

TESTING MEASUREMENT INVARIANCE OF THE LEARNING PROGRAMME MANAGEMENT AND EVALUATION (LPME) SCALE ACROSS GENDER USING MULTI-GROUP CONFIRMATORY FACTOR ANALYSIS

*MC Tshilongamulenzhe **

Abstract

The purpose of this study was to test measurement invariance of the LPME scale across gender using multi-group CFA. The LPME scale was developed to measure the effectiveness of management and evaluation practices pertaining to occupational learning programmes in the South African skills development context. A non-experimental cross-sectional survey was conducted with 389 human resource practitioners and apprentices/learners. The results indicate that the LPME scale is invariant between males and females at the levels of configural, metric and strong invariance. The number of factors/constructs, pattern of item factor loading, latent constructs variances and covariances, and the reliability of the LPME scale and its dimensions are equivalent between males and females.

Key Words: Measurement Invariance; Learning Programme; Management, Evaluation; Factor Analysis

** Department of Human Resource Management, University of South Africa, South Africa*

Introduction

Testing the invariance of a measurement scale is a fundamental requirement in both applied and scientific use of measurement instruments (Blankson and McArdle, 2013). This aspect is a necessary part of psychometric evaluation of new scales in order to establish as to whether or not population/sample sub-groups interpret scale/sub-scale items the same way. During the scale development process, researchers often invest a lot of time writing items that are clear and free from ambiguity and thereafter subject such items to rigorous item, content and construct validity analysis in order to select the best. However, it cannot just be assumed that such items carry the same meaning and connotations for all target population/sample sub-groups. Thus, the invariance of scale items across sub-groups must be scientifically tested since failure by the researcher to satisfy this requirement may fuel speculation as to whether or not groups or individuals interpret the scale/sub-scale items differently, and as a result factor means may not be compared in a meaningful way (Drasgow and Kanfer, 1985; Horn and McArdle, 1992; Jöreskog, 1971; Vandenberg and Lance, 2000).

According to Byrne and van de Vijver (2010), very often researchers tend to assume that both the measuring instrument and the construct being measured are operating in the same way across a population of interest, and this assumption can be very lethal if not scientifically tested. Failure to prove

measurement invariance may lead the researcher to make inaccurate inferences and erroneous conclusions. Vandenberg and Lance (2000) have cautioned that failure to establish measurement and structural equivalence is as damaging to substantive interpretations of findings as is the inability to demonstrate reliability and validity of research. By testing for measurement invariance the researcher examines whether an instrument has the same psychometric properties across heterogeneous groups (Chen, 2007).

However, in order to prove measurement invariance, individuals with the same standing on a latent variable but sampled from different population sub-group should have the same expected observed score on a test of that variable. Blankson and McArdle (2013) indicate that invariance has been more often neglected in behavioral science research than it has been evaluated. It is therefore critical for the researchers involved in scale development and psychometric analysis to understand that invariance of a scale may not hold; that it cannot simply be assumed; and, that it is a hypothesis that can be tested.

In view of the foregoing, this study seeks to test the measurement invariance of a Learning Programme Management and Evaluation (LPME) across gender using multi-group confirmatory factor analytic procedure. A learning programme is a learnership, an apprenticeship, a skills programme or any other prescribed learning programme that

includes a structured work experience component (Coetzee, Botha, Kiley, Truman and Tshilongamulenzhe., 2012; Republic of South Africa, 2008; Van Rooyen, 2009). These programmes are provided for in the Skills Development Act 97 of 1998 (as amended in 2008). The learning programme pathway system came into being in an effort to address the inequalities and inconsistencies associated with the apartheid training regime in South Africa. These programmes are deemed as the most important innovation in the intermediate skills development space in the post-apartheid South Africa. They are administered by the Sector Education and Training Authorities (SETAs), which are in effect, a set of newly created institutions that have yet to develop capacity to drive skills development (Marock et al., 2008).

