Testing measurement invariance of the Learning Programme Management and Evaluation scale across academic achievement

CrossMark
←click for updates

Author:
Maelekanyo C. Mulaudzi[1]

Affiliation:
[1]Department of Human Resource Management, University of South Africa, South Africa

Corresponding author:
Maelekanyo Mulaudzi, tshilmc@unisa.ac.za

Orientation: Measurement invariance is one of the most precarious aspects of the scale development process without which the interpretation of research findings on population sub-groups may be ambiguous and even invalid. Besides tests for validity and reliability, measurement invariance represents the hallmark for psychometric compliance of a new measuring instrument and provides the basis for inference of research findings across a range of relevant population sub-groups.

Research purpose: This study tested the measurement invariance of a Learning Programme Management and Evaluation (LPME) scale across levels of academic achievement.

Motivation for the study: It is important for any researcher involved in new scale development to ensure that the measurement instrument and its underlying constructs have proper structural alignment and that they both have the same level of meaning and significance across comparable heterogeneous groups.

Research design, approach and method: A quantitative, non-experimental, cross-sectional survey design was used, and data were obtained from 369 participants who were selected from three public sector organisations using a probabilistic simple random sampling technique. The Statistical Package for Social Sciences and Analysis of Moment Structures software (versions 21.0.0) were used to analyse the data.

Main findings: The findings show that all the four invariance models tested have achieved acceptable goodness-of-fit indices. Furthermore, the findings show that the factorial structure of the LPME scale and the meaning of its underlying constructs are invariant across different levels of academic achievement for human resource development (HRD) practitioners and learners or apprentices involved in occupational learning programmes.

Practical implications: The findings of this study suggest practical implications for HRD scholars as they are enabled to make informed decisional balance comparisons involving educational attainment sub-groups.

Contributions and value addition: This study contributes methodologically to the sub-field of HRD by enabling scholars to make comparisons of mean differences or other structural parameters across sub-groups.

Introduction

Social science researchers are increasingly concerned with testing for measurement invariance; that is, determining if items used in survey-type instruments mean the same thing to members of different groups (Cheung & Rensvold, 2002). This concern is inevitable because measurement invariance is a key aspect of the scale development process, especially where the target population is heterogeneous. According to Van de Schoot, Schmidt, De Beuckelaer, Lek and Zondervan-Zwijnenburg (2015), a meaningful comparison of latent factor means between groups of a heterogeneous population could be achieved if the measurement structures of the latent factor and their survey items are stable, that is, invariant. Very often, researchers ignore measurement invariance issues and compare latent factor means across groups or measurement occasions even though the psychometric basis for such a practice does not hold. However, measurement invariance is a fundamental requirement in both applied and scientific use of measurement instruments (Blankson & McArdle, 2015). Researchers invest their time in writing scale items that are unambiguous and clear, and item analyses are carried out to select the best items that comprise a measurement instrument. However, it has often been assumed that the items of a measurement

instrument will have the same connotations and meanings for all people and therefore the scale is invariant in comparisons of people of different classifications (Horn & McArdle, 1992). If this assumption of invariance is incorrect, then conclusions of group comparisons based on results from studies applying such a measurement instrument are likely to be incorrect.

As Blankson and McArdle (2015) state, regardless of its importance, measurement invariance has been more often neglected in behavioural sciences than it has been evaluated. The invariance assumption has rarely been stated as a hypothesis and tested, although increasing attention is being paid by researchers to this issue in recent times (Adolf, Schuurman, Borkenau, Borsboom & Dolan, 2014; Bowden, Saklofske & Weiss, 2011; Chiorri, Day & Malmberg, 2014; Guenole & Brown, 2014; Hox, De Leew & Zijlmans, 2015; Savage-McGlynn, 2012; Tshilongamulenzhe, 2015; Van de Schoot, Lutgtig & Hox, 2012; Van de Schoot et al., 2013, 2015; Wang, Merkle & Zeileis, 2014; Zercher, Schmidt, Cieciuch & Davidov, 2015). Consequently, the question of measurement invariance should be considered in all behavioural science research wherein analyses are directed at showing that measured attributes, and the relationships among such attributes, are different for different classes of people or for the same people measured under different circumstances (Blankson & McArdle, 2015). To this end, the primary objective of this study is to heighten awareness among researchers involved in new scale development in order to ensure that the measurement instruments they design and their underlying constructs have proper structural alignment, and that they both have the same level of meaning and significance across comparable heterogeneous groups. Such awareness is particularly important if the success of studies applying newly developed measures hinge on the possibility of making meaningful comparisons across groups.

