Testing for measurement invariance for employee engagement across different demographic groups in South Africa

M DU PLESSIS

Department of Human Resource Management, University of South Africa

vannim@unisa.ac.za

N MARTINS *

Department of Industrial and Organisational Psychology, University of South Africa

martin@unisa.ac.za * *corresponding author*

Abstract

The purpose of this study was to assess the measurement invariance of the employee engagement instrument (EEI), as a factor of race and gender. Previous research has advised that existing engagement instruments cannot be used with confidence to make comparisons between different demographic groups in South Africa. Given the demographic differences between individuals, the focus of this study was to determine whether the instrument could be used with confidence among the mentioned demographic groups.

A quantitative research approach was followed in this study. The EEI was administered electronically to a sample of business people, who were representative of the South African working population. 4 099 completed questionnaires were received. Differential item functioning (DIF) was used to test for measurement invariance within an item response theory (IRT) framework.

The results revealed that the assumption for metric invariance does not hold across all six sub-scales. It was further found that the DIF for race is less problematic than for gender. Eight items demonstrated DIF across gender. The Organisational Commitment, Immediate Manager and Strategy and Implementation sub-scales demonstrated the highest degree of invariance for race.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 24

1.   INTRODUCTION

Employee engagement is a popular topic that continues to attract attention in modern management practices. Despite its popularity in the workplace, the concept is still highly fragmented and has little academic underpinning (Christian, Garza & Slaughter 2011:90; Macey & Schneider 2008:3-4; Nienaber & Martins 2014:485). Researchers (Goliath-Yarde & Roodt 2011:2; Van Rooy, Whitman, Hart and Caleo 2011:147; Viljevac, Cooper-Thomas & Saks 2012:3692) are further of the opinion that existing engagement instruments are poorly understood and fall short of identifying actionable insights and solutions.

Consequently, Nienaber and Martins (2014:404) identified the need to further clarify existing employee engagement theories and to develop and refine an instrument that measures employee engagement at both the individual and organisational level within a multicultural context. The results of their research accumulated in a validated and reliable instrument (Nienaber & Martins 2015b:418) that demonstrates measurement invariance across various sectors (Martins 2015:771) within the South African context.

Testing for measurement invariance is an important prerequisite for making meaningful comparisons between groups, especially within the South African context (Meiring, Van de Vijver, Rothmann & Barrick 2015:1; Moerdyk 2009:75). Researchers (Heyns & Rothmann 2016:84; Marais, Mostert & Rothmann 2009:7; Meiring, Van de Vijver, Rothmann & Barrick 2005:3) have reported that demographic differences, such as race and gender, affect the psychometric properties of instruments and should thus, be considered when standardising an instrument (Visser & Viviers 2010:7). Consequently, the purpose of this study was to assess the measurement invariance of the EEI as a factor of gender and race.

2.   CURRENT THEORETICAL PERSPECTIVES

2.1   Employee engagement and measurement

Employee engagement has become a popular topic in recent years and continues to attract attention in modern management practices (Albrecht, Bakker, Gruman, Macey & Saks 2015:7; Griffith 2009:7; Imandin 2015:23). Employee engagement is a critical driver of organisations' success (Macey & Scheider 2008:3; Markos & Sridevi 2010:90) because of

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 25

the role it plays in sustaining a competitive advantage, which could lead to improved business results and successful organisational performance (Nienaber & Martins 2014:485).

Studies have further linked employee engagement with a number of favourable organisational outcomes such as increased productivity, organisational commitment and loyalty, organisational citizenship behaviour, job satisfaction, customer satisfaction, reduced employee turnover and occupational accidents and improved health and wellness outcomes (Nienaber & Martins 2015a:18-19; Viljevac *et al.* 2012:3692).

Griffith (2009:7) consequently labels employee engagement as the "human resource craze", "the silver bullet" and a "magical formula" to enhance employee performance. As a result, employers want engaged, happier, healthier and more fulfilled employees who deliver improved business results and are associated with successful, high performing organisations.

Despite its popularity in the workplace, a precise definition of employee engagement remains elusive because of continued research and redefinition surrounding the topic (Albrecht *et al.* 2015:8; Imandin 2015:23). This confusion is further complicated by the misuse of the terms "work engagement" and "employee engagement" (Nienaber & Martins 2016:3). According to Schaufeli and Salanova (2011:42) employee engagement is a broader concept than work engagement.

Employee engagement is conceptualised by Kahn (1990:694) as a positive psychological state that consists of cognitive, emotional and behavioural dimensions; and consists of two district, yet related types, namely, job engagement and organisational engagement. Job engagement is conceptualised as the extent to which an employee is psychologically present in the performance of his/her work roles; whilst organisational engagement, on the contrary, is defined as the extent to which the employee is psychologically present in their role as an active member of the organisation (Saks 2008:41).

In contrast, work engagement is the preferred concept used by academics because of its focus on the relationships employees have with their work activities. It is defined as a positive, fulfilling, work-related state of mind that is characterised by vigour, dedication and absorption (Bakker & Demerouti 2008:209; Saks 2008:42). It captures how employees experience their work as stimulating and energetic and something to which they really want

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 26

to devote time and effort; as a significant and meaningful pursuit and as engrossing and something on which they are fully concentrated (Nienaber & Martins 2015a:4). Engaged employees have high levels of energy and are enthusiastic and often fully immersed in their work.

Regarding the measurement of work engagement, Schaufeli, Salanova, González Romá and Bakker (2002) developed the self-report Utrecht Work Engagement Scale (UWES). Vigour, dedication and absorption are assessed by six, five and six items respectively. This 17-item scale, also known as the UWES-17, has been validated and utilised in a number of countries, including South Africa. Findings have shown that the UWES and more specifically, the three engagement scales, have sufficient internal consistencies, ranging between 0.80 and 0.90 (Schaufeli *et al.* 2002). However, when work engagement measures are applied to different cultural groups, issues of measurement bias and equivalence become important (Storm & Rothmann 2003:64).

2.2   Measurement invariance

Measurement invariance is a psychometric method that tests whether the properties of an instrument varies across groups (Moore, Neale, Silberg & Verhulst 2016:2). Kimber, Rehm and Ferro (2015:3) further explain that measurement invariance is a prerequisite for making meaningful comparisons between groups, because according to Milfont and Fischer (2010:12), researchers often assume that an instrument measures the construct in all groups and this assumption cannot be justified without invariance testing.

Testing for measurement invariance is a particularly important prerequisite within the South African context because of its diverse composition (Meiring *et al.* 2015:1; Moerdyk 2009). Studies conducted within a South African context have reported that issues of race, education and language are the main factors that impact on the construct and item comparability of psychometric tests (Meiring *et al.* 2005:3).

Marais *et al.* (2009:7), and Heyns and Rothmann (2016:84) further explain that the lack of measurement invariance within a South African context can be attributed to (1) semantic differences, (2) cultural differences, (3) personality differences, (4) socio-economic differences and (5) stereotyping. Consequently, Moerdyk (2009:11) and Visser and Viviers (2010:7) stress the importance of considering individual differences in psychological

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 27

assessment. This statement is confirmed in South African validation studies that have determined measurement invariance of the UWES.