Brief overview of the LPME scale

The development of the LPME scale followed the established scale development conventions suggested by DeVellis (2012) and has complied with the basic psychometric expectations in relation to validity and reliability as reported by Tshilongamulenzhe, Coetzee and Masenge (2013). The researcher is not aware of any study undertaken to date that has sought to examine the measurement invariance of the LPME scale across gender in South Africa. Literature evidence shows no existence of any such studies in the South African skills development context.

Tshilongamulenzhe *et al* (2013) provided a detailed description of the process followed in the development of the LPME scale including the assessment of its psychometric properties. The scale is comprised of 81 items spread over 11 dimensions, that is: administrative processes, strategic leadership, policy awareness, stakeholders' inputs, quality assurance, environmental scanning, observation and problem solving, learning programme design and development, learning programme specifications, monitoring and evaluation and occupational competence.

The LPME scale was subjected to exploratory factor analysis, Rasch analysis and confirmatory factor analysis as part of psychometric evaluation. The Cronbach reliability coefficient of the scale was found to be .86 and .87 during the exploratory and confirmatory factor analyses respectively. The goodness of fit indices of the model were found to be acceptable and fitting the data well ($\chi^2/df = 2.68$; TLI = .97; IFI = .98; NFI = .97; CFI = .98; SRMR = .03; RMSEA = .06). However, the issue of factorial invariance of the LPME scale did not receive attention during the cross-validation stage, hence the current study.

Trends from the measurement invariance literature

The literature points to various recommendations regarding the sequences of measurement invariance tests (Bollen, 1989; Byrne, Shavelson and Muthén, 1989; Cheung, 2008; Cheung and Rensvold, 2002; Drasgow and Kanfer, 1985; Jöreskog and Sörbom, 1993; Little, 1997; Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000) but none of these is absolute as the decision for choice by the researcher is reliant on the research question to be answered. In the current study, factorial invariance was assessed to fit a series of hierarchically nested factor structures, that is, configural, metric, strong and strict models.

Configural invariance (Unconstrained model)

Configural invariance is established if a CFA model that allows the same set of items to form a factor in each group shows good model fit (Lim and Townsend, 2012). This type of invariance requires that an instrument represents the same number of common factors across groups, and that each common factor is associated with identical item sets across groups (Van de Velde, Levecque and Bracke, 2009). To test for configural invariance, the manifest variable intercepts, residual variance and factor loadings must be freely estimated for the sub-groups (e.g. male and female); factor means must be fixed at zero.

Metric/weak invariance (Measurement weights)

Metric or weak invariance exists if the strength of the relationship (factor loading) between each item and the latent construct under consideration are invariant across groups. It tests whether the common factors have the same meaning across sub-groups, that is, whether the factor loadings are equal across sub-groups. Factor loadings represent the strength of the linear relation between each factor and its associated items (Bollen, 1989). To test for metric/weak invariance, loadings have to be constrained to be equal across sub-groups for all indicators, with all other aspects of the model specified as described for the configural model. If only weak invariance holds, then meaningful comparisons across groups can be made of the variances and covariances among latent variables but not of the latent means or observed means, covariances, and variances (Bontempo and Hofer, 2007; Gregorich, 2006; McArdle, 2010; McArdle, Smith and Willis, 2011; Widaman and Reise, 1997).

Strong invariance (Measurement intercepts)

Strong invariance exists with item similarity across groups. This type of invariance addresses the question of whether or not there is differential additive response bias (Cheung and Rensvold, 2000; Rorer, 1965; Steenkamp and Baumgartner, 1998). External forces unrelated to the common factor tend to cause such a bias which manifest itself through higher or lower-valued item response in one population sub-group compared to another. For strong invariance, manifest variable intercepts have to be constrained to be equal across sub-groups; factor means have to be fixed at zero in the first group and have to be freely estimated in the second group. Evidence of strong measurement invariance is all that is required to ensure meaningful comparisons in latent means across groups (Widaman and Reise, 1997).

Strict invariance (Measurement residuals)

Strict invariance tests the unique error variances associated with each item across groups (Horn and McArdle, 1992; Meredith, 1993). Strict factorial invariance has been argued to be indicative of true measurement invariance (Wicherts and Dolan, 2010). However, for most researchers, only comparison of sub-group means is of main interest (Van de Velde, Levecque and Bracke, 2009). Residual or strict invariance allows for the comparison of observed variance or covariance across sub-groups. To test for strict invariance, manifest variable residuals have to be constrained to be equal across sub-groups.