Brief context of the study and an overview of the Learning Programme Management and Evaluation scale

Guided by the scale development protocol suggested by DeVellis (2012), Tshilongamulenzhe (2012) developed a new Learning Programme Management and Evaluation (LPME) scale which seeks to enhance the effectiveness of management and evaluation practices pertaining to occupational learning programmes (OLPs) in South Africa. These programmes have been proclaimed by the South African government as a fundamental mechanism to address skills shortages; hence, vocational and occupational certification via learnership and apprenticeship programmes is at the core of the new skills creation system (Tshilongamulenzhe, Coetzee & Masenge, 2013). An OLP includes a learnership, apprenticeship, skills programme or any other prescribed learning programme that includes a structured work experience component (Republic of South Africa, 2008).

A number of challenges have been raised regarding the co-ordination and management of skills development training projects in South Africa, including poor quality of training and lack of mentorship (Du Toit, 2012). Mummenthey, Wildschut and Kruss (2012) revealed the prevalence of difference in standards across the different occupational learning routes (learnership, apprenticeship, skills programme), which brought about inconsistencies regarding procedures to implement occupational training. These shortcomings are indicative of the management and evaluation weaknesses impacting the South African skills development system, and they raise serious concerns about the quality of occupational learning (Tshilongamulenzhe et al., 2013).

Learnerships and apprenticeships are potentially significant routes to critical vocational and occupational qualifications in South Africa and the promise of future employment (Wildschut, Kruss, Janse van Rensburg, Haupt & Visser, 2012). They represent important alternative routes to enhance young peoples' transition to the labour market and to meet the demand for scarce and critical skills. If these interventions are not effectively managed and evaluated, the dreams and aspirations of many young people who hope to acquire skills would be adversely affected. Equally important, poor management and evaluation of these interventions have the potential to further drown the economy which is struggling from persistent skills shortages. It is for some of these reasons that Tshilongamulenzhe (2012) developed a measure for LPME which is relevant to the South African occupational learning context. No evidence was found that shows the existence of such a measure in the South African skills development context (Tshilongamulenzhe, 2012). The newly developed LPME scale was necessitated by the need for an integrated and coherent approach towards occupational LPME with a view to effectively promote the alignment of skills development goals with the needs of the workplace in support of the goals of the National Skills Development Strategy.

A detailed account regarding the scientific process followed in the development of the LPME scale, including its underlying theoretical constructs, and evidence of exploratory factor and Rasch analyses are provided by Tshilongamulenzhe et al. (2013). Although adequate scientific evidence has been presented which show good psychometric properties for the LPME scale, including its validity and reliability, the degree to which the conceptual foundation and the underlying constructs measured in this scale can be validly and reliably compared across heterogeneous population groups is yet to be tested; hence, this study which seeks to test invariance across educational achievements.

Theoretical perspectives regarding measurement invariance

The diversity of the population from which the samples were drawn for both the development (Tshilongamulenzhe et al., 2013) and the cross-validation studies related to the LPME scale necessitate a further scientific examination of data in order to make informed decisional balance comparisons involving sub-groups of this population. The two samples were heterogeneous with inherent differences across a range

of demographic indicators. It is therefore significant for the researcher to examine the influence of these heterogeneous characteristics on the meaning that participants in these studies ascribe to the underlying constructs and/or items measured by this new scale. The premise that sample sub-groups comprehend the items or sub-scales in a particular measure in the same manner has to be proven first before any conclusions are made (Yen & Lan, 2013).

As Cheung and Rensvold (2002) state, if measurement invariance of an instrument cannot be established, then the finding of a between-group difference cannot be unambiguous. Thus, legitimate comparison of means or structural relations across groups requires equivalence of the measurement structure underlying the indicators (Ployhardt & Oswald, 2004; Thompson & Green, 2006). Otherwise, comparisons of mean differences or other structural parameters across groups are meaningless without evidence of measurement invariance (Schmitt & Kuljanin, 2008). Consequently, the objective of this study is to test the measurement invariance of the LPME scale based on academic achievement of participants using multiple-group confirmatory factor analysis. This is to ensure that group differences, if they exist, are interpreted in terms of differences regarding participants' academic achievement relative to the underlying constructs of the LPME scale.

