrimination against one or more groups (Theron, 2007).

Another concern, therefore, is whether criterion-referenced inferences result in fair (selection) decision making. Various authors have proposed models to address the issue of selection fairness (i.e. Cleary, 1968; Einhorn & Bass, 1971; Huysamen, 1996; Huysamen, 2002). The Cleary (1968) fairness model in particular is attractive as it argues that a selection decision rule will be considered unfair when systematic prediction errors are present in a common regression equation (i.e. the regression lines vary for two or more groups in terms of intercept and/or slope), and these are not considered by the practitioner. Practitioners can therefore avoid systematic group-related prediction errors by taking group membership into account, when it is evident that bias in the dimension scores is resulting in a significant difference across the groups in terms of the regression slopes and/or intercepts (Theron, 2007).

Evidently, practitioners cannot circumvent the fact that decisions are made on the dimension level, thus it needs to be considered how bias at the item level in the present study translates to bias at the

dimension level. To assess the implications thereof, this study considered a recommendation made by Van der Bank (2019), i.e. fitting a multigroup measurement model with dimension scores, as opposed to the item-level. Van der Bank (2019) explained that ample literature exists in terms of recommendations for evaluating MI and ME at the item level. Example of these techniques which assess the statistical and practical significance of fit of the respective multigroup measurement models include the Satorra-Bentler scaled chi-square difference test and the Cheung and Rensvold (2002) criteria.

Van der Bank (2019) argued however, that a gap exists in MI/ME testing at the dimension level, and hence a need exists to apply the same approaches to assess measurement bias at the dimension level (Van der Bank, 2019). When a measurement model is fitted at the dimension level, the latent first-order factors are represented by a single dimension score as opposed to a number of items. The issue that arises then, however, is if the number of unique elements in the observed variance-covariance matrix are less than the number of freed parameters in the measurement model, i.e. if the measurement model was under-identified with negative degrees of freedom (Diamantopoulos & Siguaw, 2000; Van der Bank, 2019). If the number of unique elements in the observed variance-covariance matrix exceed the number of freed parameters in the measurement model however, the researcher would be able to fit the multigroup measurement model at the dimension level and follow the same procedure as proposed by Dunbar et al. (2011). If a lack of invariance was detected in any level of the Mi/ME analyses, the same methodology could be applied to identify biased scales/dimensions resulting in a level of MI/ME not being achieved (Van der Bank, 2019).

In line with these recommendations, the present study attempted to fit the PCQ-24 multigroup measurement model with dimension scores. As the measurement model comprised only 4 latent dimensions however, the resulting model was under-identified with negative degrees of freedom and could not be fitted with LISREL (Diamantopoulos & Siguaw, 2000). Therefore, unfortunately the present study was not able to assess the impact of the uniform and error variance bias identified in the items of the PCQ-24, at the dimension level. Future measurement invariance studies should proceed with investigating the impact of DIF at the scale level as proposed by Van der Bank (2019), and not terminate the analyses at the item level. This will enable the researcher to quantify the impact of the DIF at the dimension level, to assess whether the DIF will have practical implications in decision making. In the absence of this evidence in the present study, the item bias which was demonstrated remains highly troublesome from a psychometric workmanship perspective. Although the bias might possibly (although not necessarily) affect practical decision making; it is still concerning that so many items are not operating in the same manner across groups.

If item bias is identified in a measure, a few options are available to test developers to address the issue. Firstly, test developers could decide to remove the implicated items from the relevant subscale (Van de Vijver & Tanzer, 2004; Van der Bank, 2019). Such an approach would be unappealing in a case such as the PCQ-24 however, where the measure only has 6 items per subscale. Alternatively, the item content of the affected items could be adjusted or substituted with new items to address the bias (Van de Vijver & Tanzer, 2004). Analysis of the new item content, however, could reveal that it too suffers some form of item bias; this second recommendation therefore does not guarantee a solution to the problem (Van der Bank, 2019). The decision to assess the compounding effect of item bias at the dimension level is therefore an attractive option. This is because it enables the researcher to assess whether the item level bias has practical implications at the dimension level, and if so, the practitioner can account for the biased effects without removing the biased items from a measure.

It is evident that when the PCQ-24 is transported to the South African environment, some items do not work as well as intended. The present research could therefore contribute to making other researchers and practitioners aware of the fact that significant uniform bias appears to be evident in the scale when applied to the South African environment. From the results, it was very clear that items 13, 20 and 23 (the negatively keyed items in the scale) created problems in the analysis, starting from the item analysis, through to the MI and ME analyses. Although these were not the only items shown to be problematic in the study; in the partial strong invariance analysis, the negatively keyed items (PCQ20; PCQ23) were the first to be freely estimated in the Black and Coloured groups. As a result, when this instrument is used in South African environments, researchers should consider rewording these items so that they do not contain negatively keyed effects.

5.4 Limitations of the Study

Methodological limitations were identified in terms of the Dunbar et al. (2011) MI/ ME procedure. This was firstly due to the multigroup measurement model obtaining close fit in the configural invariance analyses when the entire Resilience subscale obtained nonsignificant loadings for the Black sample. Secondly, the same subscale did not show any evidence of non-uniform bias in the weak invariance and metric equivalence analyses. As discussed, due to these findings, it was expected that bias would have been shown in the Resilience subscale for the Black group. It is concerning therefore that no items were flagged as demonstrating non-uniform bias in both the weak invariance and the more stringent, metric equivalence analyses.

Another possible flaw in the methodology was identified in the partial MI/ME analyses. In this process, the factor loadings/intercepts or error variances with the largest differences between groups were identified and freed one at a time in subsequent analyses. This was an iterative process, and after each

measurement model parameter was freed, the fit of the respective measurement model was judged to assess if close (good) fit was achieved, or the fit of the model was not (practically) significantly poorer than the baseline model (Dunbar et al., 2011). In the current study, a potential limitation therefore is that the White group was used as the reference group in the partial MI/ME analyses; as the difference in measurement model parameters was calculated between the White and Black group, as well as the White and Coloured group. The difference between the Black and Coloured group was therefore never directly tested. At the partial strong invariance analysis for example, we inferred that if the tau of item 2 had the largest difference and needed to be freed to be estimated; a significant difference was evident in the intercept of item 2 between the Black and White groups. If we did not do so for the Coloured group, the intercept for item 2 may have differed from the Black group; but not the White group. If the tau for item 2 was freed for both the Black and the Coloured groups, one could infer that both differed significantly from the White group; but we would not know if they differed significantly from each other. It is recommended that future studies testing measurement bias across three or more groups ensure that the differences between all respective groups are tested, i.e. in this case it would mean calculating the difference between the White and Black groups, the White and Coloured groups, as well as the Black and Coloured groups.

Furthermore, due to the different administration methods used in the three studies from which the archival data was gathered, method bias could be present in the present study. According to Van de Vijver and Poortinga (1997, p. 30), “method bias occurs when a cultural factor that is not relevant to the construct studied affects most or all items of a test in a differential way across the cultures studied. More specifically, administration bias is caused by “differences in the environmental administration conditions whether physical, technical or social” (Van de Vijver & Tanzer, 2004, p. 125). Traditionally, the administration process of a measure is standardised and controlled, ensuring that the instructions, time keeping, testing conditions and scoring procedure do not vary for individuals across testing conditions (Kaplan & Saccuzzo, 2009). As the data collection procedures in the preceding studies included online as well as paper and pencil administration (Roux, 2014; Herbert, 2011; Boers, 2014), the administration procedure was not standardised across the three studies. Furthermore, Van de Vijver and Tanzer (2004) explain that differential familiarity with the measurement process could cause substantial cross-cultural differences in non-target variables, including respondents’ level of curiosity (as a result of the novelty of the situation) and willingness to self-disclosure. The method bias resulting from the differing modes of administration used by the preceding studies may therefore have impacted the validity of the results obtained from the archival data (Van de Vijver & Poortinga, 1997; Van de Vijver & Tanzer, 2004).

In terms of sample demographics, the final sample comprised of 30% Black respondents, 37% White respondents and 33% Coloured respondents. Although the three sample groups were adequately represented for the purposes of the study, the sample demographics were not aligned to the general South African population in terms of ethnic distribution. According to the 2016 Census survey, the population consists of 79% Black Africans, 9% Coloureds and 9% Whites (Statistics South Africa, 2016). Furthermore, in terms of first language, most respondents indicated Afrikaans (47%), followed by English (25%) and African languages (27% - Zulu = 7%; Sotho = 7%; Xhosa = 6%). This is not in line with the language distribution of the SA general population, where 23% of individuals indicate Zulu as their first language, followed by Xhosa (16%), Afrikaans (14%) and English (10%). Regarding the education levels of respondents, it is noted that the sample consisted of relatively well-educated, working adults. Evidently, the sample was therefore skewed in comparison to the average education level of the general SA population. The results may therefore not be fully representative of the SA population. Lastly, it is acknowledged that archival data was used in the study. As the sampling method in the three preceding studies was a convenience sampling method, the data can be regarded as a non-probability sample (Roux, 2014; Boers, 2014; Herbert, 2011). Due to the non-random nature of this sampling technique, the sample may therefore not provide a precise representation of the sampling population.

5.5 Conclusion

The present study aimed to assess the MI and ME of the PCQ-24 across Black, White and Coloured ethnic groups, to contribute to the responsible use of the measure in SA. The results revealed that the PCQ-24 demonstrated metric, partial scalar and partial conditional probability equivalence. This indicated that significant uniform bias was prevalent across groups, as well as evidence of error variance bias in terms of the Dunbar et al. (2011) methodology. While the results suggested that the PCQ-24 measured the same underlying constructs across the three groups, it was noted that the procedure may have been deficient in its ability to identify construct bias in the Resilience subscale for the Black sample. Furthermore, due to this finding, it was also expected that the analyses would have shown non-uniform bias in the factor loadings, however this was not flagged in either the weak invariance or metric equivalence analyses. Recommendations in this regard were provided to address this issue in future studies. Furthermore, it was evident that the negatively keyed items in the PCQ-24 were shown to be problematic throughout the analyses, starting from the item analyses through to the MI and ME analyses. These findings were in line with previous research on the PCQ-24 (Avey et al., 2010a; Dawkins et al., 2013; Görgens-Ekermans & Herbert, 2013; Luthans et al., 2007a). Lastly, the study contributed to MI and ME literature by providing recommendations which contribute to improving the Dunbar et al. (2011) methodology for future measurement invariance studies.

