DEVELOPING AND VALIDATING A MEASURE OF MONITORING AND EVALUATION FOR THE SOUTH AFRICA SKILLS DEVELOPMENT CONTEXT

Maelekanyo Christopher TSHILONGAMULENZHE

University of South Africa, Pretoria, South Africa
e-mail: tshilmc@unisa.ac.za

Abstract: As nations around the world grapple with the deepening challenge of skills crunch, policies, strategies and interventions are being devised to develop and retain a competitive workforce. South Africa is not spared from the global talent war, and the persistence of skills shortages across most economic sectors makes the country vulnerable economically and socially. Legislative instruments and strategies were formulated and implemented to ameliorate the situation, but scientific evidence suggests a looming crisis due to poor monitoring and evaluation systems regarding the success of these instruments and strategies. This study develops and validates a monitoring and evaluation (ME) scale for the South African skills development context. A quantitative nonexperimental cross-sectional survey design was used to collect data from 557 participants. Data were analyzed using SPSS and AMOS software (version 23.0). The findings show that the ME scale is scientifically valid and reliable and can be used with confidence in the South African skills development context. The findings provide scope for a validation study on an independent sample, and an evaluation of structural invariance of the ME scale across sample subgroups.

Keywords: skills development, occupational learning, learnership, apprenticeship, monitoring, evaluation,

1   Introduction

The skills development battlefield is a complex, ruthless and demanding environment driven by environmental forces, policies, international trends and obligations, lifelong learning principles, human capital investments and effective training provision benchmarks. South Africa is experiencing a persistent challenge of skills shortage, which is affecting the level of economic productivity and is reducing the country's capacity to develop a knowledge society [1]. The South African government is giving the issue of skills shortages considerable attention by placing skills development at the core of socio-economic development strategy [2]. The enactment of the Skills Development Act 97 of 1998 (as amended in 2008) marks a significant policy decision to transform the skills of the South African workforce. The Act provides for the establishment of an institutional architecture to coordinate skills development at the national and sectoral levels; and for the establishment of learning programs. The goals of this legislation have to be monitored and implemented through a National Skills Development Strategy (NSDS) which is reviewed every 5 years (the current one from 2011 to 2016 being the third phase).

Developed as a cornerstone component of the skills development strategy, learning programs are supposed to provide workplace learning in a structured and systematic form; to link structured learning to multiple sites and to do so in a manner that results in a nationally-recognized qualification [2]. A learning program obligates the employer to employ the learner, to provide specified work experience and to allow the learner to attend education and training [3]. The compulsory practical work-based component aims to prepare learners for employability and to create closer synergy between education and the world of work thereby narrowing the gap between theory and practice and between classroom learning and work experience.

International experience has shown that in order to successfully implement skills training programs in the workplace, there must be processes for implementing, evaluating and monitoring such training to ensure that these programs continue to meet the skills needs of employers and society effectively [4, 5].

Thus, systematic monitoring and evaluation is the main tool to achieve effective evidence-based goals [6]. However, it appears that before implementing the learning program model, South Africa did not evaluate the efficacy of such a model in the original countries. This is a frequent omission in cases of policy emulation [7].

Learning programs in South Africa are administered by 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. There are 21 existing SETAs in South Africa currently, each being responsible for the implementation of the NSDS in its own economic sector. SETAs have experienced considerable difficulties and problems to date – such as the complexities of the implementation of the NSDS, lack of capacity in monitoring and evaluation and a lack of or poor quality assurance and management information system [8, 9].

Studies regarding the effectiveness of learning programs in South Africa have found problems inherent at both the conceptualization and implementation levels [10, 11]. A study by Kraak [12] found that little monitoring and evaluation work has been done on the success and impact of learnerships in South Africa. In their study, Karlsson and Berger [13] found that the ETDPSETA had not established a system to effectively monitor learners participating in the learnership program. There was no tracking system to enable stakeholders to monitor and evaluate learners' progress or to alert them when learners' experience problems.