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 (Byrne & Van de Vijver, 2010). That is, there is a presumed equality of (1) factorial structure, (2) perceived item content, (3) factor loadings and (4) item intercepts. The scientific reality is that this assumption of measurement invariance has to be tested first before any conclusions can be drawn in this regard. 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.

Test of measurement invariance examines whether an instrument has the same psychometric properties across heterogeneous groups (Chen, 2007). According to Brown (2006) and Meredith (1993), appropriate comparisons of group means rest on the assumption of configural and metric invariance, as well as 'scalar' or what is sometimes referred to as 'strong' invariance. Evidence of these tests is fundamental to establish the overall measurement invariance and conceptual interpretation of an instrument. Chen (2007) points out that when groups are compared based on instruments that do not measure the same construct, inference problems occur. In other words, the conclusions drawn from such studies may be biased or invalid if measures that are relied upon do not have the same meanings across different groups.

Nevertheless, the literature points to various recommendations regarding the sequences of measurement invariance tests (Bollen, 1989; Byrne, Shavelson & Muthén, 1989; Cheung, 2008; Cheung & Rensvold, 2002; Drasgow & Kanfer, 1985;

Jöreskog & Sörbom, 1993; Little, 1997; Steenkamp & Baumgartner, 1998; Vandenberg & 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. An appropriate answer to the research question will depend on the corresponding level or levels of measurement invariance test, that is, whether the test is at configural, metric and/or scalar levels. The current study followed the procedure described by Meredith (1993) and Widaman and Reise (1997) in order to test a series of models to establish measurement invariance of the LPME scale across academic achievement of participants. The procedural levels tested are as follows.

The first level: Configural (form invariance)

This level requires that the same item must be associated with the same factor in each group; however, the factor loadings may differ across groups. Participants belonging to different groups are assumed to conceptualise the constructs the same way (Riordan & Vandenberg, 1994). According to Widaman and Reise (1997), this level indicates that similar, but not identical, latent constructs have been measured in the group. Only the extent to which the same number of factors and patterns (configuration) of fixed and freely estimated parameters holds across groups is of interest and thus no equality constraints are imposed (Byrne & Van de Vijver, 2010). In other words, for each group, the same model of hypothesised factorial structure is tested.

As suggested by Meredith (1993), where configural invariance exist, the data collected from each group breaks down into the same number of factors, with similar items associated with each factor. However, when concepts are abstracted such that participants' perceptions of the construct depend on their cultural context, configural non-invariance manifest itself (Tayeb, 1994). Equally, Millsap and Everson (1991), Millsap and Hartog (1988) and Riordan and Vandenberg (1994) indicate that configural non-invariance manifest itself when participants from different groups use different conceptual frames of reference and attach different meaning to constructs.

This level of invariance is important in that it serves as a baseline against which all subsequent tests for equivalence are compared and, thus, acceptable goodness-of-fit between this test and the multi-group data is imperative. The next levels of tests for invariance involve the specification of cross-group equality constraints for particular parameters (Byrne & Van de Vijver, 2010).

The second level: Factor loading (weak invariance)

At this level, it is hypothesised that all factor loading parameters are equal across groups. According to Bollen (1989) and Jöreskog and Sörbom (1999) factor loadings represent the strength of the linear relationship between each factor and its associated items. As soon as the factor load of each item on the underlying factor shows equality in more than two groups, the resultant effect is that the underlying

factor has the same unit or same interval. Chen (2007) suggests that this level of invariance is required for comparison of regression slopes. However, conceptual agreement regarding the type and number of underlying constructs and the items associated with each construct may be obtained from data originating from samples drawn from two populations (Cheung & Rensvold, 2002). Despite this, there may be differences in the strength of the relations between specific scale items and the underlying constructs. In this instance, disagreement regarding how the constructs manifested may show from the data.

The third level: Intercept (strong invariance)

At this level, it is assumed that the vectors of item intercepts are equal across groups. According to Chen (2007), intercepts represent the foundation of the scale. This level of invariance is achieved when the scores from different groups depict the same factor loading as well as the same intercept. Widaman and Reise (1997) indicate that this level of invariance is required for comparing latent mean differences across groups.