REFERENCE LIST

- Abiola, T. & Udofia, O. (2011). Psychometric assessment of the Wagnild and Young's resilience scale in Kano, Nigeria. *BMC Research Notes*, 4(509) <https://doi.org/10.1186/1756-0500-4-509>
- Abrahams, F. (1996). *The cross-cultural comparability of the Sixteen Personality Factor Inventory (16PF)*. Unpublished doctoral thesis, University of South Africa.
- Allen, M. (2017). *The sage encyclopedia of communication research methods* (Vols. 1-4). Sage Publications, Inc. doi: 10.4135/9781483381411
- Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. J. Jackson & E. F. Borgatta (Eds.), *Factor analysis and measurement in sociological research* (pp. 249-279). Sage Publications, Inc.
- Anthony, E.J., Cohler, B.J. (1987). *The invulnerable child*. Guilford Press.
- Avci, A. & Erdem, R. (2017). Investigation of the Psychological Capital of Security Employees by Nationality and Status. *European Journal of Multidisciplinary Studies*, 2(5). Pp 204-208.
- Avey, J. B., Avolio, B. J. & Luthans, F. (2011). Experimentally analyzing the impact of leader positivity on follower positivity and performance. *The Leadership Quarterly*, 22, 282–294. doi:10.1016/j.leaqua.2011.02.004
- Avey, J. B., Hughes, L. W., Norman, S. M., & Luthans, K. W. (2008). Using positivity, transformational leadership and empowerment to combat employee negativity. *Leadership & Organizational Development Journal*, 29, 110–126. doi:10.1108/01437730810852470
- Avey, J. B., Luthans, F., & Jensen, S. M. (2009). Psychological capital: A positive resource for combating employee stress and turnover. *Human Resource Management*, 48 (5), 677–693. doi:10.1002/hrm.20294
- Avey, J. B., Patera, J. L., & West, B. J. (2006). The implications of positive psychological capital on employee absenteeism. *Journal of Leadership and Organizational Studies*, 13, 42–60. doi:10.1177/10717919070130020401
- Avey, J.B., Luthans, F. & Youssef, C.M. (2010a) The Additive Value of Positive Psychological Capital in Predicting Work Attitudes and Behaviours. *Journal of Management*, 36(2). P430-452
- Avey, J.B., Luthans, F., Smith, R.M. & Palmer, N.F. (2010b) Impact of Positive Psychological Capital on Employee Well-Being Over Time. *Journal of Occupational Health Psychology*, 15(1). P.17-28.

- Babbie, E. & Mouton, J. (2001). *The practice of social research*. Oxford University Press.
- Babbie, E. (2013). *The Practice of Social Research*, Thirteenth Edition. Cengage Learning.
- Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, 1, 45-87.
- Bakker, A. B., & Oerlemans, W. G. M. (2012). Subjective well-being in organizations. In K. Cameron & G. M. Spreitzer (Eds.). *The Oxford handbook of positive organizational scholarship* (pp. 178–189). Oxford University Press.
- Bandalos, D. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9(1), 78-102.
- Bandalos, D. L., Geske, J. A. & Finney, S. J. (2001). A model of statistics performance based on achievement goal theory. *Journal of Educational Psychology*, 95(3), 604- 616.
- Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. Freeman.
- Barendse, M.T., Albers, C.J., Oort, F.J., & Timmerman, M.E. (2015). *Measurement bias detection through Bayesian factor analysis*. In R. Van de Schoot, P. Schmidt & A. De Beuckelaer (Eds.), *Measurement invariance* (pp. 23–31). Lausanne: Frontiers.
- Barmola, K.C. (2013). Gender and Psychological Capital of Adolescents. *The Indian Journal of Applied Research*, 3 (10). Pp1-3.
- Benson, J., & Brown, M. (2011). Generations at work: Are there differences and do they matter? *The International Journal of Human Resource Management*, 22(9), 1843-1865. doi:10.1080/09585192.2011.573966
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modelling. *Sociological Methods & Research*, 16(1), 78-117.
- Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588 - 606.
- Bernstein, C. & Volpe, S. (2016). The implications of sex role identity and psychological capital for organisations: A South African study. *SA Journal of Industrial Psychology*, 42(1) a1311.
- Berry, C.M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 435–463. <https://doi.org/10.1146/annurevpsych-032414-111256>

- 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.
- Beutel, N. J., & Wittig-Berman, U. (2008). Work-family conflict and work-family synergy for generation X, baby boomers, and matures. *Journal of Managerial Psychology*, 23(5), 507-523.
- Bialosiewicz, S., Murphy, K., & Berry, T. (2013). *An introduction to measurement invariance testing: Resource packet for participants* (pp. 1–370). American Evaluation Association.
- Boers, M. (2014). *Empirical Evaluation of the Steyn-Boers Structural Model of Psychological Wellbeing at Work*. Master's Thesis, Stellenbosch University.
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley
- Bonaccio, S., & Reeve, C. L. (2006). Differentiation of cognitive abilities as function of neuroticism level: A measurement equivalence / invariance analysis. *Intelligence*, 34(4), 403-417.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.
- Bujo, B. (2009). Is there a specific African ethic? Towards a discussion with Western thought. In M. F. Murove (Ed.), *African ethics: An anthology of comparative and applied ethics* (pp. 113-128). Scottsville, South Africa: University of KwaZulu-Natal Press.
- Byrne, B. (1998). *Structural Equation Modelling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Lawrence Erlbaum Associates.
- Byrne, B. M. (2009). Structural equation modeling: Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(2), 272–300. https://doi.org/10.1207/s15328007sem1102_8
- 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, 456-466.
- Byrne, B., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of cross-cultural psychology*, 34(2), 155-175.
- Byrnes, R. M. (Ed.). (1996). *South Africa: A country study*. GPO for the Library of Congress. <http://countrystudies.us/south-africa/47.htm>
- Callegari C, Bertù L, Lucano M, Ielmini M, Braggio E, Vender S. (2016). Reliability and validity of the Italian version of the 14-item Resilience Scale. *Psychology Research and Behaviour Management*.9, p.277-284 <https://doi.org/10.2147/PRBM.S115657>

- Carver, C.S. & Scheier, M.S. (2005). Optimism. In Snyder, C.R. & Lopez, S.J. (Eds.), *Handbook of positive psychology* (p. 231-243). Oxford University Press.
- Cattell, R. B. (1978). *The Scientific Use of Factor Analysis*. Plenum Press.
- Chen, F., Carolina, N., Curran, P.J., Bollen, K.A., & Kirby, J. (2009). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Social Methods Research, 36*(4), 1–31. <https://doi.org/10.1177/0049124108314720>.
- Cheung, F. M., Cheung, S. F., Wada, S., & Zhang, J. (2003). Indigenous Measures of Personality Assessment in Asian Countries: A Review. *Psychological Assessment, 15*(3), 280–289. <https://doi.org/10.1037/1040-3590.15.3.280>
- Cheung, F., Tang, C. S., & Tang, S. (2010). Psychological capital as a moderator between emotional labor, burnout, and job satisfaction among school teachers in China. *International Journal of Stress Management, 18*, 348–371. doi:10.1037/a0025787
- Cheung, G. W., & Rensvold, R. B. (1999). Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. *Journal of Management, 25*(1), 1–27. <https://doi.org/10.1177/014920639902500101>
- Cheung, G., & Rensvold, R. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255.
- Cleary, T.A. (1968). Test bias: prediction of grades of Negro and white students in integrated Colleges. *Journal of Educational Measurement, 5*, 115-124.
- Cohen, L., Manion, L. & Morrison, K. (2000). *Research Methods in Education*. Routledge Falmer.
- Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research, 20*, 389-417.
- Combs, G. M., Milosevic, I., Jeung, W., & Griffith, J. (2012). Ethnic identity and job attribute preferences: The role of collectivism and psychological capital. *Journal of Leadership & Organizational Studies, 19* (1), 5–16. doi:10.1177/1548051811433359
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4). P. 281–302.
- Dalal, D. K., & Carter, N. T. (2015). Negatively worded items negatively impact survey research. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (p. 112–132). Routledge/Taylor & Francis Group

- Davis, S. (2014). *The measurement invariance and measurement equivalence of the Sources of Work Stress Inventory (SWSI) across gender groups in South Africa*. Stellenbosch: Unpublished masters' thesis, University of Stellenbosch.
- Dawkins, S., Martin, A., Scott, J. & Sanderson, K. (2013) Building on the positives: A psychometric review and critical analysis of the construct of Psychological Capital. *Journal of Occupational and Organisational Psychology*, 86. P. 348-370
- De Kock, F. (2018). Cross-cultural test adaptation and translation. In Foxcroft, C & Roodt, G. (Eds.), *Introduction to psychological Assessment in the South African Context*, Oxford University Press.
- Diamantopoulos, A. & Siguaw, J. A. (2000). *Introducing LISREL: A guide for the uninitiated*. Sage
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55, 34–43
- DiStefano, C. & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys, structural equation modeling, *A Multidisciplinary Journal* 13(3), 440–464, https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., Zhu, M. & Mîndrilă, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher, *Practical Assessment, Research, and Evaluation*, 14(20).
- Drasgow, F. & Hulin, C.L. (1990). Item response theory. In M.D. Dunnette & L.M. Hough (Eds.). *Handbook of industrial and organizational psychology* (Second edition). Consulting Psychologists Press.
- Du Plessis, Y., & Barkhuizen, N. (2012). Psychological capital, a requisite for organisational performance in South Africa. *South African Journal of Economic and Management Sciences*, 15(1), 16–30.
- Dunbar, H., Theron, C., & Spanenberg, H. (2011). A cross-validation study of the Performance Index. *Management Dynamics*, 20(3). P.2-24.
- Einhorn, H.J. & Bass, A.R. (1971). Methodological considerations relevant to discrimination in employment testing. *Psychological Bulletin*, 75. p. 261-269.
- Ekman, P., & Friesen, W. (1975). *Unmasking the human face: a guide to recognizing emotions from facial expressions*. Englewood Cliffs, NJ: Prentice-Hall.
- Fallan, L. & Opstad, L. (2016). Student Self-Efficacy and Gender-Personality Interactions. *International Journal of Higher Education*. 5,(3). P. 32-44.