Although the three-factor structure of the UWES was confirmed, with suitable internal consistency (Coetzee & Rothmann 2005:30; Storm & Rothmann 2003:67-68), significant differences were found in the engagement levels of employees in different demographic groups (e.g. language, culture and education) (Barkhuizen & Rothmann 2006:44; Coetzee & Rothmann 2005:30; Goliath-Yarde & Roodt 2011:3). Goliath-Yarde and Roodt (2011:10) further suggests that the UWES-17 should not be used to compare different culture groups in South Africa, and is of the opinion that the differences relate mostly to language and education differences.

3.   PROBLEM STATEMENT

From the discussion above, it is evident that the concept of employee engagement is still highly fragmented and has little academic underpinning (Christian *et al.* 2011:90; Macey & Schneider 2008:3-4; Nienaber & Martins 2014:485). Consequently, Simpson (2009), Van Rooy *et al.* (2011) and Viljevac *et al.* (2012:3692) found that existing engagement measures (e.g. UWES) are poorly understood and fall short of identifying actionable insights and solutions. It was further suggested that the UWES not be administered within a South African context because of the country's diverse composition (Goliath-Yarde & Roodt 2011:2).

In their research, Nienaber and Martins (2014:494) moreover found that most engagement instruments are designed to measure engagement at either the individual or organisational level, and there is no consensus regarding the dimensions comprising engagement. Based on these findings, Nienaber and Martins (2014:494) identified the need to further clarify existing theories and to develop and refine a measurement instrument to measure employee engagement. The results of their research accumulated in the development of a newly validated and reliable instrument measuring employee engagement concurrently at both the individual and organisational level within a diverse, multicultural context (Nienaber & Martins 2015b:419).

Consequently, Martins (2015:759) conceptualised employee engagement as "engaged employees at both the individual and organisational level, who are fully absorbed by and

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 28

enthusiastic about their work, and so take positive action to further the organisation's reputation and interests". In essence, employee engagement was conceptualised at both the individual and organisational level, because it reflects the individual's role in his/her work and as a member of the organisation.

The instrument consists of two sections, one collecting demographic information (i.e. gender, qualification, experience and tenure), and one soliciting responses using a five-point Likert scale, on statements about engagement at the individual, team/departmental and organisational level. The instrument has been shown to have good psychometric properties (Nienaber & Martins 2015b:420) but indicated that multi-group invariance could not be assumed for the different sectors (Martins 2015:772). The next step in the refining of the instrument is to further test for group invariance.

Consequently, the purpose of this research study is to assess the employee engagement instrument with specific focus on measurement invariance for gender and race within a South African context.

4.   RESEARCH OBJECTIVE

The objective of this research study was to determine if any measurement invariance (measurement equivalence) exists between the different races and genders within a South African context.

5.   RESEARCH METHODOLOGY

5.1   Research participants

A research company's database of 285 000 working individuals, who represent the profile of the South African working population, was used in this study. Ethical clearance was obtained to conduct the research study and to administer the questionnaire. An electronic survey, administered by iFeedback.co.za, an online data collection portal, was used by means of a mass e-mail invitation over a period of three weeks.

Each potential participant in the database of 285 000 working individuals received an e-mail stating the purpose of the study, information regarding the completion time and a link to the electronic survey. The e-mail further explained that participation was voluntary and that

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 29

confidentiality and anonymity would be ensured. Lastly, the participants were ensured that the responses would be used for research purposes only. A total of 4 099 completed questionnaires were received in the third national engagement survey conducted. The researchers aimed to obtain at least 4 000 responses.

The demographic profile of the participants in terms of race and gender is reflected in table 1.

TABLE 1:   Demographic profile of the respondents

Item	Category	Frequency	Percentage
Race	African	778	19.0%
	Coloured	329	8.0%
	Indian	338	8.3%
	White	2,543	62.0%
	Other	29	0.7%
	Prefer not to say	82	2.0%
Gender	Male	2,387	58.2%
	Female	1,712	41.8%

Source: Calculated from survey results

The sample was comprised of 58.2% male and 41.8% female participants, and 62.0% of the sample was white participants.

Demographic items received had very few missing responses and therefore, these were treated as system-missing.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 30

5.2 The measuring instrument

The employee engagement instrument (EEI), was administered to the population described above (section 3). The validity and reliability of the instrument was reported in the second phase of the scale development process (Nienaber & Martins 2015b:420).

The instrument consists of two sections, namely, one collecting biographical/demographic information and one soliciting responses using a five-point Likert scale (1 = strongly disagree; 5 = strongly agree), on statements about engagement at the individual (50 items), team (12 items) and organisational (10 items) level.

5.3 Research approach and procedure

A quantitative research approach was followed in this study. This approach to research involves collecting data in numerical form for quantitative analysis (Garwood 2006; Lyons & Doueck 2009). Recent published analyses aimed to demonstrate the instrument's degree of factorial invariance across South African business sectors, with results indicating that invariance can be assumed for all sectors except community and manufacturing sectors (Martins 2015:772).

Thus far, the development and validation of the EEI has adopted a classical test theory (CTT) or confirmatory factor analysis (CFA) approach (Martins 2016:185).

In order to further unpack the instrument's factorial invariance, the researchers decided to assess the EEI from an item response theory (IRT) perceptive, with a specific focus on measurement invariance focusing on gender and race. By using IRT, the difficulty level and discriminatory power of an item can be accurately determined (Foxcroft & Roodt 2009:73). Thus, the item parameters are independent of the test taker's ability level and are therefore, a useful method for determining item bias (De Beer 2005:723). The principle of invariance is thus a major advantage of IRT, because it allows researchers to compare traits of individuals from different populations (Bortolotti, Tezza, De Andrade, Bornia & De Sousa Junior 2013:2344-2345).

The method that was used to establish measurement invariance within the IRT framework is known as differential item functioning (DIF) (Foxcroft & Roodt 2009:73). DIF is conceptualised as the difference in item scores between two or more groups that match the

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 31

concept of interest (Zumbo 1999:12). DIF is thus used to identify items that may be biased or unfair to test-takers from certain groups. DIF is focused on the functioning of items *across* groups. This is not to suggest that test/scale score equivalence is unimportant, as one can determine the extent to which DIF items cumulatively impact observed mean differences across groups.

A bottom-up and progressively restricted approach was thus followed in this research to test for measurement invariance (refer to figure 1). That is, the invariance requirements at each step had to be met before moving on to the higher-order form of equivalence (Dimitrov 2010:124). The steps that were followed to determine measurement invariance in this study are graphically represented in figure 2 and explained in the remainder of this section.

Undimensional IRT analysis

Analysing items per scale according to IRT model to determine factor loadings and goodness-of-fit. This can be considered the IRT equivalent of a restricted IRT.

Unrestricted MIRT

Analysing all items together to determine factor loadings and goodness-of-fit. This can be considered the IRT equivalent of an unrestricted EFA.

* MIRT – Multidimensional IRT

Restricted MIRT

Analysing all items together, but pre-specifying the factor structure by constraining the parameters of certain items on certain factors. This can be considered the IRT equivalent of CFA.