When the item slopes (factor loadings) and item intercepts are both invariant, the measurement scales share the same operational definition including the same interval and same zero points across groups, which further suggests that meaningful comparison of the latent means can be achieved (Cheung & Lau, 2011).

The fourth level: Residual invariance (strict invariance)

At this level, it is hypothesised that residual invariance is equal across groups. Residual invariance is the portion of item variance not attributable to the variance of the associated latent variable (Cheung & Rensvold, 2002). Therefore, testing for the equality of between groups residual variance determines if the scale items measure the latent construct with the same degree of measurement error. When this level of invariance holds, all group differences on the items occur owing to group differences on the common factors (Chen, 2007). Residual invariance may fail when participants belonging to one group, compared with those of another, are unfamiliar with a scale and its scoring formats, and therefore respond to it inconsistently (Mullen, 1995). Furthermore, Malpass (1977) states that differences in vocabulary, idioms, grammar, syntax and the common experiences of different cultures may produce residual non-invariance.

In this regard, configural, weak, strong and strict invariance are the most commonly tested forms of invariance and these were applied in the current study. However, in view of the complexity surrounding measurement of invariance testing, Chen (2007) offered the following guidelines to establish model fit:

> The cut-off points on the three routinely used fit indexes (i.e., Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR)) are recommended for evaluating invariance at the three commonly tested levels (configural, factor loading and intercepts). The fourth level, residual invariance has been included in this study purely to assess the influence of measurement error on participants' interpretation of the conceptual base and underlying constructs of the LPME scale. (p. 501)

According to Chen (2007, p. 501), when the sample size is small ($n = \leq 300$), sample sizes for sub-groups are unequal and the pattern of non-invariance is uniform, the following cut-off criteria are suggested: for testing loading invariance, a change of ≤ -0.005 in CFI, supplemented by a change of ≥ 0.010 in RMSEA or a change of ≥ 0.025 in SRMR would indicate non-invariance; for testing intercept or residual invariance, a change of ≤ -0.005 in CFI, supplemented by a change of ≥ 0.010 in RMSEA or a change of ≥ 0.005 in SRMR would indicate non-invariance. Similar values are suggested for CFI and RMSEA across all four levels of invariance tests, but different values are proposed for SRMR since SRMR is more sensitive to non-invariance in loadings than to non-invariance in intercepts or residual variances (Chen, 2007).

When sample size is adequate ($n = > 300$) and sample sizes are equal across the sub-groups, particularly when lack of invariance is mixed, more stringent criteria are suggested. For testing loading invariance, a change of ≥ -0.010 in CFI (Cheung & Rensvold, 2002), supplemented by a change of ≥ 0.015 in RMSEA or a change of ≥ 0.030 in SRMR would indicate non-invariance; for testing intercept and residual invariance, a change of ≥ -0.010 in CFI (Cheung & Rensvold, 2002), supplemented by a change of ≥ 0.015 in RMSEA or a change of ≥ 0.010 in SRMR would indicate non-invariance. Chen (2007) advised that caution must be exercised when applying these criteria owing to complexity surrounding measurement invariance. The current study had adequate total sample size ($n = 369$) but was constrained by the uneven academic achievement sub-group sizes (matric equivalent and below, $n = 276$; above matric, $n = 93$). In view of this, the researcher straddled between the two sets of criteria to overcome the sample sub-group size anomaly.

Method
Research approach

A quantitative approach was applied which followed a non-experimental, cross-sectional survey design. In order to achieve the study's objective, primary data were collected from two metropolitan municipalities in Gauteng Province and a provincial government department in the North West Province, South Africa.

Participants

Participants in this study comprised human resource development (HRD) practitioners and learners or apprentices from two metropolitan municipalities in Gauteng Province and a provincial government department in the North West Province. These participants were selected from their organisations through a probabilistic simple random sampling technique. Out of a target of 900 participants, a total of 579 completed questionnaires were returned, thus yielding a 64% response rate. The returned questionnaires

were subjected to a stringent data management process in order not to derail from the objective of this study. The first round of data management resulted in 187 questionnaires that were discarded as they had missing data. This was done in order to comply with the Analysis of Moment Structures (AMOS) software requirement for computation of modification indices. At the end of this first round, a total of 392 questionnaires were retained for the subsequent round of data management.