- Fontaine, J. R. J. (2008). Traditional and multilevel approaches in Cross-Cultural Research: an integration of methodological frameworks. In F. J. R. van de Vijver, D. A. van Hemert, & Y. H. Poortinga. *Multilevel Analysis of Individuals and Cultures*. Lawrence Erlbaum Associates.
- Foxcroft, C. D. (1997). Psychological testing in South Africa: Perspectives regarding ethical and fair practices. *European Journal of Psychological Assessment*, 13(3), 229–235. <https://doi.org/10.1027/1015-5759.13.3.229>
- Fredrickson, B. L. (2003). Positive emotions and upward spirals in organizations. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 63–175). Berrett-Koehler.
- French, B.F. and Finch, W.H. (2006) Confirmatory Factor Analytic Procedures for the Determination of Measurement Invariance. *Structural Equation Modeling*, 13, 378-402. http://dx.doi.org/10.1207/s15328007sem1303_3
- Frey, B. (2018). Predictive Validity, in *The SAGE Encyclopedia of Educational Research, Measurement and Evaluation*. SAGE Publications, Inc.
- Gartland, D., Bond, L., Olsson, C.A., Buzwell, S. & Sawyer, S.M. (2011). Development of a multi-dimensional measure of resilience in adolescents: The adolescent resilience questionnaire. *BMC Med Res Methodology*, 11(1):134. <http://dx.doi.org/10.1186/1471-2288-11-134>
- George, D., & Mallery, P. (2003). *Using SPSS for Windows Step by Step: A Simple Guide and Reference* (4th ed.). London: Pearson Education.
- Gibson, A & Hicks, R. (2018). Psychological Capital and Core Self Evaluations in the Workplace: Impacts on Well-being. *International Journal of Psychological Studies*, 10(2). P. 15-24.
- Gignac, G. E. (2005). *Determining the dimensionality of a self-report emotional intelligence inventory (SUEIT) and testing its unique factorial validity*. Unpublished doctoral dissertation. Swinburne University of Technology.
- Görgens-Ekermans, G., & Herbert, M. (2013). Psychological capital: Internal and external validity of the Psychological Capital Questionnaire (PCQ-24) on a South African sample. *SA Journal of Industrial Psychology* 39(2) 1-12
- Government Gazette, Republic of South Africa (1998). *Vol. 400, no 19370*. Cape Town, 19 October 1998.
- Government Gazette. (2017). List of Classified and Certified Psychological Tests. *Health Professions Act, 1974 (Act 56 of 1974)*.

- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What They Are and How to Use Them. *Educational and Psychological Measurement*, 66(6), 930–944. <https://doi.org/10.1177/0013164406288165>
- Gunn, H. (2016). *Approaches to studying measurement invariance in multilevel data with a level-1 grouping variable*. Unpublished thesis, Arizona State University.
- Hair, J. F., Black, W., Babin, B. & Anderson, R. E. (2010). *Multivariate Data Analysis (7th edition)*. USA: Prentice Hall.
- Harzing, A.W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6(2), 243–266. <https://doi.org/10.1177/1470595806066332>
- He, J. & Van de Vijver, F.J.R. (2015). The value of keeping an open eye for methodological issues in research on resilience and culture. In: Theron L, Liebenberg L, Ungar M, editors. *Youth resilience and culture: Complexities and commonalities*. Springer, p. 189–202. http://dx.doi.org/10.1007/978-94-017-9415-2_14
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Beuhner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16, 319–336.
- Heilemann, M., Kathryn, L. & Salvador, K. (2003). Psychometric Properties of the Spanish Version of the Resilience Scale. *Journal of Nursing Measurement*, 11(1), DOI: 10.1891/jnum.11.1.61.52067
- Herbert, M. (2011). An exploration of the relationships between Psychological Capital (Hope, Optimism, Self-efficacy, Resilience), Occupational Stress, Burnout and Employee Engagement. Masters Thesis, Stellenbosch University.
- Herholdt, K. (2015). *Determinants of Work Engagement and Organisational Citizenship Behaviour amongst Nurses*. Masters thesis, Stellenbosch University.
- Hobfoll, S. E. (2002). Social and psychological resources and adaptation. *Review of General Psychology*, 6, 307–324.
- Hofstede, G. (1980). *Culture's Consequences: international differences in work-related values*. Sage Publications.

- Horn, J. L. (1991). Comments on 'Issues in Factorial Invariance' in Best Methods for the Analysis of Change. In L. M. Collins & J. L. Horn (eds.), *Best Methods for the Analysis of Change* (pp. 114-125). American Psychological Association
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144
- Hox, J. & Boeije, H. (2005). Data Collection, Primary vs. Secondary. *Encyclopaedia of Social Measurement, 1*. P. 593- 599.
- Hu, L. and Bentler, P. (1998) Fit Indices in Covariance Structure modelling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453. <http://dx.doi.org/10.1037/1082-989X.3.4.424>
- Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76–99). London: Sage.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Huysamen, G.K. (1996). The socio-political context of the application of fair selection models in the USA. *Journal of Industrial Psychology, 22*, 1-6.
- Huysamen, G.K. (2002). The relevance of the new APA standards for educational and psychological testing for employment testing in South Africa. *South African Journal of Psychology, 32*, 26-33.
- Jackson Barnette, J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems, *Educational and Psychological Measurement 60*(3), 361–370. <https://doi.org/10.1177/00131640021970592>
- Johnson, T. P., Shavitt, S., & Holbrook, A. L. (2011). Survey response styles across cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Culture and psychology. Cross-cultural research methods in psychology* (p. 130–175). Cambridge University Press.
- Jones, C. (2010). Archival Data: Advantages and Disadvantages for Research in Psychology. *Social and Personality Psychology Compass, 4* (11). P. 1008-1017
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.

- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Kanengoni, H. Ngarambe, C. N. & Buitendach, J. H. (2018). Psychological capital and work behaviour-related outcomes among South African church ministers. *South African Journal of Psychology*, 48(4). 488-500.
- Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Sage Publications.
- Kaplan, R. & Saccuzzo, D. (2009). *Psychological Testing: principles, Applications and Issues*. Cengage Learning.
- Kelloway, E. (1998). *Using LISREL for structural equation modelling: A researcher's guide*. Sage.
- Kerlinger, F.N., & Lee, H.B. (2000). *Foundations of Behavioural Research*. (4th ed.) Harcourt College.
- Kirman, M.N. & Jahan, F. (2015). Study the Psychological Capital (PsyCap) among Private Bank Employees in Aligarh. *Indian Journal of Positive Psychology*, 6(1), 79-83.
- Klika, J.B & Herrenkohl, T.I. (2013). Review of developmental research on resilience in maltreated children. *Trauma Violence Abuse*, 14(3):222–234.
<http://dx.doi.org/10.1177/1524838013487808>
- Kline, R. (2013). Exploratory and confirmatory factor analysis. *Applied quantitative analysis in the social sciences*, 171-207.
- Kline, T. J. (2005). Classical test theory: assumptions, equations, limitations, and item analyses. In Kline, T. J. *Psychological testing: A practical approach to design and evaluation* (pp. 91-106). SAGE Publications, Inc. doi: 10.4135/9781483385693
- Kotze, M. (2018) The influence of psychological capital, self-leadership and mindfulness on work engagement. *South African Journal of Psychology*, 48(2), 279-292.
- Kumar, R. & Lal, R. (2006), The Role of Self-Efficacy and Gender Difference among the Adolescents. *Journal of the Indian Academy of Applied Psychology*, 32(3), 249-254.
- Laher, S. & Cockroft, K. (2013). *Psychological Assessment in South Africa: Research and Applications*. Wits University Press.
- Lamba, P. (2017). Gender Based Comparative Analysis of Psychological Capital in R & D Organizations, *International Journal of Education and Psychological Research (IJEPR)*, 6 (4). P. 15-18.
- Langenhoven, A. (2015). *How Job Demands and Resources predict Burnout, Engagement and Intention to Quit in Call Centres*. Masters thesis, Stellenbosch University

- Laschinger, H.K.S., & Grau, A.L. (2012). The influence of personal dispositional factors and organizational resources on workplace violence, burnout, and health outcomes in new graduate nurses: A cross-sectional study. *International Journal of Nursing Studies*, 49, 282–291. <http://dx.doi.org/10.1016/j.ijnurstu.2011.09.004>, PMID:21978860
- Leedy, P.D. & Ormrod, J.E. (2005) *Practical Research: Planning and Design*. Prentice Hall, Upper Saddle River, NJ.
- Lerner, R. M. (2006). Resilience as an attribute of the developmental system: Comments on the papers of Professors Masten & Wachs. In: Lester BM, Masten MS, McEwen B, editors. *Resilience in children*. Blackwell. p. 40–51. <http://dx.doi.org/10.1196/annals.1376.005>
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173.
- Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological methods*, 18(3), 285–300. <https://doi.org/10.1037/a0033266>
- Little, T.D. (1997). Mean and Covariance Structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate behavioural research*, 32(1), 53-76.
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. New York, NY: Taylor & Francis.
- Losoi, H., Turunen, S., Wäljas, M., Helminen, M., Öhman, J., Julkunen, J. & Rosti-Otajärvi, E. (2013). Psychometric Properties of the Finnish Version of the Resilience Scale and its Short Version. *Psychology, Community & Health*, 2(1), 1–10, doi:10.5964/pch.v2i1.40
- Lundman, B., Strandberg, G. Eisemann, M., Gustafson, Y. & Brulin, C. (2007). Psychometric properties of the Swedish version of the Resilience Scale. *Scandinavian Journal of Caring Sciences*, 21. P. 229–237.
- Luthans, F. (2002a). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 23(6), 695-706.
- Luthans, F. (2002b). Positive organizational behavior: Developing and managing psychological strengths. *Academy of Management Perspectives*, 16(1), 57-72.