FIGURE 1:   Research process: introductory analysis

Source: Martins 2016:186

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 32
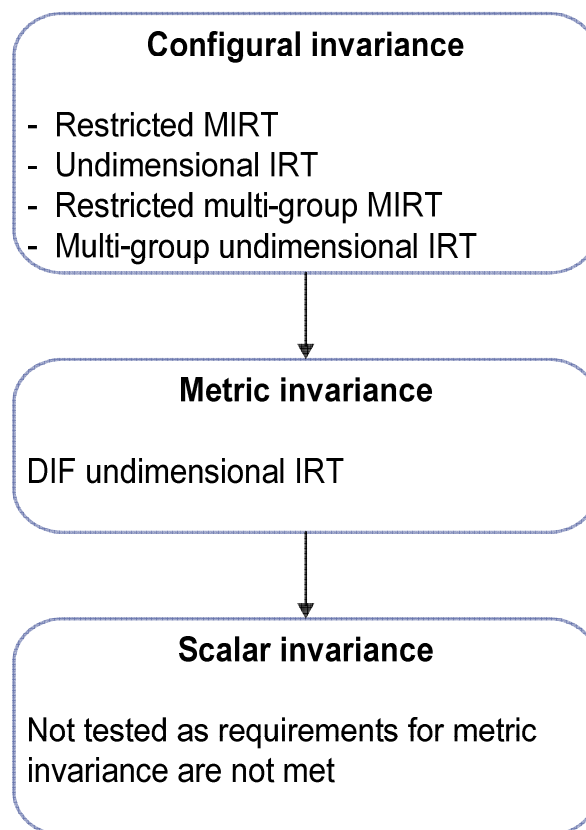
FIGURE 2:   Measurement invariance: an IRT approach

Source: Authors' compilation

5.3.1   Configural invariance

The first test of measurement invariance tests is to determine whether configuration holds across groups (Hortensius 2012:8). If a model exhibits configural non-invariance, it is not possible to compare across groups. To demonstrate configural invariance, the researcher has to show that a factor is unidimensional across groups. Finding unidimensionality across groups implies that there is configural measurement invariance (Tay, Meade & Cao 2015:8). Fundamentally, one would have to prove local independence by establishing that observed item responses are uncorrelated after controlling for the underlying construct/trait.

Configural invariance was established by conducting the following analysis (Martins 2016:186):

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 33

- A *restricted multidimensional item response theory (MIRT) and/or unidimensional model (CFA equivalent)* was conducted on each group within each DIF variable (e.g. race and gender). In other words, the overall sample was divided into two, three or four depending on the number of groups. In the case of gender, separate models were conducted for males and females, and factor loadings were examined to assess if there were any significant differences.

- Thereafter, a *restricted multi-group MIRT model* was conducted. All groups within the same model were analysed to determine if differences in terms of factor loadings and means existed. It is important to note that only item loadings on designated factors were freely estimated similar to multi-factor CFA analysis. This output assists in determining whether form equivalence exists for the total model of engagement.

- In order to identify specific areas of non-invariance, each sub-scale was analysed separately to examine differences that were flagged in the previous step. Unidimensional multi-group analysis also provided item-level statistics to assess the requirement of local independence.

The results of the above analysis were collated to assess whether the data met the requirements of configural invariance. Thereafter, metric invariance was assessed via a DIF analysis.

5.3.2   Metric invariance

Metric invariance is the next step of equivalence and requires that the size of the factor loadings be equivalent across groups under consideration. This model therefore tests if different groups respond to items in the same way (Milfont & Fischer 2010:115). If metric invariance is satisfied, obtained ratings can be compared across groups and observed item differences will indicate group differences in the underlying construct (Milfont & Fischer 2010:115).

Within IRT, this is assessed via DIF in order to determine whether items function differently across groups. To assess this requirement, DIF analyses were conducted at a unidimensional level to determine goodness-of-fit and item functioning. The item slopes for each sub-scale/model were set equal to the assumption of equivalent loadings/slopes.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 34

M DU PLESSIS
N MARTINS

Testing for measurement invardiance for
employee engagement across different
demographic groups in South Africa

5.3.3   Scalar invariance

Scalar invariance is required to compare means. Scalar invariance indicates that observed scores are related to the latent scores. That is, individuals who have the same score on the latent construct would obtain the same score on the observed variable regardless of their group membership (Milfont & Fischer 2010:115). Scalar invariance assumes that the size of intercepts is equivalent across groups under consideration. This is assessed by conducting a DIF analysis and constraining both the slopes and item intercepts in order to test the assumption(s).

6.   RESULTS

The previous analysis of the data provided strong evidence of the factor analytical structure of the measure (Martins 2015b:763). In order to address the objective of this study, the results of the measurement invariance testing of the EEI in relation to gender and race is reported in this article.

6.1   Gender

6.1.1   Unidimensional analyses

Males and females were analysed in separate models on a unidimensional level. This was done in order to obtain a more detailed picture of the item functioning within each sub-scale. Table 2 provides a summary of the goodness-of-fit indices for each model. Two analysis were conducted per sub-scale – one for males and one for females.

Results of the unidimensional analysis show a difference in male and female models in terms of -2loglikelihood, AIC and BIC. Overall, Customer Service, Immediate Manager and Organisational Commitment show better model fit in terms of the aforementioned indices. However, these may be affected by the difference in sample size, and therefore, M_2 and RMSEA should be used to provide a better indication of model fit.

As mentioned, a significant M_2 value (as seen in table 2) demonstrates poor model fit.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 35

TABLE 2:   Comparing unidimensional models: male and female

Analysis	Sample size	-2loglikelihood	AIC	BIC	M2	d.f.	RMSEA
Customer service – Male	2 387	33346.85	33406.85	33580.19	1315.82***	234	0.04
Customer service – Female	1 712	23971.78	24031.78	24195.15	1015.61***	234	0.04
Immediate manager – Male	2 387	34939.75	35009.75	35211.97	4334.33***	329	0.07
Immediate manager – Female	1 712	25456.89	25526.89	25717.48	3361.32***	329	0.07
Organisational commitment – Male	2 387	30888.29	30948.29	31121.62	2084.16***	234	0.06
Organisational commitment – Female	1 712	22407.47	22467.47	22630.83	2811.18***	234	0.08
Organisational satisfaction – Male	2 387	38399.28	38489.28	38749.28	5989.46***	567	0.06
Organisational satisfaction – Female	1 712	28110.09	28200.09	28445.13	6441.55***	567	0.08
Strategy & implementation – Male	2 387	54204.48	54304.48	54593.37	5253.58***	710	0.05
Strategy & implementation – Female	1 712	38921.36	39021.36	39293.63	4897.48***	710	0.06
Team – Male	2 387	47854.1	47974.10	48320.76	7063.97***	1044	0.05
Team – Female	1 712	33102.02	33222.02	33548.74	5926.86***	1044	0.05

*** - p<0.001; -2loglikelihood – Proxy for chi-squared; AIC - Akaike Information Criterion; BIC - Bayesian Information Criterion; M2 – Relative comparison between models to assess fit; d.f. – degrees of freedom; RMSEA - Root Mean Square Error of Approximation

Source: Calculated from survey results

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 36

6.1.2   Restricted MIRT

A restricted MIRT was analysed for males and females separately, in order to investigate whether there were any significant differences in the pattern and magnitude of factor loadings.