At each level of measurement invariance testing briefly discussed above, additional constraints have to be applied to the multi-group measurement model, using nested-model comparisons to determine whether the added constraints contribute to poor model fit. The comparison of (a) configural invariance to metric/weak factorial invariance, (b) weak to strong factorial invariance, and (c) strong to strict factorial invariance involves the addition of loading constraints, latent intercept constraints and unique factor variance constraints respectively across gender sub-groups (White, Umana-Taylor, Knight and Zeiders, 2011)

Goodness-of-fit indices have to be used to make decisions about the accuracy of the models in accordance with the guidelines provided by Cheung and Rensvold (2002) and Chen (2007). Consequently, the procedure adopted for multi-group CFA model testing in each of the 11 sub-scales of the LPME scale in the current study was as follows:

Invariance testing procedure

All models were estimated from a baseline configural model in which the sub-scales were treated as

indicators of the latent construct. For model identification purposes, the factor loading for one indicator was set at 1 (keeping this consistent across males and females). When model fit was deemed acceptable, the next step was to test metric/weak invariance model. In doing so, loading constraints were added requiring that the factor loadings for individual items be equal across the two sub-groups (males and females). When model fit was acceptable, the next step was to test strong invariance model. In the strong invariance model the loading constraints were retained and latent intercept constraints were added. When model fit was acceptable the last step was to test strict invariance models. In the strict invariance models the loading and latent intercept constraints were retained.

Blackson and McArdle (2013) attest that the question of measurement invariance should be considered in all research in which analyses are directed at showing that measurement attributes (states and traits), and the relationships among such attributes are different for different classifications of people or for the same people measured under different circumstances (times and places). As a result, this study seeks to test the following measurement invariance assumptions regarding the LPME scale across gender using multi-group CFA, that:

- a) The number of factors/constructs for the LPME scale is equivalent between males and females.
- b) The pattern of item factor loading is equivalent between males and females, that is, the set of items in each dimension of the LPME scale measures the same construct between males and females.
- c) The latent construct variances and covariances are the same between males and females, that is, the true score variance for each construct and relations among these constructs are the same for males and females.
- d) The reliability of the LPME scale and its dimensions is equivalent between males and females, that is, each dimension/construct is measuring what it was designed to measure in the same way for males and females.

A multi-group CFA was conducted to test the invariance of the LPME scale between the target sample sub-groups (*i.e.*, male and female). Multi-group CFA compares groups by measurement weights, measurement intercepts and measurement residuals. Examining differences in these additional parameters provides a clearer understanding of the nature of any potential moderating effects. Moderation is indicated by a significant change in model fit when the coefficients were constrained to be equal between groups. In testing for measurement invariance, the models of interest are necessarily nested and thus can be compared in pairs by computing the difference in their overall Chi-Square

values and the related degrees of freedom (df) ($p \geq .01$ for significance) (Cheung and Rensvold, 2002).

Methodology

Research approach

This study followed a quantitative approach using a non-experimental, cross-sectional survey design. Primary data for the study were collected from two metropolitan municipalities in Gauteng Province and a provincial government department in the North West Province, South Africa.

Participants

The population in this study comprised human resource practitioners and learners/apprentices from two metropolitan municipalities in Gauteng Province and a provincial government department in the North West Province, South Africa. A probabilistic simple random sampling technique was used to select participants from their organisations' databases. The target sample was 900 participants to whom questionnaires were distributed. About 579 completed questionnaires were returned yielding a 64% response rate.

All returned questionnaires were carefully analysed in two phases: Phase 1 focused on any possible anomalies related to incompleteness and non-compliance with the instruction, whereas Phase 2 focused on whether or not participants completed the key variable of the current study (gender).