- Luthans, F., & Jensen, S. M. (2002). Hope: A New Positive Strength for Human Resource Development. *Human Resource Development Review*, 1(3), 304–322. <https://doi.org/10.1177/1534484302013003>
- Luthans, F., & Youssef, C. M. (2004). Human, Social, and Now Positive Psychological Capital Management: Investing in people for competitive advantage. *Organizational Dynamics*, 33(2), 143–160. <https://doi.org/10.1016/j.orgdyn.2004.01.003>
- Luthans, F., Avey, J.B., Avolio, B.J., & Peterson, S.J. (2010). The development and resulting performance impact of positive psychological capital. *Human Resource Development Quarterly*, 21, 41–67. <http://dx.doi.org/10.1002/hrdq.20034>
- Luthans, F., Avolio, B.J., Avey, J.B. & Norman, S.M. (2007a). Positive Psychological Capital: Measurement and Relationship with Performance and Satisfaction. *Personnel Psychology*; 60: p. 541-572
- Luthans, F., Norman, S.M., Avolio, B.J. & Avey, J.B. (2008). The Mediating Role of Psychological Capital in the Supportive Organisational Climate: Employee Performance Relationship. *Journal of Organisational Behaviour*, 29(2). P. 219-238.
- Luthans, F., Youssef, C. M., Sweetman, D., & Harms, P. (2013). Meeting the leadership challenge of employee well-being through relationship PsyCap and health PsyCap. *Journal of Leadership and Organizational Studies*, 20, 114–129.
- Luthans, F., Youssef, C.M., & Avolio, B.J. (2007b). Psychological capital: Developing the human competitive edge. Oxford, UK: Oxford University Press
- Luthans, F., Youssef-Morgan, C. M., & Avolio, B. (2015). *PsyCap and beyond*. New York: Oxford University Press.
- Luthar, S.S., Cicchetti, D. & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*, 71(3):543–562. <http://dx.doi.org/10.1111/1467-8624.00164>
- Marsh, H. W. (1994). Confirmatory factor models of factorial invariance: A multi-faceted approach. *Structural Equation Modeling*, 1, 5-34.
- Masten, A.S. & Reed, M.J. (2002). Resilience in development. In C.R. Snyder & S. Lopez (Eds.). *Handbook of positive psychology* (pp. 74-88). Oxford University Press.

- Masten, A.S. (2011). Resilience in children threatened by extreme adversity: Frameworks for research, practice, and translational synergy. *Developmental Psychopathology*, 23(2):493–506. <http://dx.doi.org/10.1017/S0954579411000198>
- McNeish, D. J. & Hancock, G. R. (2018) The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, 100(1), 43-52, doi: 10.1080/00223891.2017.1281286
- Meade, A. W. and Lautenschlager, G.J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organisational Research Methods*, 7(4): 361-388.
- Meade, A., Johnson, E. & Braddy, P. (2008). Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology*, 93(3) pp. 568–592
- Meiring, D., Van de Vijver, A.J.R., Rothmann, S., & Barrick, M.R. (2005). Construct, item, and method bias of cognitive and personality measures in South Africa. *South African Journal of Industrial Psychology*, 31(1), p. 1–8.
- Meiring, D., Van de Vijver, F.J.R., & Rothmann, S. (2006). Bias in an adapted version of the 15FQ+ in South Africa. *South African Journal of Psychology* 36(2), p. 340-356.
- Mels, G. (2010). *Structural Equation Modelling with LISREL 9 for Windows*. Scientific Software International.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, p. 525-543.
- Middleton, K. L., & Jones, J. L. (2000). Socially desirable response sets: The impact of country culture. *Psychology & Marketing*, 17(2), 149-163.
- Millsap, R.E. & Everson, H.T. (1993). Methodological review: statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, p. 297-334.
- Mind Garden, Inc. (n.d.). *Psychological Capital Questionnaire (PCQ) - Assessments, Tests | Mind Garden - Mind Garden*. <https://www.mindgarden.com/136-psychological-capital-questionnaire>
- Mishra, S. & Shafiq, M. (2018). Psychological Capital, Social Support and Psychological Well-being among Adults. *Delhi Psychiatry Journal*, 21(1). P. 80-84

- Mkhize, N. (2006). African traditions and the social, economic and moral dimensions of fatherhood. In L. M. Richter & R. Morrell (Eds.), *Baba: Men and fatherhood in South Africa* (pp. 183-198). HSRC Press.
- Mokwena, M. (2007). African cosmology and psychology. In M. Visser (Ed.), *Contextualising community psychology in South Africa* (pp. 66-78). Van Schaik.
- Moyo, S., & Theron, C. (2011). A preliminary factor analytic investigation into the first-order factor structure of the Fifteen Factor Plus (15FQ+) on a sample of Black South African managers. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 37(1), 1-22.
- Mulaik, S. A. (1975). Confirmatory factor analysis. In D. J. Amick & H. J. Walberg (Eds.), *Introductory multivariate analysis* (pp. 170-207). McCutchan.
- Munyaka, S.A., Boshof, A.B., Pietersen, J. & Snelgar, R. (2017). The relationships between authentic leadership, psychological capital, psychological climate, team commitment and intention to quit. *SA journal of Industrial Psychology*, 43(0), a1430.
- Murove, M. F. (2009). Beyond the savage evidence ethic: A vindication of African ethics. In M. F. Murove (Ed.), *African ethics: An anthology of comparative and applied ethics* (pp. 14-32). University of KwaZulu-Natal Press.
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171-189.
- Narasimha. M.L & , Reddy, L. K. (2017). Impact of Emotional Maturity, Intelligence and Self-Efficacy on the Academic Achievement of Teacher Trainees. *Indian journal of positive psychology*, 8(2), 395-397.
- Nelson, D. & Cooper, C. (2007). *Positive Organisational Behaviour*. SAGE Publications LTD. <http://dx.doi.org/10.4135/9781446212752>
- Ngwenya, B. & Pelsler, T. (2020). Impact of psychological capital on employee engagement, job satisfaction and employee performance in the manufacturing sector in Zimbabwe. *SA Journal of Industrial Psychology*, 46(0). A1781
- Nigah, N., Davis., A.J., & Hurrell, S.A. (2012). The impact of buddying on psychological capital and work engagement: an empirical study of socialization in the professional services sector. *Thunderbird International Business Review*, 54(6), 891–905. <http://dx.doi.org/10.1002/tie.21510>

- Nolzen, N. (2018). The concept of psychological capital: a comprehensive review. *Management Review Quarterly*, 68. P 237–277
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), pp. 625-632.
<https://link.springer.com/article/10.1007%2Fs10459-010-9222-y#citeas>
- Nunnally, J. C. (1978). *Psychometric Theory (2nd ed.)*. New York: McGraw-Hill.
- Nzimande, B. (1995). *To test or not to test? Paper presented at the Congress on Psychometrics for Psychologists and Personnel Practitioners*. Council for Scientific and Industrial Research, p. 13–14.
- Odendaal, A. (2015). Cross-cultural differences in social desirability scales: Influence of cognitive ability. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 41(1), Art. #1259, 13 pages.
- Oladipo, S. & Idemudia, E. (2017) Reliability and Validity Testing of Wagnild and Young’s Resilience Scale in a Sample of Nigerian Youth. *Journal of Psychology*, 6(1), p. 57 – 65
- Peterson, S. J., Luthans, F., Avolio, B. J., Walumbwa, F. O., & Zhang, Z. (2011). Psychological capital and employee performance: A latent growth modelling approach. *Personnel Psychology*, 64, 427–450. doi:10.1111/j.1744-6570.2011.01215.x
- Pinheiro, M., Matos, A., Pestana, C., Oliveira, S. & Costa, J. (2015). The Resilience Scale: A study in a Portuguese adult sample. *The European Proceedings of Social & Behavioural Sciences*. P. 67 – 80. <http://dx.doi.org/10.15405/epsbs.2015.08.7>
- Portzky, M., Wagnild, G., De Bacquer, D., & Audenaert, K. (2010). Psychometric evaluation of the Dutch Resilience Scale RS-nl on 3265 healthy participants: a confirmation of the association between age and resilience found with the Swedish version. *Scandinavian Journal of Caring Sciences*, 24(1), 86–92. doi:10.1111/j.1471-6712.2010.00841.x
- Prinsloo, J. (2013). *Modification, Elaboration and Empirical evaluation of the Burger Learning Potential Structural Model*. Masters Thesis, Stellenbosch University.
- Prozesky, M. H. (2009). Cinderella, survivor and saviour: African ethics and the quest for a global ethic. In M. F. Murove (Ed.), *African ethics: An anthology of comparative and applied ethics* (pp. 3-13). University of KwaZulu-Natal Press.
- Putnick, D., & Bornstein, M. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90