In each case, the results were reported both on the original sample and a randomised sample of equal size. The latter procedure was followed to mitigate any large changes in fit indices that may be due to unbalanced samples. Since the information indices (AIC and BIC) are largely used to compare nested model variance, it is not prudent to compare non-nested models directly. As such, the differences in these indices between the four models in Table 3, may not be very informative. However, the -2loglikelihood can be compared. Although the original male and female samples differ quite significantly based on the -2loglikelihood index, the balanced and random samples show small differences in overall model fit.

TABLE 3:   Goodness-of-fit comparison: male and female restricted MIRT

Analysis	Sample size	d.f.	-2loglikelihood	AIC	BIC
Restricted MIRT – Males only	2 387	265	230 690.05 $\pm$ 62.39	231 220.05 $\pm$ 62.39	232 751.16 $\pm$ 62.39
Restricted MIRT – Males randomised sample	1 000	265	96 311.10 $\pm$ 39.94	96 861.10 $\pm$ 39.94	98 161.65 $\pm$ 39.94
Restricted MIRT – Females only	1 712	265	166 551.49 $\pm$ 57.81	167 081.79 $\pm$ 57.81	168 524.82 $\pm$ 57.81
Restricted MIRT – Females randomised sample	1 000	265	97 953.18 $\pm$ 45.96	98 483.18 $\pm$ 45.96	99 783.73 $\pm$ 45.96

Source: Calculated from survey results

The results further demonstrate that factor loadings between males (n = 2 387) and females (n = 1 712) are similar across models, with loading differences not greater than 0.05.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 37

After analysing male and female models separately, both gender groups were combined into a multi-group restricted MIRT. A multi-group analyses both groups simultaneously and provides an indication of their factor loadings relative to one another.

In the multi-group analysis, males were treated as the reference group with the latent mean and variance constrained to 0 and 1 respectively, for each sub-scale. Females were treated as the focal group, with latent means and variances freely estimated.

The results of the restricted MIRT between male and female respondents in a multi-group model indicate factor loading differences between the gender groups. Factor loadings for females were generally 0.15 lower than those for males.

To better understand the multi-group effects seen in the MIRT model, separate multi-group analyses were conducted at a unidimensional level, with item parameters again being estimated via the Bock-Aitkin (BAEM) method (Bock & Aitkin 1981). A visual inspection of the factor loadings across the unidimensional and multidimensional analyses shows similarly strong factor loadings and low standard errors (+ 0.02 to + 0.04) between males and females

Table 4 provides a breakdown of S- χ^2 item fit statistics, where highly significant values ($p<0.01$) indicate that the expected values of the model deviate from the observed values of the data. Table 5 provides a breakdown of goodness-of-fit results.

TABLE 4:   Item diagnostic tables for multi-group unidimensional analysis

Item	Customer service			
	Male		Female	
	χ^2	d.f.	χ^2	d.f.
Q32	116.77***	64	114.16***	61
Q33	108.94***	61	100.45***	58
Q43	103.36***	58	96.36***	55
Q44	78.52***	52	121.49***	47

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 38

| Q55 | 129.03*** | 66 | 176.3*** | 61 |
| Q56 | 99.1*** | 59 | 141.66*** | 60 |

	Immediate manager			
Item	Male		Female	
	χ^2	d.f.	χ^2	d.f.
Q45	199.61***	53	149.9***	49
Q46	230.08***	51	159.92***	54
Q47	177.58***	57	148.99***	53
Q48	289.34***	75	279.41***	68
Q52	218.52***	65	172.1***	60
Q57	183.16***	71	149.97***	68
Q58	310.93***	76	247.64***	76

	Organisational commitment			
Item	Male		Female	
	χ^2	d.f.	χ^2	d.f.
Q14	196.59***	66	192.11***	63
Q15	191.51***	55	277.3***	51
Q16	285.05***	46	144.64***	44
Q17	368.01***	61	317.35***	60
Q18	150.68***	62	214.24***	59

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 39

Q19	107.79***	62	179.81***	54
Item	**Organisational satisfaction**			
	Male		**Female**	
	χ^2	d.f.	χ^2	d.f.
Q9	218.36***	74	142.8***	73
Q10	161.94***	71	147.55***	69
Q11	89.08*	63	100.68***	53
Q12	107.24***	62	82.88*	57
Q13	80.37	67	82.33*	59
Q20	130.01***	68	127.29***	66
Q23	264.17***	81	187.82***	77
Q30	156.95***	72	129.68***	70
Q34	252.46***	105	166.59***	99

*= $p<0.05$; **= $p<0.01$; ***= $p<0.001$

Source: Calculated from survey results

TABLE 5:   Goodness-of-fit indices from multi-group unidimensional analyses

Analysis	-2loglike-lihood	AIC	BIC	M2	d.f.	RMSEA
Customer Service – Multi-group	57318.64	57442.64	57834.39	2305.16***	466	0.03

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 40

Immediate Manager – Multi-group	60396.64	60540.64	60995.57	7779.17***	656	0.05
Organisational Commitment – Multi-group	53295.76	53419.76	53811.5	4292.66***	466	0.04
Organisational Satisfaction – Multi-group	66509.37	66693.37	67274.68	12264.52***	1132	0.05
Strategy & Implementation – Multi-group	93125.84	93329.84	93974.33	9893.55***	1418	0.04
Team – Multi-group	80956.11	81200.11	81970.96	12933.98***	2086	0.04

***= $p<0.001$;

Source: Calculated from survey results

Based on the foregoing tables, it is clear that some of the individual items do not fit the data very well. That is, they are not very good items. What is pleasing to see though, is that the items that appear to be problematic are consistent across males and females. Nonetheless, these items should be considered for adaptation or deletion to improve the overall discrimination of the individual sub-scale and the overall assessment.

6.1.3   Configural invariance

A visual inspection of the factor loadings across the unidimensional and multidimensional analyses shows similarly strong factor loadings and low standard errors (± 0.02 to ± 0.04) between males and females. At a surface level, the questionnaire demonstrates configural invariance.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 41

However, detailed analysis of the goodness-of-fit indices (table 5) from the multi-group unidimensional models show adequate RMSEA values (close to 0), but highly significant M_2 metrics. This indicates a significant source of error in the model that has not been addressed.

Examining other relevant metrics from the output show significant deviation between expected and observed frequencies across both males and females (table 4). An examination of the underlying frequency table is necessary to determine the reason for this difference. An analysis of the local independence statistic (not represented here) also shows evidence of local dependence, indicating likely multidimensionality, and the presence of one or more additional latent variables that have not been modelled.

These issues need to be addressed before configural invariance can be proven. However, given that the factor loadings are similar across groups, it was decided to proceed to the next stage of invariance testing by running a unidimensional DIF analysis on each sub-scale.

6.1.4   Metric invariance

A differential item functioning (DIF) analysis is similar to a multi-group model, but specifically measures whether the expected item responses (in terms of scope and intercept) differ between the groups under consideration.

DIF analyses were conducted at a unidimensional level on each sub-scale. The item slopes were set to equal in order to test whether goodness-of-fit remained similar to the baseline model (freely estimated multi-group analysis). Table 6 provides a goodness-of-fit comparison between the multi-group and DIF unidimensional models. The relative measures (-2loglikelihood, AIC and BIC) show slight increases between the baseline and DIF models. Similarly, there were larger M_2 values for the more constrained models, with Immediate Manager, Organisational Satisfaction, and Strategy and Implementation showing a M_2 increase of more than 10.