The second round of data management was carried out as necessitated by the focus and objective of the current study which sought to test measurement invariance of participants across academic achievement. Only questionnaires that had an indication of the level of academic achievement were retained during this second round. Consequently, about 23 questionnaires were eliminated as participants who completed them did not indicate their level of academic achievement. The final pool comprised 369 questionnaires that were split between two categories of academic achievement (Group 1 – matric level or equivalent and below [n = 276]; Group 2 – above matric level [n = 93]). A matric level certificate is a school leaving certificate. About 86% of the participants were younger than age 35 and 56% were male. In terms of academic achievement, 74% of the participants had acquired a matric level certificate, its equivalent or below. While 84.9% of participants had exposure to learnerships, 79% were learners or apprentices.

The disparity in the sample size for academic achievement was anticipated in this study since the target beneficiaries of OLPs are young people (learners or apprentices) who are mainly at the lower levels in their academic pursuits. Participants who have reported educational achievement above matric level were mainly HRD practitioners who are one of the key stakeholders in the occupational learning context.

Measuring instrument

The 11-dimensional LPME scale developed by Tshilongamulenzhe (2012), comprising 81 items, was used to collect data for this study. The 11 dimensions of the LPME scale were: administrative processes (AP), environmental scanning (ES), monitoring and evaluation (ME), observation and problem solving (OPS), policy awareness (PA), quality assurance (QA), stakeholder inputs (SI), strategic leadership (SL), learning programme design and development (LPDD), occupational competence (OC) and learning programme specifications (LPS). The LPME scale had achieved an acceptable Cronbach's alpha coefficient of 0.86 during the exploratory factor analysis phase (Tshilongamulenzhe *et al.*, 2013) and a 0.87 during the confirmatory factor analysis phase. The reliability coefficient for the 11 LPME dimensions ranged from 0.82 to 0.93.

Research procedure

Permission to undertake this study was sought from the three participating organisations. The researcher collected data using a self-administered questionnaire which was distributed through a drop-in and pick-up method. The questionnaire had clear instructions and telephone numbers of the research team in case participants would need further clarity at the point of completion. In addition to this, a short briefing session was held with each participant at the point of questionnaire drop-in to ensure that they understood the purpose of the study, their rights as well as the instructions on how to complete the questionnaire.

Statistical analysis

Data obtained in this study were analysed using Statistical Package for Social Sciences (SPSS) (IBM, 2013) and AMOS software (versions 21.0) (Arbuckle, 2013). In view of the study's objective, descriptive statistics (mainly frequencies), scale reliability analysis, as well as multiple-group confirmatory factor analysis were computed.

Results

Table 1 depicts the results of the goodness-of-fit index for both the total sample and academic achievement sub-groups in this study. The CFI (0.980) of the total sample shows a good fit for the model to the empirical data whereas the RMSEA (0.078) shows an acceptable fit. All four models tested for invariance had a good CFI (ranging from 0.980 to 0.984) and a good RMSEA (ranging from 0.055 to 0.041). The SRMR of 0.022 across all models and the total sample is indicative of a good fit. Taken together, these results support measurement invariance of the LPME scale and show that the model provides a good fit to the data.

The ΔCFI (measurement weight = 0.001; structural covariance = 0.000 and measurement residual = 0.003), ΔRMSEA (measurement weight = -0.005; structural covariance = 0.000 and measurement residual = -0.009) and ΔSRMR (measurement weight = 0.000; structural covariance = 0.000 and measurement residual = 0.000) across all successive models after items were constrained met the cut-off criteria suggested by Chen (2007) and showed strong invariance between the two sub-groups. These results support the strong factor loading, intercepts and residual invariance of the LPME scale across academic achievement.

The results of the standardised regression weights (z), squared multiple correlations (R^2) and Cronbach's alpha coefficient (α) are depicted in Table 2.

It is clear from the table that all sub-scales of the LPME scale had a high reliability coefficient during the confirmatory factor analysis phase with α values ranging from 0.82 to 0.93. The results in Table 2 depict that all the sub-scales of the LPME scale are good indicators of the effectiveness of management and evaluation practices for OLPs. PA, programme design and development, stakeholders' inputs and occupational competence are the best indicators of OLP, and their standardised regression weights are 0.838, 0.914, 0.901 and 0.803, respectively. This means that OLP explains about 70%, 83%, 81% and 64% of variance in PA, programme

TABLE 1: Measurement invariance by academic achievement.