- Ramsden, K. (2019). *The effects of psychological capital and job satisfaction on work engagement of support staff at a Holdings establishment (Meridian Holdings)*. Masters thesis, Stellenbosch University.
- Rani, E.K. & Chaturvedula, S. (2018). Psychological Capital: Gender Differences and its Relationship with Job Involvement. *Defence Life Science Journal*, 3(4). 383-387.
- Rego, A., Sousa, F., Marques, C., & Cunha, M.P. (2012). Authentic leadership promoting employees' psychological capital and creativity. *Journal of Business Research*, 65, 429–437. <http://dx.doi.org/10.1016/j.jbusres.2011.10.003>
- Renner, W., Peltzer, K., & Phaswana, M. G. (2003). The structure of values among Northern Sotho speaking people in South Africa. *South African Journal of Psychology*, 33, 103-108.
- Republic of South Africa (RSA). (1998). *Employment Equity Act 55 of 1998*. Juta Law.
- Republic of South Africa (RSA). (2006). *Ethical and professional rules of conduct for practitioners registered under the health professions act (Act 56 of 1974)*.
- Reyes, J., Chavez, J., Beltran, F., Guerrero, A. & Enriques, M. (2017). A gender study on college students' academic self-efficacy. *Science Journal of Education*, 2(6): 180-184.
- Roberts, S. J., Scherer, L. L., & Bowyer, C. J. (2011). Job stress and incivility: What role does psychological capital play? *Journal of Leadership & Organizational Studies*, 18(4), 449–458. doi:10.1177/1548051811409044
- Roche, M., Haar, J., & Luthans, F. (2014). The role of mindfulness and psychological capital on the wellbeing of leaders. *Journal of Occupational Health Psychology*, in press.
- Rock, D. A., Werts, C. E., & Flaugh, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, 13, 403-418.
- Roemer, A. & Harris, C. (2018). Perceived organisational support and well-being: The role of psychological capital as a mediator. *SA Journal of Industrial Psychology*, 44(0)
- Roux, S. (2014). *The impact of positive organisational factors on the career success of black employees in the South African work environment: An exploratory study*. PhD Dissertation, Stellenbosch University.
- Schaie, K.W., & Hertzog, C. (1985). Measurement in the psychology of adulthood and aging. In J. E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 61- 92). Van Nostrand Reinhold.

- Schermelleh-Engel, K., Moosbrugger, H. & Mueller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measured. *Methods of Psychological Research Online*, 8(2). 23 – 74.
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367–373. <https://doi.org/10.1177/014662168500900405>
- Schneider, S. L. (2001). In search of realistic optimism: Meaning, knowledge, and warm fuzziness. *American Psychologist*, 56(3), 250–263. <https://doi.org/10.1037/0003-066X.56.3.250>
- Schuttleworth-Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G. L., Hartman, N. P., & Radloff, S. E. (2004). Cross-Cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, 26(7), 903-920.
- Sehlapelo, M. & Terre Blanche, M. (1996). Psychometric testing in South Africa: Views from above and below. *Psychology in Society*, 21, 49–59.
- Sekaran, U. (1983) Methodological and Theoretical Issues and Advancements in Cross-Cultural Research. *Journal of International Business Studies*, 14, 61–73. <https://doi.org/10.1057/palgrave.jibs.8490519>
- Selig, J.P., Card, N.A., & Little, T.D. (2008). *Latent variable structural equation modelling in cross-cultural research: Multigroup and multilevel approaches*. In F.J.R. Van de Vijver, D.R. Van Hemert & A. Poortinga (Eds.), *Multilevel analysis of individuals and cultures* (pp. 93–119). Psychology Press.
- Seligman, M. E. P. (1998). *Learned optimism*. New York: Pocket Books
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55(1), 5–14. <https://doi.org/10.1037/0003-066X.55.1.5>
- Sepeng, W., Stander, M.W., Van der Vaar, L. & Coxen, L. (2020). Authentic leadership, organisational citizenship behaviour and intention to leave: The role of psychological capital. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 46(0), a1802. <https://doi.org/10.4102/sajip.v46i0.1802>
- Simons, J.C., & Buitendach, J.H. (2013). Psychological capital, work engagement and organisational commitment amongst call centre employees in South Africa. *SA Journal of Industrial*

- Psychology/SA Tydskrif vir Bedryfsielkunde*, 39(2), Art. #1071, 12 pages.
<http://dx.doi.org/10.4102/sajip.v39i2.1071>
- Sireci, S.G. (2011). Evaluating test and survey items for bias across languages and cultures. In D. Matsumoto & F.J.R. van de Vijver, *Cross-cultural Research Methods in Psychology* (pp. 216-240). Cambridge University Press
- Snyder, C. R. (2000). *Handbook of hope*. Academic Press
- Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., Gibb, J., Langelle, C., & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology*, 60(4), 570–585. <https://doi.org/10.1037/0022-3514.60.4.570>
- Staples, H. (2014) *The Generational Divide: Generational Differences in PsyCap*. Thesis, University of the Incarnate Word.
- Statistics South Africa. (2016). *Community Survey 2016, Statistical release*.
http://cs2016.statssa.gov.za/wp-content/uploads/2016/07/NT-30-06-2016-RELEASE-for-CS-2016-Statistical-releas_1-July-2016.pdf
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78 - 90.
- Stellenbosch University. (2013). *Policy for responsible research conduct at Stellenbosch University*. Stellenbosch University.
- Steyn, R., & De Bruin, G.P. (2020). An investigation of gender-based differences in assessment instruments: A test of measurement invariance. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 46(0), a1699. <https://doi.org/10.4102/sajip.v46i0.1699>
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L., García-Cueto, E. Cuesta, M. & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2) p. 149 – 158.
- Sullivan, G. & Artino Jr., A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*. 5(4), pp. 541-542.
- Surzykiewicz, J., Konaszewski, K., & Wagnild, G. (2019). Polish Version of the Resilience Scale (RS-14): A Validity and Reliability Study in Three Samples. *Frontiers in psychology*, 9, 2762. <https://doi.org/10.3389/fpsyg.2018.02762>
- Tabachnick, B. & Fidell, L. (2013). *Using Multivariate Statistics*, Sixth edition. Pearson Education, Inc.

- Theron, C. (2007). Confessions, scapegoats and flying pigs: psychometric testing and the law. *South African Journal of Industrial Psychology*, 33(1), 102-117.
- Theron, L., Theron, A. & Malindi, M. (2013). Toward an African Definition of Resilience: A Rural South African Community's View of Resilient Basotho Youth. *Journal of Black Psychology*, 39(1), 63-87.
- Theron, C.C (2017) *Research Methodology*. Unpublished lecture slides. Stellenbosch University
- Tol, A., Song, S. & Jordans, M.J.D. (2013). Annual research review: Resilience and mental health in children and adolescents living in areas of armed conflict – A systematic review of findings in low- and middle-income countries. *Journal of Child Psychology and Psychiatry*, 54(4). 445–460. <http://dx.doi.org/10.1111/jcpp.12053>
- Tripathi, P. (2011). Employee well-being: Role of psychological capital. *Amity Journal of Applied Psychology*, 2(1), 18–22.
- Ungar M. (2013). Resilience, trauma, context, and culture. *Trauma Violence Abuse*, 14(3). 255–266. <http://dx.doi.org/10.1177/1524838013487805>
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. <https://doi.org/10.1080/17405629.2012.686740>
- Van de Vijver, A., & Poortinga, Y. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13(1), 29-37.
- 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.
- Van de Vijver, F. J. R. & Rothmann, S. (2004). Assessment in multicultural groups: The South African case. *South African Journal of Industrial Psychology*, 30(4), 1–7.
- Van de Vijver, F. J. R., & Leung, K. (1997a). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed.). pp. 257-300. Allyn & Bacon.
- Van de Vijver, F. J. R., & Leung, K. (1997b). *Methods and data analysis for cross-cultural research*. Sage.
- Van de Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto & F.J.R. van de Vijver, *Cross-cultural Research Methods in Psychology* (pp. 17-45). Cambridge University Press.

- Van de Vijver, F., & Tanzer, N. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology, 54*, 119-135.
- Van der Bank, S. (2019). *An investigation into the Measurement Invariance and Measurement Equivalence of the South African Personality Inventory across gender groups in South Africa*. Masters Thesis, Stellenbosch University.
- Van der Merwe, B. (2016). *Salient predictors of intention to quit among sales employees within the South African financial services industry*. Masters Thesis, Stellenbosch University.
- Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Correction: Ineffectiveness of Reverse Wording of Questionnaire Items: Let's Learn from Cows in the Rain. *PLoS ONE, 8*(9), 10.1371/annotation/af78b324-7b44-4f89-b932-e851fe04a8e5.
<https://doi.org/10.1371/annotation/af78b324-7b44-4f89-b932-e851fe04a8e5>
- Van Wyk, N. (2016). *The role of Psychological Capital in protecting the psychological well-being of individuals working in call centres*. Masters Thesis, Stellenbosch University.
- 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*(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Wagnild, G.M., & Young, H.M. (1993). Development and Psychometric evaluation of the resiliency scale. *Journal of Nursing Management, 1*(2), 165-178.
- Walsh, V.A., Dawson, J. & Mattingly, M.J. (2010). How are we measuring resilience following childhood maltreatment? Is the research adequate and consistent? What is the impact on research, practice, and policy? *Trauma Violence Abuse, 11*(1):27–41.
<http://dx.doi.org/10.1177/1524838009358892>
- Wan, W. & Hu, X. (2017). *A Review of Literature on PsyCap*. International Conference on Economics and Management Engineering.
- Wang, Y., Chang, Y., Fu, J., & Wang, L. (2012). Work-family conflict and burnout among Chinese female nurses: The mediating effect of psychological capital. *BMC Public Health, 12*, 915–923.
<http://dx.doi.org/10.1186/1471-2458-12-915>
- Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reverse item bias: An integrative model. *Psychological Methods, 18*, 320-334.
- Windle, G., Bennett, K. & Noyes, J. (2011). *A methodological review of resilience measurement scales*. Health Quality Life Outcomes: <http://www.hqlo.com/content/9/1/8>.