The chi-squared difference test is also reported in table 6. This measure, annotated as χ^2 Diff, assesses the statistical significance of the difference between the DIF and baseline models for each sub-scale. It is calculated by obtaining the difference between the -2loglikelihood for the DIF model and the -2loglikelihood for the baseline model. This value is

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 42

then referenced to the chi-squared table, along with the associated degrees of freedom, to determine statistical significance. If a significant difference is found based on the -2loglikelihood, one can assume that assumption of metric invariance does not hold for the more constrained model. Since the -2loglikelihood metric can be interpreted as a chi-square distribution, the same problems associated with using chi-square for significance testing will plague the -2loglikelihood. For this reason, it is important to also consider the information indices like the AIC and BIC to detect differences between the two nested models.

TABLE 6:   Overall model fit comparing multi-group and DIF analyses (unidimensional)

Analysis	-2loglikeli-hood	AIC	BIC	M2	d.f.	RMSEA	χ^2 diff.
Customer Service Baseline	57318.64	57442.64	57834.39	2305.16***	466	0.03	204.7 ***
Customer Service DIF	57523.34	57623.34	57939.27	2577.43***	478	0.03	
Immediate Manager Baseline	60396.64	60540.64	60995.57	7779.17***	656	0.05	1004.2* **
Immediate Manager DIF	61400.88	61516.88	61883.36	8961.48***	670	0.05	
Organisational Commitment Baseline	53295.76	53419.76	53811.5	4292.66***	466	0.04	450.6** *
Organisational Commitment	53746.42	53846.42	54162.34	4746.57***	478	0.05	

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 43

M DU PLESSIS
N MARTINS

**Testing for measurement invariance for
employee engagement across different
demographic groups in South Africa**

Analysis	-2loglikeli-hood	AIC	BIC	M2	d.f.	RMSEA	χ^2 diff.
DIF							
Organisational Satisfaction Baseline	66509.37	66693.37	67274.68	12264.52***	1132	0.05	1207.1***
Organisational Satisfaction DIF	67716.54	67864.54	68332.11	13366.82***	1150	0.05	
Strategy & Implementation Baseline	93125.84	93329.84	93974.33	9893.55***	1418	0.04	1083.3***
Strategy & Implementation DIF	94209.2	94373.2	94891.32	12298.85***	1438	0.04	
Team Baseline	80956.11	81200.11	81970.96	12933.98***	2086	0.04	421.62***
Team DIF	81377.73	81573.73	82192.95	12799.21***	2110	0.04	

***= p<0.001; χ2 Diff. = Chi-squared difference test

Source: Calculated from survey results

In the case of gender, all six DIF models are shown to be significantly different to the baseline models, therefore rendering metric invariance untenable (weak).

Table 7 presents the DIF statistics for each sub-scale under consideration. Significant values of χ^2 indicate the presence of differential functioning across groups, only the problematic items are highlighted.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 44

TABLE 7:   DIF statistics

Male	Female	Total χ²	d.f.	p	χ²a	d.f.	p	χ²c|a	d.f.	p
Customer Service										
Q32	Q32	15.8	5	0.0075	0	1	0.9884	15.8	4	0.0033
Immediate Manager										
Q46	Q46	12.6	5	0.0278	0.1	1	0.7973	12.5	4	0.0140
Q47	Q47	10	5	0.0737	0.1	1	0.7973	10	4	0.0407
Organisational Commitment										
Q15	Q15	10.5	5	0.0622	0.3	1	0.5666	10.2	4	0.0376
Organisational Satisfaction										
Q30	Q30	12.9	5	0.0241	0.1	1	0.7039	12.8	4	0.0124
Strategy and Implementation										
Q29	Q29	13.2	5	0.0218	0.2	1	0.6979	13	4	0.0112
Team										
Q51	Q51	11.3	5	0.0459	0.1	1	0.7069	11.1	4	0.0249
Q54	Q54	12.3	5	0.0305	0.1	1	0.7069	12.2	4	0.016

Source: Calculated from survey results

Overall, the DIF analyses echo the findings of the multi-group undimensional analysis with signs of misfitting models and DIF across all six sub-scales. However, given the large number of items and the large sample size, only 8 items demonstrated significant DIF across gender. It thus appears as if these 8 items should either be removed or adapted. The subsequent section of the report will look at the measurement invariance of race.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 45

6.2   Race

This section focuses on the examination of group differences based on race. The breakdown of response frequencies was provided in table 2, with the majority of test-takers being of Caucasian origin.

For the purpose of IRT analysis "Other (4)" and "Prefer not to say (5)" were excluded due to their small sample size. Due to the unbalanced size of the "White (4)" sample, a randomised selection of 500 cases was instead utilised.

6.2.1   Unidimensional analyses

Unidimensional analyses were conducted on each group within the race variable to assess factor loadings and goodness-of-fit between said groups.

Table 8 provides a breakdown of the goodness-of-fit of the unidimensional models run on each racial group within each sub-scale. A total of 24 different models were analysed to account for the four ethnic categories and the six sub-scales.

TABLE 8:   Goodness-of-fit comparison: race unidimensional analyses

Customer Service (6 items)							
Group	**Sample Size**	**-2loglikelihood**	**AIC**	**BIC**	**M_2**	**d.f.**	**RMSEA**
African (1)	778	11103.95	11163.95	11303.65	730.57***	234	0.05
Coloured (2)	329	4608.21	4668.21	4782.09	501.44***	234	0.06
Indian (3)	338	4417.67	4477.67	4592.36	493.01***	234	0.06
White (4)	500	6891.61	6951.61	7078.05	355.59***	234	0.03
Immediate Manager (7 items)							
Group	**Sample**	**-2loglikelih**	**AIC**	**BIC**	**M_2**	**d.f.**	**RMSEA**

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 46

	Size	**ood**					
African (1)	778	11665.27	11735.27	11898.2	2120.6***	329	0.08
Coloured (2)	329	4836.82	4906.82	5039.69	1519.2***	329	0.1
Indian (3)	338	4915.87	4985.87	5119.67	1984.4***	329	0.12
White (4)	500	7237.7	7307.7	7455.21	1035.5***	329	0.07
Organisational Commitment (6 items)							
Group	**Sample Size**	**-2loglikelihood**	**AIC**	**BIC**	**M2**	**d.f.**	**RMSEA**
African (1)	778	10103.7	10163.7	10303.41	1110.5***	234	0.07
Coloured (2)	329	4311.02	4371.02	4484.9	654.55***	234	0.07
Indian (3)	338	4221.17	4281.17	4395.86	1111.7***	234	0.11
White (4)	500	6456.02	6516.02	6642.46	568.93***	234	0.05
Organisational Satisfaction (9 items)							
Group	**Sample Size**	**-2loglikelihood**	**AIC**	**BIC**	**M2**	**d.f.**	**RMSEA**
African (1)	778	13436.85	13526.85	13736.4	3273.68***	567	0.08
Coloured (2)	329	5587.78	5677.78	5848.61	2370.7***	567	0.1
Indian (3)	338	5083.57	5173.57	5345.61	2525.9***	567	0.1
White (4)	500	7958.44	8048.44	8238.1	1129.1***	567	0.04