Model	X^2	ΔX^2	df	Δdf	CFI	ΔCFI	RMSEA	PCLOSE	ΔRMSEA	SRMR	ΔSRMR
Total sample (n = 369)	118.324	-	31	-	0.980	-	0.078	0.002	-	0.022	-
Unconstrained	236.648	-	70	-	0.980	-	0.055	0.130	-	0.022	-
Measurement weights	236.648	0.000	80	10	0.981	0.001	0.050	0.483	-0.005	0.022	0.000
Structural co-variances	236.648	0.000	81	1	0.981	0.000	0.050	0.525	0.000	0.022	0.000
Measurement residuals	236.648	0.000	101	20	0.984	0.003	0.041	0.980	-0.009	0.022	0.000

df, degrees of freedom; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation; PCLOSE, Closeness-of-fit statistic; SRMR, Standardized Root Mean Square Residual.

TABLE 2: Regression weights, squared multiple correlations and Cronbach's alpha.

Observed indicator or Sub-scale	Relationship	Latent variable	Estimate	SE	C.R.	z	R^2	α
Policy awareness	<---	OLP	1.000	-	-	0.838	0.702	0.89
Observation and problem solving	<---	OLP	0.700	0.027	25.728	0.859	0.737	0.85
Quality assurance	<---	OLP	0.414	0.023	17.984	0.761	0.579	0.85
Administrative processes	<---	OLP	0.575	0.027	21.102	0.845	0.714	0.82
Programme design and development	<---	OLP	1.722	0.071	24.098	0.914	0.835	0.91
Stakeholders inputs	<---	OLP	1.808	0.077	23.437	0.901	0.812	0.93
Strategic leadership	<---	OLP	0.476	0.027	17.537	0.804	0.647	0.85
Learning programme specifications	<---	OLP	0.372	0.019	19.770	0.707	0.500	0.85
Monitoring and evaluation	<---	OLP	0.579	0.030	19.283	0.798	0.637	0.85
Occupational competence	<---	OLP	1.031	0.053	19.549	0.803	0.645	0.88
Environmental scanning	<---	OLP	0.700	0.033	20.923	0.838	0.702	0.90

OLP, occupational learning programme.

design and development, stakeholders' inputs and occupational competence, respectively. Therefore, the null hypothesis that the model is not a good fit to the data is easily rejected.

A close inspection of these results reveals that the conceptual foundation and factorial structure of the LPME scale as proposed by Tshilongamulenzhe (2012) and further reported by Tshilongamulenzhe *et al*. (2013) are invariant across a range of academic achievements in the occupational learning context.

Discussion

The main objective of this study was to test the measurement invariance of the LPME scale across academic achievement of participants. Following multiple-group confirmatory factor analysis, the findings of this study support the measurement invariance of the 11-dimensional LPME scale, and these findings meet the criteria suggested by Cheung and Rensvold (2002) and Chen (2007). As Byrne (2008) stated, goodness-of-fit related to multi-group parameterisation as carried out in this study are indicative of a well-fitting model. The equivalence of factors related to the LPME scale and their related items relative to academic achievement have been positively tested in this study.

The findings show an acceptable fit of the models to the empirical data both for the total sample and for the sub-groups as identified by levels of academic achievement. Given the paucity of previous studies in South Africa focusing on the effectiveness of management and evaluation practices related to OLPs (except a study reported by Tshilongamulenzhe *et al*., 2013), this study is very profound as it lays a solid foundation for future studies that will further examine the phenomenon of occupational learning. The study succinctly

followed the measurement invariance testing procedure outlined by Meredith (1993) and Widaman and Reise (1997). The configural model was used as a baseline against which all subsequent invariance tests were compared with. Metric invariance which is a strong test (Vandenberg & Lance, 2000) was conducted in order to guarantee that the factor loadings between factors and indicators are similar across groups, and the results from this test support metric invariance of the LPME scale. The results of both the scalar and residual invariance tests also support the invariance of the LPME scale. Thus, the observed scores are related to the latent scores in such a way that group differences in means were meaningfully compared.

Overall, the findings of this study suggest that the same constructs are measured across groups and that the units and origins of the LPME scale are the same. These findings provide adequate validity evidence regarding the factorial structure and measurement properties of the LPME scale.