- Wright, T. A., & Hobfoll, S. E. (2004). Commitment, psychological well-being and job performance: An examination of conservation of resources (COR) theory and job burnout. *Journal of Business and Management, 9*, 389–406
- Wu, A., Li, Z. & Zumbo, B. (2007) Decoding the Meaning of Factorial Invariance and Updating the Practice of Multi-group Confirmatory Factor Analysis: A Demonstration With TIMSS Data. *Practical Assessment Research & Evaluation, 12*(3). University of British Columbia.
- Yoon, M., & Kim, E. (2014). A comparison of sequential and nonsequential specification searches in testing factorial invariance. *Behavioural research methods, 46*, 1199-1206.
- Youssef, C. M., & Luthans, F. (2013). Developing psychological capital in organizations: Cognitive, affective, conative, and social contributions of happiness. In S. A. David, I. Boniwell, & A. C. Ayers (Eds.), *Oxford handbook of happiness* (pp. 751–766). Oxford University Press.
- Youssef-Morgan, C. M. & Luthans, F. (2015). Psychological Capital and Well-being. *Stress and Health, 31*. 180-188.
- Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology, 34*, 390-400.

**APPENDIX A: GOODNESS OF FIT STATISTICS FOR THE PCQ-24 MULTIGROUP CONFIGURAL
INVARIANCE MEASUREMENT MODEL**

Degrees of Freedom = 738
 Minimum Fit Function Chi-Square = 1778.4192 (P = 0.0)
 Normal Theory Weighted Least Squares Chi-Square = 1793.6968 (P = 0.0)
 Satorra-Bentler Scaled Chi-Square = 1251.3448 (P = 0.0)
 Estimated Non-centrality Parameter (NCP) = 513.3448
 90 Percent Confidence Interval for NCP = (419.5286 ; 615.0256)
 Minimum Fit Function Value = 2.4262
 Population Discrepancy Function Value (F0) = 0.7003
 90 Percent Confidence Interval for F0 = (0.5723 ; 0.8391)
 Root Mean Square Error of Approximation (RMSEA) = 0.05336
 90 Percent Confidence Interval for RMSEA = (0.04823 ; 0.05840)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.1379
 Expected Cross-Validation Index (ECVI) = 2.3456
 90 Percent Confidence Interval for ECVI = (2.1194 ; 2.3861)
 ECVI for Saturated Model = 0.8186
 ECVI for Independence Model = 29.9610
 Chi-Square for Independence Model with 828 Degrees of Freedom = 21913.4216
 Independence AIC = 22057.4216
 Model AIC = 1719.3448
 Saturated AIC = 1800.0000
 Independence CAIC = 22460.7102
 Model CAIC = 3030.0327
 Saturated CAIC = 6841.1071
 Normed Fit Index (NFI) = 0.9429
 Non-Normed Fit Index (NNFI) = 0.9727
 Parsimony Normed Fit Index (PNFI) = 0.8404
 Comparative Fit Index (CFI) = 0.9757
 Incremental Fit Index (IFI) = 0.9758
 Relative Fit Index (RFI) = 0.9359
 Critical N (CN) = 487.3731
 Contribution to Chi-Square = 568.9627
 Percentage Contribution to Chi-Square = 31.9926
 Root Mean Square Residual (RMR) = 0.08923
 Standardized RMR = 0.06667
 Goodness of Fit Index (GFI) = 0.8416

APPENDIX B: GOODNESS OF FIT STATISTICS FOR THE PCQ-24 MULTIGROUP WEAK INVARIANCE**MEASUREMENT MODEL**

Degrees of Freedom = 786
Minimum Fit Function Chi-Square = 1884.4263 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 1881.3256 (P = 0.0)
Satorra-Bentler Scaled Chi-Square = 1302.2169 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 516.2169
90 Percent Confidence Interval for NCP = (420.9336 ; 619.3820)
Minimum Fit Function Value = 2.5708
Population Discrepancy Function Value (F0) = 0.7043
90 Percent Confidence Interval for F0 = (0.5743 ; 0.8450)
Root Mean Square Error of Approximation (RMSEA) = 0.05185
90 Percent Confidence Interval for RMSEA = (0.04682 ; 0.05679)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.2677
Expected Cross-Validation Index (ECVI) = 2.2841
90 Percent Confidence Interval for ECVI = (2.0558 ; 2.3266)
ECVI for Saturated Model = 0.8186
ECVI for Independence Model = 29.9610
Chi-Square for Independence Model with 828 Degrees of Freedom = 21913.4216
Independence AIC = 22057.4216
Model AIC = 1674.2169
Saturated AIC = 1800.0000
Independence CAIC = 22460.7102
Model CAIC = 2716.0457
Saturated CAIC = 6841.1071
Normed Fit Index (NFI) = 0.9406
Non-Normed Fit Index (NNFI) = 0.9742
Parsimony Normed Fit Index (PNFI) = 0.8929
Comparative Fit Index (CFI) = 0.9755
Incremental Fit Index (IFI) = 0.9756
Relative Fit Index (RFI) = 0.9374
Critical N (CN) = 497.0024
Contribution to Chi-Square = 587.7705
Percentage Contribution to Chi-Square = 31.1910
Root Mean Square Residual (RMR) = 0.09983
Standardized RMR = 0.08381
Goodness of Fit Index (GFI) = 0.8379

APPENDIX C: GOODNESS OF FIT STATISTICS FOR THE PCQ-24 MULTIGROUP STRONG INVARIANCE**MEASUREMENT MODEL**

Degrees of Freedom = 834
Minimum Fit Function Chi-Square = 2220.4165 (P = 0.0)
Normal Theory Weighted Least Squares Chi-Square = 2264.0134 (P = 0.0)
Satorra-Bentler Scaled Chi-Square = 1645.4537 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 811.4537
90 Percent Confidence Interval for NCP = (700.0014 ; 930.6705)
Minimum Fit Function Value = 3.0292
Population Discrepancy Function Value (F0) = 1.1070
90 Percent Confidence Interval for F0 = (0.9550 ; 1.2697)
Root Mean Square Error of Approximation (RMSEA) = 0.06310
90 Percent Confidence Interval for RMSEA = (0.05861 ; 0.06758)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0000
Expected Cross-Validation Index (ECVI) = 2.6214
90 Percent Confidence Interval for ECVI = (2.3711 ; 2.6858)
ECVI for Saturated Model = 0.8186
ECVI for Independence Model = 29.9610
Chi-Square for Independence Model with 828 Degrees of Freedom = 21913.4216
Independence AIC = 22057.4216
Model AIC = 1921.4537
Saturated AIC = 1800.0000
Independence CAIC = 22460.7102
Model CAIC = 2694.4234
Saturated CAIC = 6841.1071
Normed Fit Index (NFI) = 0.9249
Non-Normed Fit Index (NNFI) = 0.9618
Parsimony Normed Fit Index (PNFI) = 0.9316
Comparative Fit Index (CFI) = 0.9615
Incremental Fit Index (IFI) = 0.9615
Relative Fit Index (RFI) = 0.9255
Critical N (CN) = 416.1573
Contribution to Chi-Square = 636.1025
Percentage Contribution to Chi-Square = 28.6479
Root Mean Square Residual (RMR) = 0.1042
Standardized RMR = 0.08371
Goodness of Fit Index (GFI) = 0.8312

APPENDIX D: GOODNESS OF FIT STATISTICS FOR THE PCQ-24 MULTIGROUP PARTIAL STRONG INVARIANCE MEASUREMENT MODEL

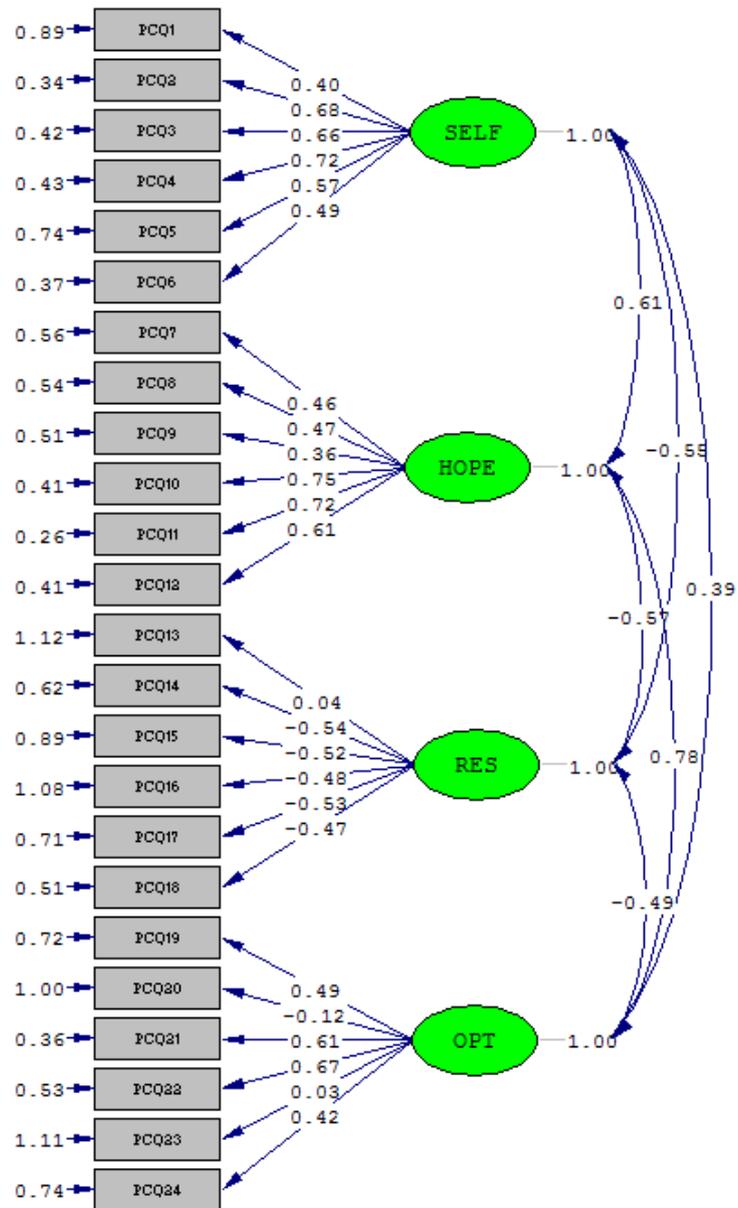
Degrees of Freedom = 821
 Minimum Fit Function Chi-Square = 2010.2415 (P = 0.0)
 Normal Theory Weighted Least Squares Chi-Square = 1998.8636 (P = 0.0)
 Satorra-Bentler Scaled Chi-Square = 1419.8166 (P = 0.0)
 Estimated Non-centrality Parameter (NCP) = 598.8166
 90 Percent Confidence Interval for NCP = (498.1981 ; 707.2840)
 Minimum Fit Function Value = 2.7425
 Population Discrepancy Function Value (F0) = 0.8169
 90 Percent Confidence Interval for F0 = (0.6797 ; 0.9649)
 Root Mean Square Error of Approximation (RMSEA) = 0.05464
 90 Percent Confidence Interval for RMSEA = (0.04984 ; 0.05938)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.05590
 Expected Cross-Validation Index (ECVI) = 2.3490
 90 Percent Confidence Interval for ECVI = (2.1135 ; 2.3988)
 ECVI for Saturated Model = 0.8186
 ECVI for Independence Model = 29.9610
 Chi-Square for Independence Model with 828 Degrees of Freedom = 21913.4216
 Independence AIC = 22057.4216
 Model AIC = 1721.8166
 Saturated AIC = 1800.0000
 Independence CAIC = 22460.7102
 Model CAIC = 2567.6023
 Saturated CAIC = 6841.1071
 Normed Fit Index (NFI) = 0.9352
 Non-Normed Fit Index (NNFI) = 0.9714
 Parsimony Normed Fit Index (PNFI) = 0.9273
 Comparative Fit Index (CFI) = 0.9716
 Incremental Fit Index (IFI) = 0.9716
 Relative Fit Index (RFI) = 0.9347
 Critical N (CN) = 475.0385
 Contribution to Chi-Square = 628.7221
 Percentage Contribution to Chi-Square = 31.2759
 Root Mean Square Residual (RMR) = 0.1006
 Standardized RMR = 0.08388
 Goodness of Fit Index (GFI) = 0.8347