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 47

M DU PLESSIS
N MARTINS

Testing for measurement invariance for
employee engagement across different
demographic groups in South Africa

Strategy and Implementation (10 items)							
Group	Sample Size	-2loglikelihood	AIC	BIC	M2	d.f.	RMSEA
African (1)	778	18211.11	18311.11	18543.94	2445.1***	710	0.06
Coloured (2)	329	7611.87	7711.87	7901.67	2304.3***	710	0.08
Indian (3)	338	7427.14	7527.14	7718.29	1688.6***	710	0.06
White (4)	500	11252.03	11352.03	11562.76	1846.8***	710	0.06
Team (12 items)							
Group	Sample Size	-2loglikelihood	AIC	BIC	M2	d.f.	RMSEA
African (1)	778	16446.15	16566.15	16845.55	3656.1***	1044	0.06
Coloured (2)	329	6521.33	6641.33	6869.09	2292.6***	1044	0.06
Indian (3)	338	6472.59	6592.59	6821.98	3036.5***	1044	0.08
White (4)	500	9420.43	9540.43	9793.3	3026.2***	1044	0.06

***= p<0.001

Source: Calculated from survey results

Similar to the results seen with gender (table 2), the various models all demonstrate an acceptable RMSEA model fit indicator. However, the M_2 statistic demonstrated that the theoretical configuration of the sub-scale may not replicate the sample data with an adequate degree of accuracy (p<0.001). As a whole, Team showed the highest degree of misfit according to the -2loglikelihood metric. In general, the African group demonstrated higher levels of misfit compared to the other three groups.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 48

6.2.2 Restricted MIRT

A race multi-group MIRT was conducted according to the same methodology described above. The African group was treated as the reference group with the latent mean and variance constrained to 0 and 1, respectively. The other three categories were treated as focal groups with means and variances freely estimated.

The results of the factor loadings for each group showed that the African group demonstrates the strongest factor loadings across each sub-scale. As a whole factor, loadings ranged from 0.54 to 0.95 with no aberrantly low loadings, and standard error ranged between 0.01 and 0.05. This instils confidence in the cross-cultural robustness of the measure.

Table 9 demonstrates the latent mean of the reference group (African) was constrained to zero for comparative purposes. Since all the latent mean values for the Coloured, Indian and White groups are positive, it would suggest that these groups have higher levels of endorsement on the six dimensions of engagement. It is important to remember that it is not prudent at this point to compare the latent mean scores of groups since measurement invariance has not yet been demonstrated. This will only be possible after the DIF analysis has been concluded.

TABLE 9: Multi-group restricted MIRT group parameter estimates: race

Label	Group	μ1	s.e.	μ2	s.e.	μ3	s.e.	μ4	s.e.	μ5	s.e.	μ6	s.e.
1	African	0	-----	0	-----	0	-----	0	-----	0	-----	0	-----
2	Coloured	0.48	0.08	0.26	0.08	0.4	0.07	0.49	0.06	0.23	0.06	0.76	0.06
3	Indian	0.54	0.08	0.59	0.07	0.7	0.07	0.81	0.08	0.29	0.06	0.92	0.07
4	White	0.36	0.04	0.24	0.04	0.52	0.07	0.54	0.05	0.21	0.05	0.8	0.03

μ1-6 – Latent mean variance; s.e. – standard error

Source: Calculated from survey results

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 49

M DU PLESSIS
N MARTINS

Testing for measurement invariance for
employee engagement across different
demographic groups in South Africa

Table 10 demonstrates the covariance between latent variables across the four race groups. As would be expect, most of the latent factors are positively correlated with one another. Unfortunately, no latent covariance can be examined for the African group since the latent means are constrained to zero.

TABLE 10:   Multi-group restricted MIRT latent variable variance-covariance matrix: race

θ_1	s.e.	θ_2	s.e.	θ_3	s.e.	θ_4	s.e.	θ_5	s.e.	θ_6	s.e.
Group 1: African											
1	-----										
0	-----	1	-----								
0	-----	0	-----	1	-----						
0	-----	0	-----	0	-----	1	-----				
0	-----	0	-----	0	-----	0	-----	1	-----		
0	-----	0	-----	0	-----	0	-----	0	-----	1	-----
Group 2: Coloured											
1.86	0.13										
1.41	0.15	2.2	0.16								
1.35	0.13	1.22	0.13	1.67	0.13						
1.18	0.11	1.1	0.12	1.27	0.11	1.32	0.11				
1.27	0.12	1.2	0.1	1.26	0.11	1.02	0.08	1.3	0.1		
1.13	0.11	1.03	0.11	0.7	0.07	0.69	0.08	0.66	0.08	1.2	0.1

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 50

M DU PLESSIS
N MARTINS

**Testing for measurement invariance for
employee engagement across different
demographic groups in South Africa**

Group 3: Indian											
θ_1	s.e.	θ_2	s.e.	θ_3	s.e.	θ_4	s.e.	θ_5	s.e.	θ_6	s.e.
2.02	0.19										
1.38	0.15	1.91	0.14								
1.48	0.13	1.22	0.12	1.62	0.12						
1.55	0.13	1.27	0.11	1.47	0.13	1.91	-----				
1.4	0.12	1.24	0.11	1.29	0.1	1.15	0.09	1.3	0.09		
1.45	0.13	1.02	0.09	0.98	0.11	1.19	0.08	0.95	0.08	1.54	0.11
Group 4: White											
θ_1	s.e.	θ_2	s.e.	θ_3	s.e.	θ_4	s.e.	θ_5	s.e.	θ_6	s.e.
1.62	0.02										
0.9	0.03	1.35	0.09								
1.66	0.11	1.29	0.1	2.46	0.16						
1.04	0.07	0.87	0.08	1.48	0.09	1.3	0.06				
1.26	0.08	1.05	0.08	1.7	0.11	1.08	0.03	1.51	0.09		
0.86	-----	0.59	0.05	0.9	0.07	0.6	0.05	0.69	0.05	0.88	0.04

θ1-6 - Variances variables; s.e. – standard error

Source: Calculated from survey results

As discussed above, multi-group analyses were conducted on each of the six sub-scales to assess the factor loadings and goodness-of-fit of models that included all race groups. The results indicate that the factor loadings follow a similar pattern as revealed in the restricted multi-group MIRT analyses discussed above. However, at times, the loadings were higher than those seen in the separate group analysis with some differences of up to 0.25. The

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 51

Team sub-scale reported two items (Q39 and Q42), with factor loadings equal to 1. This may be due to potential Haywood cases or items with very restricted variances. These items were flagged as potentially problematic in the subsequent analyses.

The group parameter estimates and model fit indices for each of the multi-group analyses, varied in a narrow range, across the four race groups and the six dimensions.

6.2.3 Configural invariance

The results seen when comparing race groups are similar to those seen in the instance of gender and are outlined in section 4.1.2. It seems that the issue of local dependence persists when examining item scores by race. While factor loadings are high across sub-scales and groups, together with low standard error, goodness-of-fit indices consistently report some degree of misfit. This may indicate that some items poorly reflect the underlying construct or violate the assumption of local independence. This may result in the detection of DIF, which may not be due to real differences across ethnic groups.