In addition, Smith, Jennings and Solanki [14] found little mention of either monitoring or evaluating learnerships by SETAs. They also reported that other SETAs have failed woefully; are not in a position to either administer the learnership or monitor the performance of learners during the skills development phase; and that they have set aside insufficient time to critically reflect on whether the learnerships are in fact achieving their overall objectives. Accordingly, SETAs are going to need to develop far more complex monitoring and evaluation systems to assess progress toward achieving the learning program objectives and outcomes [14]. To this end, this study seeks to contribute towards

an effective learning program monitoring and evaluation regime in South Africa by developing a monitoring and evaluation (ME) scale, which could be used by stakeholders in the occupational learning context.

2   Theoretical perspectives regarding monitoring and evaluation

According to Cracknell [15], the term "monitoring and evaluation" has come into common usage over the last three decades. The mainstream position is that monitoring is an ongoing process of data capture and analysis for the purpose of control whereas evaluation is a periodic process of assessment for the purpose of learning [16]. Monitoring has an internally focused, management-driven emphasis on the efficiency of the project, while evaluation primarily has an externally focused, stakeholder-driven emphasis on the effectiveness of the project.

Monitoring and evaluation is not just a question of whether people completed an apprenticeship or traineeship and obtained any relevant qualification. It is also a matter of the skills they have acquired that build on the skills they already possessed before they took part [17]. According to Lange and Luescher [18] for monitoring and evaluation to have a function beyond mere accountability and resource allocation, they have to transcend the generation of baseline data and venture into the more complicated and contested terrain of explanation. In order to achieve this, monitoring and evaluation systems need to be deeply embedded in the socio-economic and socio-political dynamics of the societies in which organizations operate both at the conceptual and the design level.

The OECD/DAC [19] defines evaluation as "the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results". Monitoring is defined as "a continuous management function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an on-going development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated resources" [19]. Monitoring is thus

rather descriptive and assesses whether different levels of an intervention (inputs, activities, outputs, outcomes and impact) are realized as expected. According to UNICEF [20] monitoring is "the periodic oversight of the implementation of any activity which seeks to establish the extent to which input deliveries, work schedules, other required actions and targeted outputs are proceeding according to plan so that timely action can be taken to correct deficiencies detected." Evaluation is "a process which attempts to determine as systematically and objectively as possible the relevance, effectiveness, efficiency and impact of activities in the light of specified objectives. It is a learning and action-oriented management tool and organizational process for improving the current activities and future planning, programming and decision making" [20]. In this conceptualization, monitoring and evaluation represent qualitatively different yet complimentary activities.

While monitoring is viewed as the continuous observation of an activity and aims to identify the need for corrective action by measuring change (input, output, process, instruments) over time, evaluation is preoccupied with the interpretation of monitoring data, and the attempt to discern, explain and assess change patterns and causalities [21]. There is a blurring line between monitoring and evaluation, especially if the object of monitoring is not simply to produce data regularly but to produce intelligence on data, which in turn has some role in pointing to incipient problems and trends that cannot be simply signalled but need some level of explanation and interpretation. Explanation and interpretation in a monitoring and evaluation system imply that the system itself has to be embedded in social dynamics [18].

Evaluation is a complement to monitoring in that when a monitoring system sends signals that the efforts are going off track then good evaluative information can help clarify the realities and trends noted with the monitoring system [19]. According to Annecke [22], monitoring and evaluation are conducted for two major reasons: (1) to track implementation progress; (2) is to learn from the experience and use this learning to design future projects better. Monitoring usually implies a continuing operation conducted by organisations during project imple-

mentation to ensure that the project stays on track to achieve its objectives. Thus, monitoring exercise may be used to improve program efficacy as it is implemented: the project should be flexible and able to change and adapt to conditions on the ground as indicated by the exercise. Evaluation is a periodic assessment of the relevance and performance of the project.