APPENDIX E: GOODNESS OF FIT STATISTICS FOR THE PCQ-24 MULTIGROUP STRICT INVARIANCE**MEASUREMENT MODEL**

Degrees of Freedom = 840
 Minimum Fit Function Chi-Square = 2130.2628 (P = 0.0)
 Normal Theory Weighted Least Squares Chi-Square = 2141.3385 (P = 0.0)
 Satorra-Bentler Scaled Chi-Square = 1428.0642 (P = 0.0)
 Estimated Non-centrality Parameter (NCP) = 588.0642
 90 Percent Confidence Interval for NCP = (487.5168 ; 696.4724)
 Minimum Fit Function Value = 2.9062
 Population Discrepancy Function Value (F0) = 0.8023
 90 Percent Confidence Interval for F0 = (0.6651 ; 0.9502)
 Root Mean Square Error of Approximation (RMSEA) = 0.05353
 90 Percent Confidence Interval for RMSEA = (0.04874 ; 0.05825)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.1111
 Expected Cross-Validation Index (ECVI) = 2.3084
 90 Percent Confidence Interval for ECVI = (2.0730 ; 2.3581)
 ECVI for Saturated Model = 0.8186
 ECVI for Independence Model = 29.9610
 Chi-Square for Independence Model with 828 Degrees of Freedom = 21913.4216
 Independence AIC = 22057.4216
 Model AIC = 1692.0642
 Saturated AIC = 1800.0000
 Independence CAIC = 22460.7102
 Model CAIC = 2431.4266
 Saturated CAIC = 6841.1071
 Normed Fit Index (NFI) = 0.9348
 Non-Normed Fit Index (NNFI) = 0.9725
 Parsimony Normed Fit Index (PNFI) = 0.9484
 Comparative Fit Index (CFI) = 0.9721
 Incremental Fit Index (IFI) = 0.9721
 Relative Fit Index (RFI) = 0.9358
 Critical N (CN) = 482.6103
 Contribution to Chi-Square = 657.6034
 Percentage Contribution to Chi-Square = 30.8696
 Root Mean Square Residual (RMR) = 0.1016
 Standardized RMR = 0.08122
 Goodness of Fit Index (GFI) = 0.8169