The DIF analysis was undertaken to determine whether the EEI items show evidence of metric invariance. In doing so, the DIF results are compared to the baseline models (multi-group unidimensional analyses) to determine the overall fit of the more constrained models. If the overall model fit deteriorates significantly, it is assumed that DIF may be prevalent in the given sub-scale. DIF statistics are also reported on item level.

6.2.4 Metric invariance

Wald's χ^2 statistic implies significant differences between groups. In the four group analyses, contrast 1 compares group 1 with 2, 3, and 4 (1 vs. 2, 3, and 4). This can be regarded as an overall omnibus significant test similar to ANOVA. Contrast 2 compares group 2 to 3 and 4 (2 vs. 3, 4), while contrast 4 compares group 3 to 4 (3 vs. 4).

In order to obtain a more detailed picture of the DIF between specific race groups, post-hoc DIF analyses were performed. These compared the DIF statistic of items between groups 1 and 2, 1 and 3, and 1 and 4, in three separate models per sub-scale, the results of which are provided below.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 52

- Five items on the customer service sub-scale demonstrated DIF according to the overall test (contrast 1).

- Four items on the immediate manager sub-scale demonstrated DIF according to the overall test (contrast 1). In general, only one item on the Organisational Commitment scale demonstrates DIF according to the overall test (contrast 1).

- Eight items on the Organisational Satisfaction sub-scale demonstrated DIF according to the overall test (contrast 1) and two items according to contrast 2. Clearly this sub-scale is plagued by DIF. Paired DIF post-hoc analyses on the Organisational Satisfaction sub-scale found

- Post-hoc analyses on the Strategy and Implementation sub-scale found that the items showed signs of DIF on four items (contrast 1) and one item on contrast 2. Similarly to Organisational Commitment above, there were no item differences between the African and Coloured groups.

- Eight items in the Team sub-scale demonstrated DIF according to the overall test (contrast 1). Unlike previous analyses, a post-hoc analysis of race pairs with Team found evidence of DIF between the African and Coloured groups, while there were no significant item differences between the African and Indian groups. The biggest differences in item functioning were found between the African and White groups.

It is clear from the post-hoc test that the biggest item differences for the Immediate Manager sub-scale are between the African and White groups. Comparisons between the Indian and African groups also demonstrated some significant differences. Organisational Commitment showed relatively little DIF. Clearly the sub-scale is functioning quite well and latent mean differences between groups can be made due to the strong evidence for metric invariance. A possible reason for detecting DIF on so many items in the Strategy and Implementation sub-scale may be due to the local independence function that was violated. In the unidimensional analyses per group, it was evident that some of the items were highly correlated after the influence of the latent variable was taken into consideration.

A comparison of the baseline and DIF models is presented in table 11. Goodness-of-fit on the DIF models is generally poorer, although the difference between these metrics is not overly large.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 53

M DU PLESSIS
N MARTINS

**Testing for measurement invariance for
employee engagement across different
demographic groups in South Africa**

TABLE 11:   Comparison between DIF and baseline models: race

Analysis	-2loglikeli-hood	AIC	BIC	M2	d.f.	RMSEA	χ2Diff.
Customer Service Baseline	27021.44	27273.44	27975.64	2117.53***	930	0.03	90.83***
Customer Service DIF	27112.27	27312.27	27869.57	2229.36***	956	0.03	
Immediate Manager Baseline	28655.67	28947.67	29761.33	6535.29***	1310	0.05	497.9***
Immediate Manager DIF	29153.6	29385.6	30032.07	6115.43***	1340	0.04	
Organisational Commitment Baseline	25091.9	25343.9	26046.1	3375.03***	930	0.04	215.6***
Organisational Commitment DIF	25307.49	25507.49	26064.79	3430.88***	956	0.04	
Organisational Satisfaction Baseline	32066.62	32438.62	33475.2	9294.11***	2262	0.04	592.6***
Organisational Satisfaction DIF	32659.22	32955.22	33780.03	9714.78***	2300	0.04	
Strategy & Implementation – Baseline	44502.15	44914.15	46062.19	8105.98***	2834	0.03	603.7***
Strategy & Implementation DIF	45105.8	45433.8	46347.77	10026.76***	2876	0.04	

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 54

Analysis	-2loglikeli-hood	AIC	BIC	M2	d.f.	RMSEA	χ2Diff.
Team Baseline	38860.43	39352.43	40723.39	11750.66***	4170	0.03	246.2***
Team DIF	39106.61	39498.61	40590.92	11848.47***	4220	0.03	

*** - p<0.001; χ2Diff. – Chi-squared difference test

Source: Calculated from survey results

As discussed in section 5.2.4 the -2loglikelihood difference test in table 11 demonstrates statistically significant differences between the DIF and baseline models at p<0.001. The differences are greatest for the Strategy and Implementation, Organisational Satisfaction and Immediate Manager sub-scales. The differences between the baseline and DIF analysis for Team are not overly large, relative to the other sub-scales, but the high values of the -2loglikelihood demonstrate the sub-scale is poorly fitting in both analyses. Due to the significant differences in the -2loglikelihood ratio, one can presume that the strict assumption of metric invariance does not hold across all six sub-scales. However, the differences in fit were significantly lower compared to the analysis with ethnicity as the grouping variable. This may suggest that the DIF is less problematic for race than for gender. Similar to the DIF analysis using race – the Organisational Commitment and Strategy and Implementation sub-scales seem to demonstrate the highest degree of invariance.

7.   CONCLUSION

The purpose of this study was to assess the measurement invariance of the EEI as a factor of gender and race. Previous results confirmed the validity and reliability of the instrument and successfully demonstrated evidence of configural invariance using a CFA (Martins 2015). The data were further used to test for measurement invariance among race and gender.

The three steps of measurement invariance according to the IRT framework (refer to figure 2) were used to test for measurement invariance among race and gender. The IRT analysis found support for the assumption of configural invariance for the total measure, as well as for the sub-scales; however, this was tempered by poor model fit (i.e. some of the individual

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 55

items do not fit the data very well) and violation of the local independence requirement of IRT. This violation was most pronounced for the Team sub-scale. It is posited that certain problematic items leading to a source of error in analysis that has not been accounted for, such as multidimensionality. It was, however, pleasing to see that the items that appear to be problematic are consistent across males, females, and race.

Unidimensional DIF analysis was undertaken to determine whether the EEI items show evidence of metric invariance.

The -2loglikelihood difference test displayed in tables 6 and 11 demonstrates a statistically significant difference in the model fit between the baseline and restricted DIF models. Due to the significant differences in the -2loglikelihood ratio, it is assumed that the assumption of metric invariance does not hold across all six sub-scales. However, the difference in fit was significantly lower compared to the analysis with suggesting that the DIF is less problematic for race than for gender. Nevertheless, given the large number of items and sample size, only eight items demonstrated significant DIF across gender (refer to table 7).

The EEI can thus be used with confidence for genders if the eight items are removed or adapted. The Organisational Commitment, Immediate Manager and Strategy and Implementation sub-scales demonstrated the highest degree of invariance for race. The results accordingly demonstrated that for the most part, it is not permissible to compare latent means directly across race. In order to remedy the problem, items demonstrating DIF should be removed or adapted and re-tested. These results thus confirm researchers concern that instruments are used without invariance testing and that testing for measurement invariance is a particularly important prerequisite within the South African context.