Limitations of the study

Like other studies, this study has some limitations. The first limitation relates to the generalisability of the findings to the relevant population. It would be inappropriate to assume invariance of the LPME scale across other demographic indicators that have not been tested in this study. Therefore, the findings of this study must be limited to academic achievement of the relevant population from which the sample was drawn. The second limitation relates to the unequal sub-group sizes pertaining to academic achievement, which may have influenced the magnitude of change in fit statistics. Chen (2007) uncovered a number of factors that researchers may need to be wary of when testing measurement invariance, namely pattern of non-invariance, sample size, ratio of sample size and model complexity. The current study

is mainly concerned with the ratio of academic achievement sub-groups sample size.

Implications and recommendations for future research

This study has the following implications:

- It will provide an opportunity for empowerment to management scientists and methodologists in the sub-field of training management or HRD by laying a solid and scientifically tested foundation which should ignite renewed thinking about the occupational learning construct in order to propel the necessary critique and modification of this new instrument.
- It will support researchers in the sub-field of training management or HRD with a scientifically developed and validated measure which can be practically applied to a relevant population with different levels of academic achievement.

In conclusion, the invariance assumption regarding the LPME scale across different levels of academic achievement has been successfully tested scientifically in this study. To this end, it can be confidently stated that participants in this study ascribe the same meaning and understanding towards the LPME scale and its sub-scales irrespective of their levels of academic achievement.

Future studies may explore the measurement invariance of the LPME scale on other different but even population sub-groups in order to ascertain the actual meaning that participants ascribe to the construct of an occupational learning programme and its attendant dimensions.

Acknowledgements

Competing interest

References

Adolf, J., Schuurman, N., Borkenau, P., Borsboom, D., & Dolan, V. (2014). Measurement invariance within and between subjects: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, 5, 883. http://dx.doi.org/10.3389/fpsyg.2014.00883

Arbuckle, J.L. (2013). *IBM SPSS Amos 21 user guide*. New York: Amos Development Corporation.

Blankson, A.N., & McArdle, J.J. (2015). Measurement invariance of cognitive abilities across ethnicity, gender and time among older Americans. *Journal of Gerontology, Series B: Psychological Sciences and Social Sciences*, 70(3), 386–397.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bowden, S.C., Saklofske, D.H., & Weiss, L.G. (2011). Invariance of the measurement model underlying the Wechler Adult Intelligence Scale-IV in the United States and Canada. *Educational and Psychological Measurement*, 71, 186–199. http://dx.doi.org/10.1177/0013164410387382

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.

Byrne, B.M. (2008). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. (2nd edn). New York: Routledge.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456–466. http://dx.doi.org/10.1037/0033-2909.105.3.456

Byrne, B.M., & Van de Vijver, F.J.R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of non-equivalence. *International Journal of Testing*, 10, 107–132. http://dx.doi.org/10.1080/15305051003637306

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14 (3), 464–504. http://dx.doi.org/10.1080/10705510701301834

Cheung, G.W. (2008). Testing equivalence in the structure, means and variances of higher-order constructs with structural equation modelling. *Organizational Research Methods*, 11, 593–613. http://dx.doi.org/10.1177/1094428106298973

Cheung, G.W., & Lau, R.S. (2011). A direct comparison approach for testing measurement invariance. *Organizational Research Methods*, 1–32.

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9 (2), 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Chiorri, C., Day, T., & Malmberg, L.E. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Frontiers in Psychology*, 5, 983. http://dx.doi.org/10.3389/fpsyg.2014.00983

DeVellis, R.F. (2012). *Scale development: Theory and application.* (2nd edn). Thousand Oaks, CA: Sage.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70, 662–680. http://dx.doi.org/10.1037/0021-9010.70.4.662

Du Toit, R. (2012). *The NSF as a mechanism to address skills development of the unemployed in South Africa*. Pretoria: HSRC.

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology*, 5, 980. http://dx.doi.org/10.3389/fpsyg.2014.00980

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research Special Issue: Quantitative Topics in Research on Aging*, 18, 117–144. http://dx.doi.org/10.1080/03610739208253916

Hox, J.J., De Leeuw, E.D., & Zijlmans, E.A.O. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology*, 6, 87. http://dx.doi.org/10.3389/fpsyg.2015.00087

IBM. (2013). *IBM SPSS statistics 21 brief guide.* New York: IBM Corporation.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.