APPENDIX F: PATH DIAGRAM FOR THE CONFIGURAL INVARIANCE MEASUREMENT MODEL: BLACK

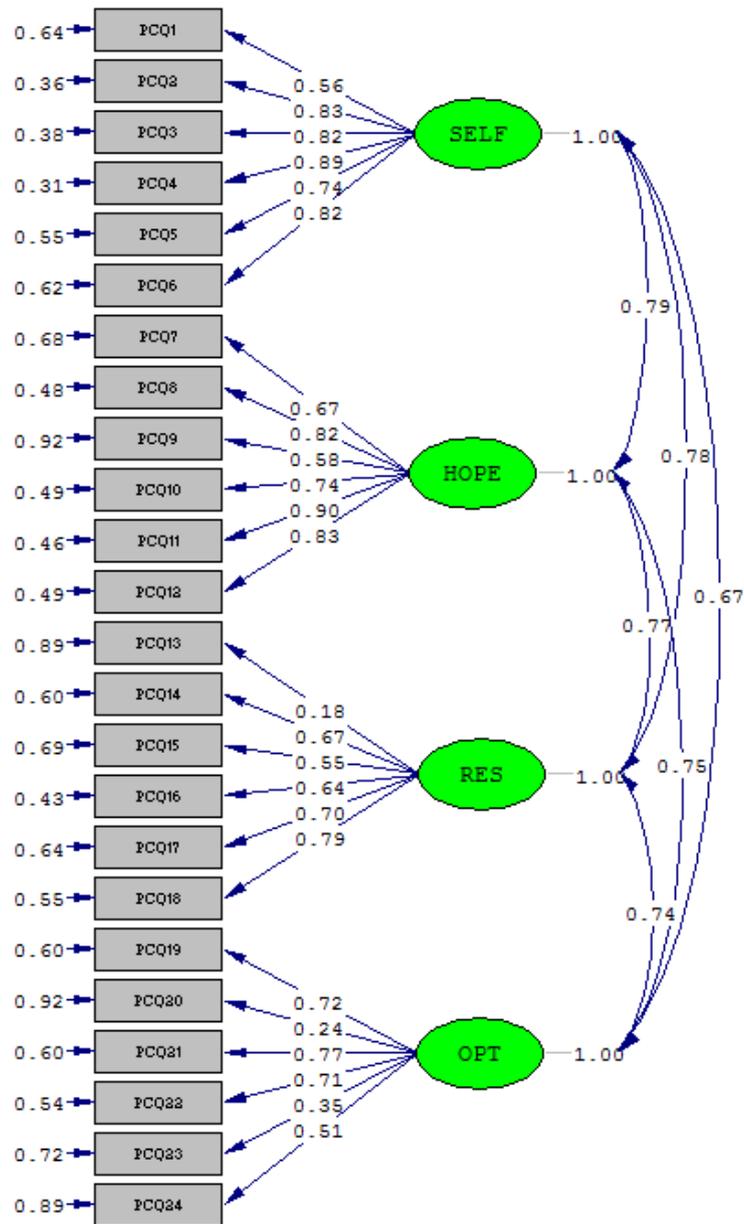
GROUP



Chi-Square=1251.34, df=738, P-value=0.00000, RMSEA=0.053

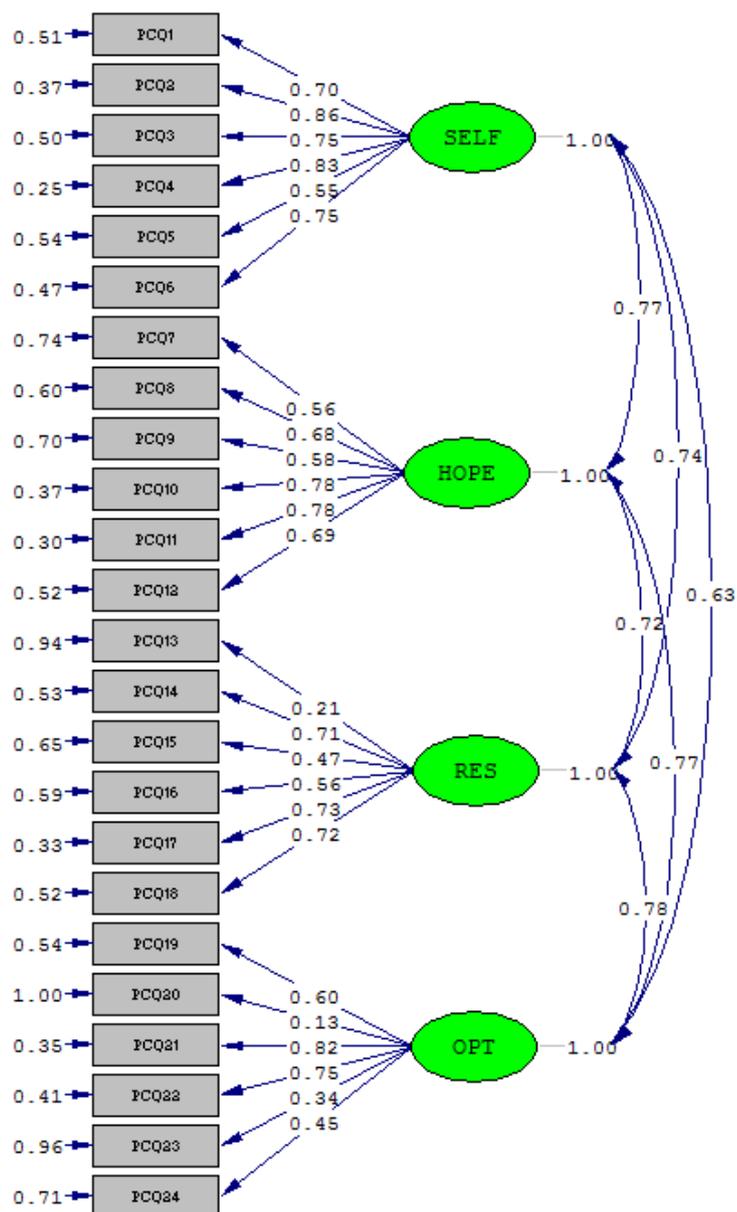
APPENDIX G: PATH DIAGRAM FOR THE CONFIGURAL INVARIANCE MEASUREMENT MODEL: WHITE

GROUP



Chi-Square=1251.34, df=738, P-value=0.00000, RMSEA=0.053

**APPENDIX H: PATH DIAGRAM FOR THE CONFIGURAL INVARIANCE MEASUREMENT MODEL:
COLOURED GROUP**



Chi-Square=1251.34, df=738, P-value=0.00000, RMSEA=0.053

**APPENDIX I: LISREL SYNTAX FOR THE CLASSICALLY PARALLEL PSYCAP CONFIGURAL INVARIANCE
MULTIGROUP MEASUREMENT MODEL**

```

GROUP WHITE
TI CLASSICALLY PARALLEL PSYCAP CONFIGURAL INVARIANCE MULTIGROUP MEASUREMENT
MODEL
DA NI=4 NO=272 NG=3 MA=CM
RA='C:\LISREL88_NICCI\WHITE_DIM.psf' NG=3
AC FI=C:\LISREL88_NICCI\WHITE_DIM.ACM
MO NX=4 NK=4 TD=DI TX=FI KA=FI
LK
SELF HOPE RES OPT
FR LX(1,1) LX(2,2) LX(3,3) LX(4,4)
FR TX(1) TX(2) TX(3) TX(4)
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4)
EQ LX(4,4) LX(1,1) LX(2,2) LX(3,3)
EQ TX(1) TX(2) TX(3) TX(4)
EQ TD(1,1) TD(2,2) TD(3,3) TD(4,4)
PD
OU RS SS SC IT=900 AD=900 ND=4

```

```

GROUP BLACK
DA NI=4 NO=218 NG=3 MA=CM
RA='C:\LISREL88_NICCI\BLACK_DIM.psf' NG=3
AC FI=C:\LISREL88_NICCI\BLACK_DIM.ACM
MO NX=4 NK=4 LX=IN PH=IN TD=DI TX=IN KA=IN
LK
SELF HOPE RES OPT
FR LX(1,1) LX(2,2) LX(3,3) LX(4,4)
FR TX(1) TX(2) TX(3) TX(4)
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4)
FR PHI(2,1) PHI(3,1) PHI(4,1) PHI(3,2) PHI(4,2) PHI(4,3)
EQ LX(4,4) LX(1,1) LX(2,2) LX(3,3)
EQ TX(1) TX(2) TX(3) TX(4)
EQ TD(1,1) TD(2,2) TD(3,3) TD(4,4)
OU

```

```

GROUP COLOURED
DA NI=4 NO=246 NG=3 MA=CM
RA='C:\LISREL88_NICCI\COLOURED_DIM.psf' NG=3
AC FI=C:\LISREL88_NICCI\COLOURED_DIM.ACM
MO NX=4 NK=4 LX=IN PH=IN TD=DI TX=IN KA=IN
LK
SELF HOPE RES OPT
FR LX(1,1) LX(2,2) LX(3,3) LX(4,4)
FR TX(1) TX(2) TX(3) TX(4)
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4)
FR PHI(2,1) PHI(3,1) PHI(4,1) PHI(3,2) PHI(4,2) PHI(4,3)
EQ LX(4,4) LX(1,1) LX(2,2) LX(3,3)
EQ TX(1) TX(2) TX(3) TX(4)
EQ TD(1,1) TD(2,2) TD(3,3) TD(4,4)
OU

```

APPENDIX J: ETHICAL CLEARANCE LETTER (REC)**NOTICE OF APPROVAL**

REC: Social, Behavioural and Education Research (SBER) - Initial Application Form

13 November 2019

Project number: 10803

Project Title: Investigation into the Measurement Bias of the PCQ-24 in South Africa

Dear Miss Nichale Esterhuizen

Your REC: Social, Behavioural and Education Research (SBER) - Initial Application Form submitted on 29 October 2019 was reviewed and approved by the REC: Humanities.

Please note the following for your approved submission:

Ethics approval period:

Protocol approval date (Humanities)	Protocol expiration date (Humanities)
13 November 2019	12 November 2022

GENERAL COMMENTS:

Please take note of the General Investigator Responsibilities attached to this letter. You may commence with your research after complying fully with these guidelines.

If the researcher deviates in any way from the proposal approved by the REC: Humanities, the researcher must notify the REC of these changes.

Please use your SU project number (10803) on any documents or correspondence with the REC concerning your project.

Please note that the REC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

FOR CONTINUATION OF PROJECTS AFTER REC APPROVAL PERIOD

Please note that a progress report should be submitted to the Research Ethics Committee: Humanities before the approval period has expired if a continuation of ethics approval is required. The Committee will then consider the continuation of the project for a further year (if necessary)

Included Documents:

Document Type	File Name	Date	Version
Default	G Gorgens-Ekermans CV April 2019	27/06/2019	Final
Research Protocol/Proposal	Research Proposal Nicci Esterhuizen Final	30/06/2019	V1
Proof of Ethics Clearance	ETHICAL CLEARANCE SHAYNE ROUX	29/07/2019	V1
Proof of Ethics Clearance	ETHICAL CLEARANCE MARITSA BOERS	29/07/2019	V1
Proof of Ethics Clearance	ETHICAL CLEARANCE JESSICA PRINSLOO	30/07/2019	V1
Proof of Ethics Clearance	ETIESE KLARING M HERBERT [ENGELS]	30/07/2019	V1
Proof of permission	Written permission Jessica Prinsloo	30/07/2019	V1
Proof of permission	Written permission Maritsa Smith (Boers)	30/07/2019	V1
Default	Written permission Shayne Roux	30/07/2019	V1
Default	Informed Consent - Shayne Roux	30/07/2019	V1
Default	Informed Consent - Maritsa Boers	30/07/2019	V1
Default	INFORMED CONSENT JESSICA PRINSLOO	30/07/2019	V1

Proof of permission	Written permission - Marthine Herbert	31/07/2019	V1
Default	Permission to use data - Professor Gina Gorgens	31/07/2019	V1
Proof of permission	Written permission Shayne Roux	01/10/2019	V2
Default	Response letter NE 23.10.2019	23/10/2019	V1

Investigator Responsibilities

Protection of Human Research Participants

Some of the general responsibilities investigators have when conducting research involving human participants are listed below:

1. Conducting the Research. You are responsible for making sure that the research is conducted according to the REC approved research protocol. You are also responsible for the actions of all your co-investigators and research staff involved with this research. You must also ensure that the research is conducted within the standards of your field of research.

2. Participant Enrollment. You may not recruit or enroll participants prior to the REC approval date or after the expiration date of REC approval. All recruitment materials for any form of media must be approved by the REC prior to their use.

3. Informed Consent. You are responsible for obtaining and documenting effective informed consent using **only** the REC-approved consent documents/process, and for ensuring that no human participants are involved in research prior to obtaining their informed consent. Please give all participants copies of the signed informed consent documents. Keep the originals in your secured research files for at least five (5) years.

4. Continuing Review. The REC must review and approve all REC-approved research proposals at intervals appropriate to the degree of risk but not less than once per year. There is **no grace period**. Prior to the date on which the REC approval of the research expires, **it is your responsibility to submit the progress report in a timely fashion to ensure a lapse in REC approval does not occur**. If REC approval of your research lapses, you must stop new participant enrollment, and contact the REC office immediately.

5. Amendments and Changes. If you wish to amend or change any aspect of your research (such as research design, interventions or procedures, participant population, informed consent document, instruments, surveys or recruiting material), you must submit the amendment to the REC for review using the current Amendment Form. You **may not initiate** any amendments or changes to your research without first obtaining written REC review and approval. The **only exception** is when it is necessary to eliminate apparent immediate hazards to participants and the REC should be immediately informed of this necessity.

6. Adverse or Unanticipated Events. Any serious adverse events, participant complaints, and all unanticipated problems that involve risks to participants or others, as well as any research related injuries, occurring at this institution or at other performance sites must be reported to Malene Fouche within **five (5) days** of discovery of the incident. You must also report any instances of serious or continuing problems, or non-compliance with the REC's requirements for protecting human research participants. The only exception to this policy is that the death of a research participant must be reported in accordance with the Stellenbosch University Research Ethics Committee Standard Operating Procedures. All reportable events should be submitted to the REC using the Serious Adverse Event Report Form.

7. Research Record Keeping. You must keep the following research related records, at a minimum, in a secure location for a minimum of five years: the REC approved research proposal and all amendments; all informed consent documents; recruiting materials; continuing review reports; adverse or unanticipated events; and all correspondence from the REC

8. Provision of Counselling or emergency support. When a dedicated counsellor or psychologist provides support to a participant without prior REC review and approval, to the extent permitted by law, such activities will not be recognised as research nor the data used in support of research. Such cases should be indicated in the progress report or final report.

9. Final reports. When you have completed (no further participant enrollment, interactions or interventions) or stopped work on your research, you must submit a Final Report to the REC.

10. On-Site Evaluations, Inspections, or Audits. If you are notified that your research will be reviewed or audited by the sponsor or any other external agency or any internal group, you must inform the REC immediately of the impending audit/evaluation.