REFERENCES

ALBRECHT SL, BAKKER AB, GRUMAN JA, MACEY WH & SAKS AM. 2015. Employee engagement, human resource management practices and competitive advantage. *Journal of Organisational Effectiveness: People and Performance* 2(1):7-35.

BAKKER AB & DEMEROUTI E. 2008. Towards a model of work engagement. *Career Development International* 13(3):209-223.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 56

BARKHUIZEN N & ROTHMANN S. 2006. Work engagement of academic staff in South African higher education institutions. *Management Dynamics* 15(1):38-46.

BOCK RD & AITKIN M. 1981. Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika* 46:443-459.

BORTOLOTTI SLV, TEZZA R, DE ANDRADE DF, BORNIA AC & DE SOUSA JUNIOR AF. 2013. Relevance and advantages of using the item response theory. *Quality and Quantity* 47:2341-2360.

CHRISTIAN MS, GARZA AS & SLAUGHTER JE. 2011. Work engagement: a quantitative review and test of its relations with task and contextual performance. *Personnel Psychology* 64:89-136.

COETZEE SE & ROTHMANN S. 2005. Work engagement of employees in a higher education institution in South Africa. *South African Business Review* 9(3):23-34.

DE BEER M. 2005. Developing of the learning potential computerised adaptive test (LPCAT). *South African Journal of Psychology* 35(4):717-747.

DIMITROV DM. 2010. Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counselling and Development* 43(2):121-149.

FOXCROFT C & ROODT G (eds.) 2009. Introduction to psychological assessment in the South African context. 3rd ed. Cape Town: Oxford University Press.

GARWOOD J. 2006. Quantitative research. In Jupp V (ed). The SAGE dictionary of social research methods. London, UK: Sage. pp. 251-252.

GOLIATH-YARDE L & ROODT G. 2011. Differential item functioning of the UWES-17 in South Africa. *South African Journal of Industrial Psychology /SA Tydskrif vir Bedryfsielkunde, 37*(1), Art.[#897, 11 pages. http://dx.doi.org/10.4102. sajip.v37i1.897.]

GRIFFITH G. 2009. Conceptualising and exploring the meaning of employee engagement: an exploratory study. Johannesburg: University of the Witwatersrand. (Master's dissertation.)

HEYNS M & ROTHMANN S. 2016. Comparing trust levels of male and female managers: measurement invariance of the behavioural trust inventory. *South African Journal of Psychology* 46(1):74-87.

HORTENSIUS L. 2012. Project for introduction to multivariate statistics: measurement invariance. [Internet: www.tc.umn.edu/~horte005/docs/MeasurementInvariance.pdf; downloaded on 2016-05-30.]

IMANDIN L. 2015. Developing a conceptual framework to analyse engagement and disengagement in the workplace. Potchefstroom: North-West University. (DPhil thesis.)

KAHN WA. 1990. Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal* 33(4):692-724.

KIMBER M, REHM J & FERRO MA. 2015. Measurement invariance of the WHODAS 2.0 in a population-based sample of youth. *PLOS ONE 10* 11:1-13.

LYONS P & DOUECK HJ. 2009. The dissertation: from beginning to end. New York, NY: Oxford University Press.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 57

MACEY WH & SCHNEIDER B. 2008. The meaning of employee engagement. *Industrial and Organisational Psychology* 1:3-30.

MARAIS C, MOSTERT K & ROTHMANN S. 2009. The psychometrical properties of translated versions of the Maslach Burnout Inventory – general survey. *South African Journal of Industrial Psychology* 35(1):1-8.

MARKOS S & SRIDEVI MS. 2010. Employee engagement: the key to improving performance. *International Journal of Business and Management* 5(12):89-96.

MARTINS N. 2015. Testing for measurement invariance for employee engagement across sectors in South Africa. *Journal of Contemporary Management* 12:757-774.

MARTINS N. 2016. Confirming the measurement properties of an engagement measure. *Proceedings of the 15th European conference on research methodology for business and management studies.* London, UK: Kingston Business School, Kingston University:184-192.

MEIRING D, VAN DE VIJVER AJR, ROTHMANN S & BARRICK MR. 2005. Construct, item and method bias of cognitive and personality tests in South Africa. *South African Journal of Psychology* 31(1):1-8.

MILFONT TL & FISCHER R. 2010. Testing measurement invariance across groups: applications in cross-cultural research. *International Journal of Psychological Research* 3(1):111-121.

MOERDYK A. 2009. The principles and practices of psychological assessment. Pretoria: Van Schaik.

MOORE AA, NEALE MC, SILBERG JL & VERHULST B. 2016. Substance use and depression symptomatology: measurement invariance of the Beck Depression Inventory (BDI-II) among non-users and frequent-users of alcohol, nicotine and cannabis. *PL0S ONE* 11(4):1-14.

NIENABER H & MARTINS N. 2014. An employee engagement instrument and framework building on existing research. *Mediterranean Journal of Social Sciences* 5(20):485-496.

NIENABER H & MARTINS N. 2015a. Employee engagement in a South African context. Pretoria: KR Publishing.

NIENABER H & MARTINS N. 2015b. Validating a scale measuring engagement in a South African context. *Journal of Contemporary Management* 12:401-425.

SAKS AM. 2008. The meaning and bleeding of employee engagement: how muddy is the water? *Industrial and Organisational Psychology* 1:40-43.

SCHAUFELI W & SALANOVA M. 2011. Work engagement: on how to better catch a slippery concept. *European Journal of Work and Organisational Psychology* 20(1):39-46.

SCHAUFELI WB, SALANOVA M, GONZÁLEZ ROMÁ V & BAKKER AB. 2002. The measurement of engagement and burnout: a confirmative analytic approach. *Journal of Happiness Studies* 3:71-92.

SIMPSON M. 2009. Engagement at work: a review of the literature. *International Journal of Nursing Studies* 46:1012-1024.

STORM K & ROTHMANN S. 2003. A psychometric analysis of the Utrecht Work Engagement Scale in the South African Police Service. *South African Journal of Industrial Psychology* 29(4):62-70.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 58

TAY L, MEADE AW & CAO M. 2015. An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods* 18(1):3-46.

VAN ROOY DL, WHITMAN DS, HART D & CALEO S. 2011. Measuring employee engagement during a financial downturn: business imperative or nuisance? *Journal of Business Psychology* 26:147-152.

VILJEVAC A, COOPER-THOMAS HD & SAKS AM. 2012. An investigation into the validity of two measures of work engagement. *The International Journal of Human Resource Management* 23(17):3692-3709.

VISSER D & VIVIERS R. 2010. Construct equivalence of the OPQ32n for black and white people in South Africa. *South African Journal of Industrial Psychology* 36(1).

ZUMBO BD. 1999. A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modelling as a unitary framework for binary and Likert-type item scores. Ottawa, CAN: Department of National Defence, Directorate of Human Resources Research and Evaluation.

Journal of Contemporary Management
DHET accredited
ISSN 1815-7440

Volume 14
2017
Pages 24-59

Page 59