After the Phase 1 analysis, about 187 questionnaires were discarded as they had missing data. This was done in order to comply with the AMOS software requirement for computation of modification indices in the event where the model and data do not fit well. At the end of this phase, a total of 392 questionnaires were retained for the subsequent round of data management.

Phase 2 was carried out as necessitated by the focus and objective of the current study which sought to test measurement invariance of participants across gender. Only questionnaires in which participants indicated their gender were retained during this second phase of analysis. Consequently, about 3 questionnaires were eliminated as participants who completed them did not indicate their gender.

The final pool comprised 389 questionnaires which were split between two gender sub-groups (Group 1 – Male ($n = 220$); Group 2 – Female ($n = 169$)). About 86% of the participants were below the age of 35 years. In terms of academic achievement, 74% of the participants had acquired a matriculation level certificate, its equivalence or below. Whilst 84.9% of participants had some exposure to learnerships, about 79% were learners/apprentices.

Measure

The 81 items 11-dimensional Learning Programme Management and Evaluation (LPME) scale developed by Tshilongamulenzhe (2012) was used for data collection in this study. The Cronbach's reliability coefficient for the scale was .86 and .87 during the exploratory and confirmatory factor analyses phases respectively, whilst that of the 11 sub-scales ranged from .82 to .93 (Tshilongamulenzhe *et al.*, 2013).

Procedure

The researcher sought permission to undertake this study from the 3 participating organisations. Data were collected using a self-administered questionnaire which was distributed through a drop-in and pick-up method. Instructions to complete the questionnaire as well as the contact numbers of the research team were provided to the participants. The purpose of the study was clearly communicated to the participants including all ethical aspects such as anonymity, consent, freedom to discontinue and confidentiality of responses. Further to the written instructions, each participant was orally briefed by a member of the research team regarding the purpose of the study, his/her rights and privileges as a participant. Participants were also given the opportunity to raise any questions they had prior to completion of the questionnaire.

Data analysis

Statistical Package for Social Sciences (SPSS) (IBM, 2013) and AMOS software (versions 21.0) (Arbuckle, 2013) were used to conduct data analysis in the current study. The statistical computations were informed by the objective of this study and have included descriptive statistics (mainly frequencies), scale reliability analysis, as well as multiple-group confirmatory factor analysis.

Results

Sample characteristics

The sample comprised of 56.6% male participants relative to 43.4% females. About 59.5% of the participants were aged 35 years and below. In terms of educational achievement, 60.4% of the participants achieved a Grade 12 (Matric/Senior Certificate) or below. Occupationally, 89.5% of the participants had been involved in a learnership programme fulfilling various roles prior to or at the time of the survey. Finally, 81% of the participants were learners/apprentices.

Invariance tests

Table 1 shows that all model parameters were systematically constrained to be equal between sub-groups, with each constraint being applied in an additive manner. More specifically, the configural model constrained the strict model; the second model constrained the strong model and the strict model; the third model constrained the metric model, the strong model and the strict model. At each stage, any constraint that failed to result in a significant Chi-Square change was retained in subsequent comparisons, to improve parsimony, while narrowing the source of variability between groups and freeing

degrees of freedom in the model (Cole and Maxwell, 2003; Kline, 2005).

The CFI value of the configural model fit was estimated at .985, the RMSEA value was .057 and the SRMR value was .019 as shown in Table 1 for both males and females. These model fit indices suggest that the model shows configural invariance between males and females. The model provided the values against which all subsequently specified invariance models were compared. It was therefore, reduced from a second-order factor model to a first-order factor model. Both male and female respondents were tested separately to check for adequate model fit and the results in Table 1 show a good fit.

Table 1. Model fit summary: χ^2 , CFI, RMSEA and SRMR for males and females

Model	χ^2	$\Delta\chi^2$	Df	Δdf	CFI	ΔCFI	RMSEA	PCLOSE	$\Delta RMSEA$	SRMR	$\Delta SRMR$
Total sample (n = 389)	94.124	-	33	-	.985	-	.069	.028	-	.019	-
Configural	149.231	55.107	66	33	.981	-.004	.057	.162	-.012	.019	0
Metric	197.268	48.037	76	10	.972	-.009	.064	.017	.007	.031	.012
Strong	197.402	0.134	77	1	.972	0	.064	.021	0	.029	-.002
Strict	319.458	0.000	99	0	.949	0	.076	.000	0	.041	0

Male = 220 (56.6%); Female = 169 (43.4%)

The model fit indices for the configural model as depicted in Table 1 indicate that the eleven dimensional LPME scale fits the empirical data well for the two sub-groups of participants (*i.e.*, male and female). Thus, the LPME scale represents the same number of common factors across sub-groups, and that each common factor is associated with identical item sets across sub-groups.