Jöreskog, K., & Sörbom, D. (1999). *LISREL 8: User's reference guide.* (2nd edn.). Chicago, IL: Scientific Software International.

Little, T.D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioural Research*, 32, 53–76. http://dx.doi.org/10.1207/s15327906mbr3201_3

Malpass, R.S. (1977). Theory and method in cross-cultural psychology. *American Psychologist*, 32, 1069–1079. http://dx.doi.org/10.1037/0003-066X.32.12.1069

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. http://dx.doi.org/10.1007/BF02294825

Millsap, R.E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, 26, 479–497. http://dx.doi.org/10.1207/s15327906mbr2603_6

Millsap, R.E., & Hartog, S.B. (1988). Alpha, beta, and gamma changes in evaluation research: A structural equation approach. *Journal of Applied Psychology*, 73, 574–584. http://dx.doi.org/10.1037/0021-9010.73.3.574

Mullen, M. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, 3, 573–596. http://dx.doi.org/10.1057/palgrave.jibs.8490187

Mummenthey, C., Wildschut, A., & Kruss, G. (2012). *Assessing the impact of learnerships and apprenticeships under NSDSII: Three case studies: MERSETA, FASSET & HWSETA*. Pretoria: HSRC.

Ployhardt, R.E., & Oswald, F.L. (2004). Application of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*, 7(1), 27–65. http://dx.doi.org/10.1177/1094428103259554

Republic of South Africa. (2008). *Skills Development Amendment Act, Act No. 37 of 2008*. Pretoria: Government Printers.

Riordan, C.M., & Vandenberg, R.J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management*, 20, 643–671. http://dx.doi.org/10.1177/014920639402000307

Savage-McGlynn, E. (2012). Sex differences in intelligence in younger and older participants of the Raven's Standard Progressive Matrices Plus. *Personality and Individual Differences*, 53, 137–141. http://dx.doi.org/10.1016/j.paid.2011.06.013

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210–222. http://dx.doi.org/10.1016/j.hrmr.2008.03.003

Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90. http://dx.doi.org/10.1086/209528

Tayeb, M. (1994). Organizations and national cultures: Methodology considered. *Organization Studies*, 15, 429–446. http://dx.doi.org/10.1177/017084069401500306

Thompson, M.S., & Green, S.B. (2006). Evaluating between-group differences in latent means. In G.R. Hankok & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 119–169). Greenwich: Information Age.

Tshilongamulenzhe, M.C. (2012). *An integrated learning programme management and evaluation model for the South African skills development context*. Unpublished doctorate thesis, University of South Africa, Pretoria.

Tshilongamulenzhe, M.C. (2015). Testing measurement invariance of the Learning Programme Management and Evaluation (LPME) scale across gender using Multi-group Confirmatory Factor Analysis. *Journal of Governance and Regulation*, 4(1), 124–131.

Tshilongamulenzhe, M.C., Coetzee, M., & 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*, 39(1), 1–14. http://dx.doi.org/10.4102/sajip.v39i1.1080

Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4, 770. http://dx.doi.org/10.3389/fpsyg.2013.00770

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). Developmentrics: A checklist for testing measurement invariance. *European Journal of Developmental Psychology,* 9(4), 486–492. http://dx.doi.org/10.1080/17405629.2012.686740

Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6, 1064. http://dx.doi.org/10.3389/fpsyg.2015.01064

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practice and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. http://dx.doi.org/10.1177/109442810031002

Wang, T., Merkle, E., & Zeileis, A. (2014). Score-based tests of measurement invariance: Use in practice. *Frontiers in Psychology*, 5, 438. http://dx.doi.org/10.3389/fpsyg.2014.00438

Widaman, K.F., & Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K.J. Bryant, M. Windle & S.G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Wildschut, A., Kruss, G., Janse Van Rensburg, D., Haupt, G., & Visser, M. (2012). *Learnerships and apprenticeships survey 2010 technical report: Identifying transitions and trajectories through the learnership and apprenticeship systems*. Pretoria: HSRC.

Yen, W.K.S., & 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*, 12(2), 123–139.

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact versus approximate measurement equivalence. *Frontiers in Psychology*, 6, 733. http://dx.doi.org/10.3389/fpsyg.2015.00733