The metric model CFI value of .972, the RMSEA value of .064 and the SRMR value of .031 as depicted in Table 1 were found to support metric invariance. The change in chi square ($\Delta\chi^2 = 48.037$), the change in degrees of freedom ($\Delta df = 10$), and the difference in ΔCFI values (-.009) between the configural and metric models as depicted in Table 1 were also found to support metric invariance and are within the recommended criterion of $\leq -.010$ for significance when testing metric invariance (Cheung and Rensvold, 2002). The $\Delta RMSEA$ was found to be .007 whereas the $\Delta SRMR$ was .012. Both these values were respectively less than the $\leq .015$ (RMSEA) and the $\leq .030$ (SRMR) threshold recommended by Cheung and Rensvold (2002). These findings strongly support metric invariance and confirm that factor loading parameters of the LPME scale are invariant between males and females. Thus, there is a strong agreement between males and females regarding how the constructs for the LPME scale were manifested.

In order to test strong/intercept invariance, a change of $\geq -.010$ in CFI (Cheung and Rensvold,

2002), supplemented by a change of $\geq .015$ in RMSEA or a change of $\geq .010$ in SRMR would indicate noninvariance. The results in Table 1 show that the strong model CFI value was .097, the RMSEA value was .064 and the SRMR value was .029 and these support strong invariance. The ΔCFI and $\Delta RMSEA$ were both at .0 and $\Delta SRMR$ was -.002. These values are within the cut-off criteria of $\leq -.010$ for CFI, $\leq .015$ for RMSEA and $\leq .010$ for SRMR as recommended by Cheung and Rensvold (2002). These results support strong invariance and suggest that the vectors of item intercepts are equal between males and females. Thus, the LPME scale shares the same operational definitions including the same interval and same zero points across males and females, which further suggests that meaningful comparison of the latent means can be achieved (Cheung and Lau, 2011). In this regard, the sub-group differences in estimated factor means are therefore unbiased and the sub-group differences in observed means are directly related to sub-group differences in factor means and are not contaminated by differential additive response bias.

For strict invariance, the results in table 1 shows an acceptable CFI value of .949, an RMSEA value of .076 and an SRMR value of .041. However, since the comparison of group means was of main interest for the researcher, the test for strict invariance was found to be of limited practical value in this study.

Table 2. Nested Model Comparisons

Assuming Configural Model to be correct:

Model	DF	CMIN	P	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Metric	10	48.037	.000	.011	.011	.008	.009
Strong	33	170.227	.000	.039	.039	.024	.025
Strict	33	170.227	.000	.039	.039	.024	.025

Assuming Strict Model to be correct:

Model	DF	CMIN	P	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Strong	1	.135	.714	.000	.000	-.001	-.001
Strict	23	122.190	.000	.028	.028	.016	.016

Assuming Strong Model to be correct:

Model	DF	CMIN	P	NFI Delta-1	IFI Delta-2	RFI rho-1	TLI rho2
Strict	22	122.055	.000	.028	.028	.017	.017

Table 2 presents the results of nested model comparisons for the models tested in this study. All model parameters were systematically constrained to be equal between males and females, respectively with the $\chi^2 = 122.055$; $df = 22$; ΔNFI and $\Delta IFI = .028$, and an ΔRFI and $\Delta TLI = .017$ ($p \geq .01$ for significance) (Cheung and Rensvold, 2002).

By adding one constraint (strong) to the model, both ΔNFI and ΔIFI remained constant (.000) while the ΔRFI and ΔTLI decreased by -.001 respectively ($RFI = .016$; $TLI = .016$). By adding two constraints (strong and metric) to obtain the model, ΔNFI and ΔIFI increased by .011 respectively ($NFI = .039$; $IFI = .039$), while the ΔRFI and ΔTLI respectively changed by .008 and .009 ($RFI = .024$; $TLI = .025$).

As depicted in Table 2, a comparison of the constrained model (metric) with the non-constrained model (configural) yielded a χ^2 difference of 48.037 with a difference in degrees of freedom of 10 ($CMIN/df, = 4.80$) which is significant at $p \geq .01$ (Cheung and Rensvold, 2002). At this level of statistical significance and accepting the scale of change in model fit indices, it can be said that the LPME scale is invariant between males and females.

Discussion

Researchers and scholars have examined measurement invariance across gender (Eagle, Miles and Icenogle, 2001; Tang, Luna-Arocas and Sutarso, 2005) although such studies focused on scales measuring different other phenomena. The purpose of the current study was to test the measurement invariance of the LPME scale across gender using multi-group CFA. According to Van de Velde, Levecque and Bracke (2009), multi-group CFA allows the researchers to compare the means and variance of latent constructs by correcting for possible bias due to variation across groups in the number of common factors and the item/factor

clusters (configural invariance), factor loadings (metric invariance), item intercepts (strong invariance) and residual variances (strict invariance).

Yen and Lan (2013) posit that scientific proof must be provided first before conclusions are made that sample sub-groups comprehend the items or sub-scales in a particular measure in the same manner. It is pivotal therefore that a research instrument measures constructs with the same meaning across groups and allows defensible quantitative group comparisons (Van de Velde, Levecque and Bracke, 2009). In the current study, the LPME scale that measures effective management and evaluation practices pertaining to occupational learning programmes in the South African skills development context was used. An occupational learning programme was considered a latent construct whose properties are inferred by observing a set of dimensions compressed from variables that serve as manifest indicators. It was imperative for the researcher to test as to whether or not males and females differ in their mean score and interpretation of the dimensions and variables that comprise the LPME scale, hence this comparative quantitative research.

As argued by Ployhardt and Oswald (2004) and Thompson and Green (2006), legitimate comparison of means or structural relations across groups requires equivalence of the measurement structure underlying the indicators. Otherwise, comparisons of mean differences or other structural parameters across groups are meaningless without evidence of measurement invariance (Schmitt and Kuljanin, 2008). It has been argued that perhaps strong invariance is all that is necessary for meaningful comparisons in latent means across groups (Widaman and Reise, 1997).

The current study established measurement invariance of the LPME scale between males and females at the level of configural, metric and strong

invariance as supported by the empirical findings. Thus, the findings of this study support the assumptions that, the number of factors/constructs for the LPME scale is equivalent between males and females; the pattern of item factor loading is equivalent between males and females; the latent construct variances and covariances are the same between males and females; and that, the reliability of the LPME scale and its dimensions is equivalent between males and females. These findings support full measurement invariance of the LPME scale between males and females since all tested and significant parameters are invariant across these sub-groups. The results met the cut-off criteria to support invariance as suggested by Cheung and Rensvold (2002) for configural, metric and strong invariance.

Conclusions

The study found that number of factors/constructs, the pattern of item factor loading, the latent constructs variances and covariances, and the reliability of the LPME scale and its dimensions are equivalent between male and female participants. Although scientific strides have been made to test the psychometric properties of the LPME scale by Tshilongamulenzhe (2012) and Tshilongamulenzhe *et al* (2013), the review work on the instrument must carry on to further examine its scientific worth and level of invariance across sub-groups. The current study provides a necessary leap not only towards the understanding of measurement invariance in the South African skills development context, but through scientific value-add with regards to the broader comprehension of gender influence on the conceptualisation and interpretation of new scale items, an aspect which must be seriously tested as part of the broader psychometric evaluation of new scales. The new scale requires further scientific scrutiny to establish its invariance across different other sample sub-groups such as age, ethnicity and occupation.

References

1. Arbuckle, J. L. (2013). IBM SPSS Amos 21 User Guide, Amos Development Corporation, New York.
2. Blankson, A.N. and McArdle, J.J. (2013). "Measurement invariance of cognitive abilities across ethnicity, gender and time among older Americans". *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, pp. 1-12.
3. Bollen, K.A. (1989). *Structural equations with latent variables*, Wiley, New York.
4. Bontempo, D.E. and Hofer, S.M. (2007). "Assessing factorial invariance in cross-sectional and longitudinal studies", in Ong, A.D. and van Dulmen, M (Eds.), *Handbook of methods in positive psychology*, Oxford University Press, New York, pp. 153-175.
5. Byrne, B.M, Shavelson, R.J. and Muthén, B. (1989). "Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance". *Psychological Bulletin*, Vol. 105, pp. 456-466.
6. Byrne, B.M. and van de Vijver, F.J.R. (2010). "Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence". *International Journal of Testing*, Vol. 10, pp. 107-132.
7. Chen, F. (2007). "Sensitivity of goodness of fit indexes to lack of measurement invariance". *Structural Equation Modeling*, Vol. 14, No. 3, pp. 464-504.
8. Cheung, G.W. (2008). "Testing equivalence in the structure, means and variances of higher-order constructs with structural equation modelling". *Organizational Research Methods*, Vol. 11, pp. 593-613.
9. Cheung, G.W. and Lau, R.S. (2011). "A direct comparison approach for testing measurement invariance". *Organizational Research Methods*, pp. 1-32
10. Cheung, G.W. and Rensvold, R.B. (2002). "Evaluating Goodness-of-Fit Indexes for testing measurement invariance". *Structural Equation Modeling*, Vol. 9. No. 2, pp. 233-255.
11. Coetzee M, Botha J, Kiley J, Truman K. and Tshilongamulenzhe M.C. (2012). *Practising Training and Development in South Africa* (2nd edn), Juta & Company Ltd, Cape Town.
12. Cole, D.A. and Maxwell, S.E. (2003). "Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modelling". *Journal of Abnormal Psychology*, Vol. 112, pp. 558-577.
13. DeVellis, R.F. (2012). *Scale Development: Theory and Application* (2nd edn), Sage Publications, Thousand Oaks, CA.
14. Drasgow, F. and Kanfer, R. (1985). "Equivalence of psychological measurement in heterogeneous populations". *Journal of Applied Psychology*, Vol. 70, pp. 662-680.
15. Gregorich, S.E. (2006). "Do self-reported instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework". *Medical Care*, Vol. 44, pp. 78-94.
16. Eagle, B.W, Miles, E.W. and Icenogle, M.L. (2001). "Male and female interpretations of Bi-Directional Work-Family Conflict Scales: Testing for measurement equivalence", in Schriesheim, C.A. and Neider, L.L (Eds.), *Equivalence in measurement*, Information Age Publishing, Connecticut, USA, pp. 5-23.
17. Horn, J.L. and McArdle, J.J. (1992). "A practical and theoretical guide to measurement invariance in aging research". *Experimental Aging Research*, Vol. 18, pp. 117-144.
18. IBM. (2013). *IBM SPSS Statistics 21 Brief Guide*, IBM Corporation, New York.
19. Jöreskog, K.G. (1971). "Simultaneous factor analysis in several populations". *Psychometrika*, Vol. 36, 409-426.
20. Jöreskog, K. and Sörbom, D. (1993). *LISREL 8: User's reference guide*, Scientific Software International, Chicago.
21. Kline, R. B. (2005). *Principles and practice of structural equation modelling* (2nd edn.), The Guilford Press, New York, NY.

22. Lim, J. and Townsend, A. (2012). "Cross-ethnicity measurement equivalence of family coping for breast cancer survivors". *Research on Social Work Practice*, Vol. 22, No. 6, pp. 689-703.
23. Little, T.D. (1997). "Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues". *Multivariate Behavioural Research*, Vol. 32, pp. 53-76.
24. McArdle, J.J. (2010). "Contemporary challenges of longitudinal measurement using HRS data", in Walford, G, Tucker, E. and Viswanathan, M (Eds.), *The SAGE Handbook of Measurement*, SAGE Press, London, pp. 509-536.
25. McArdle, J.J, Smith, J.P. and Willis, R.J. (2011). "Cognition and economic outcomes in the health and retirement study", in Wise, D.A (Ed.), *Explorations in the economics of aging*, National Bureau of Economic Research, Cambridge, MA, pp. 209-236.
26. Meredith, W. (1993). "Measurement invariance, factor analysis and factorial invariance". *Psychometrika*, Vol. 58, pp. 525-543.
27. Ployhardt, R.E. and Oswald, F.L. (2004). "Application of mean and covariance structure analysis: Integrating correlational and experimental approaches". *Organizational Research Methods*, Vol. 7, No. 1, pp. 27-65.
28. Republic of South Africa. (2008). *Skills Development Amendment Act, Act No. 37 of 2008*, Government Printers, Pretoria.
29. Rorer, L.G. (1965). "The great response-style myth". *Psychological Bulletin*, Vol. 63, pp. 123-156.
30. Schmitt, N. and Kuljanin, G. (2008). "Measurement invariance: Review of practice and implications". *Human Resource Management Review*, Vol. 18, pp. 210-222.
31. Steenkamp, J.E.M. and Baumgartner, H. (1998). "Assessing measurement invariance in cross-national consumer research". *Journal of Consumer Research*, Vol. 25, pp. 78-90.
32. Tang, T.L.P, Luna-Arocas, R. and Sutarso, T. (2005). "From income to pay satisfaction: The love of money and pay equity comparison as mediators and culture and gender as moderators". *Management Research: The Journal of the Iberoamerican Academy of Management*, Vol. 3, No. 1, pp. 7-26.
33. Thompson, M.S. and Green, S.B. (2006). "Evaluating between-group differences in latent means", in Hankok, G.R. and Mueller, R.O (Eds.), *Structural Equation Modeling: A second course*, Information Age, Greenwich, pp. 119-169.
34. Tshilongamulenzhe, M.C. (2012). *An Integrated Learning Programme Management and Evaluation Model for the South African Skills Development Context*. Unpublished Doctorate Thesis, Pretoria, University of South Africa.
35. Tshilongamulenzhe, M.C, Coetzee, M. and Masenge, A. (2013). "Development of the Learning Programme Management and Evaluation Scale for the South African Skills Development Context". *South African Journal of Industrial Psychology*, Vol. 39, No. 1, pp. 1-14.
36. Vandenberg, R.J. and Lance, C.E. (2000). "A review and synthesis of the measurement invariance literature: Suggestions, practice and recommendations for organizational research". *Organizational Research Methods*, Vol. 3, pp. 4-70.
37. Van de Velde, S, Levecque, K. and Bracke, P. (2009). "Measurement equivalence of the CES-D 8 in the general population in Belgium: A gender perspective". *Archives of Public Health*, Vol. 67, pp. 15-29.
38. Van Rooyen, M. (2009). *New learning programmes leading to new qualifications*, Cutting Edge, Johannesburg.
39. White, R.M.B, Umana-Taylor, A.J, Knight, G.P. and Zeiders, K.H. (2011). "Language measurement equivalence of the Ethnic Identity Scale with Mexican American early adolescents". *Journal of Early Adolescence*, Vol. 31, No. 6, pp. 817-852.
40. Wicherts, J.M. and Dolan, C.V. (2010). "Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities". *Educational Measurement: Issues and Practice*, Vol. 29, pp. 39-47.
41. Widaman, K.F. and Reise, S.P. (1997). "Exploring the measurement invariance of psychological instruments: Applications in the substance use domain", in Bryant, K.J, Windle, M. and West, S.G (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research*, American Psychological Association, Washington, DC, pp. 281-324.
42. Yen, W.K.S. and Lan, O.S. (2013). "Measurement invariance of Students' Parents Action Questionnaire (SPAQ) among three ethnic groups in Malaysia". *The International Journal of Educational and Psychological Assessment*, Vol. 12, No. 2, pp. 123-139.