Given that there is little conceptual difference among prevailing monitoring and evaluation approaches, there is a great opportunity to work towards common standards for how to do project and program monitoring and evaluation and to agree on common indicators for measuring key variables. Monitoring is a continuous assessment of the functioning of project activities, as compared with implementation schedules, the use of project inputs by the target population, and the effects of the project as measured by physical, social or biological indicators [23]. The objectives of monitoring are to inform interested parties about the performance of the project, to adjust project development, to identify measures that can improve project quality, to make the project more cost effective, to improve planning and measuring processes, and to be part of a learning process for all involved stakeholders. Evaluation is a periodic assessment of the relevance, performance, efficiency and impact of the project in the context of its stated objectives. Evaluation organizes and appraises the information collected by the monitoring procedures, compares this information with information collected in other ways and presents the resulting analysis of the overall performance of a project at a time and place convenient to stakeholders so that they can make decisions about (1) whether to continue the project, (2) to compare the performance of different projects, (3) to make changes in the project design, and (4) to make major changes in the project's management [23]. As with other areas of the management system, the procedure for monitoring and evaluation should be constantly subject to improvement and refinement.

Monitoring and evaluation can provide public and internal accountability and help demonstrate impact [24, 25]. Monitoring and evaluation answer questions related to how well a project or strategy is working and identifies the conditions under which

an action is likely to succeed or falter [26, 27]. It can also serve as an early warning system for potential problems and lead to ideas for potential remedial actions [28].

Evaluation approaches vary by context and stakeholder interests, thus serving multiple and often overlapping purposes, including knowledge generation, programme improvement, accountability, transparency, resource allocation, advocacy and impact assessment. Mark, Henry and Julnes [29] developed an evaluation framework that identified different purposes, among others, basic research, accounting and certification, status assessment, and effective measurement. At the learning programme level, monitoring and evaluation is most important for determining whether or not the interventions are effective [30]. Apprenticeships and traineeships involve complex governance arrangements and consideration needs to be given to how monitoring and evaluation functions in this context [17]. Monitoring and evaluation needs to be part of quality assurance systems for apprenticeships and traineeships. The NQI emphasizes the importance of monitoring and evaluation of the progress made toward meeting the goals of the organization [31]. In Singapore, monitoring and evaluation are the key aspects of the success of apprenticeships because this process begins when the program commences [32]. In South Africa, the level of monitoring and evaluation of occupational learning programs must be enhanced and constant site visits may have to be introduced in this endeavour since this is an aspect in which weaknesses have been identified. The Quality Council for Trades and Occupations (QCTO) has to conduct research to monitor the effectiveness of learning interventions in the context of the broader occupational learning system. The process of monitoring and evaluation should revolve around the development and design processes, the implementation of occupational learning programs and data analysis and impact assessment (qualitative and quantitative) [33]. SETAs have to focus on monitoring and evaluation of the implementation of occupational learning programs in line with the Department of Higher Education and Training (DHET) regulations [34]. It is therefore imperative that adequate monitoring and evaluation be conducted constantly to ensure

proper conceptualization, development and implementation of occupational learning programs in South Africa, hence this study that seeks to develop a monitoring and evaluation (ME) scale.

3   Research design

3.1   Research approach

This study followed a quantitative, non-experimental, cross-sectional survey design and primary data were collected from five Sector Education and Training Authorities (SETAs) and a human resource (HR) professional body in South Africa.

3.2   Research participants

A total of 557 participants were drawn from six organizations (five SETAs and a human resource (HR) professional body) in South Africa using a probabilistic simple random sampling technique. These participants were diverse in their occupational status and included learning or training managers/employers, mentors/supervisors of learners/ apprentices, skills development officers/providers, learning assessors/moderators as well as learners/apprentices. Adequate knowledge and understanding of the South African skills development context, including the new occupational learning system was very critical to the selection of participants into the sample. The majority of participants (77%) were young people who were aged 35 years or younger. About 52% of participants were females. Fifty-six percent of the participants had achieved a Senior Certificate as the highest qualification. Only 28% of the participants had a three years bachelor's degree or diploma as their highest qualification. Seventy-nine percent of the participants were involved in learnerships. Occupationally, 62% of the participants were learners/apprentices, while 12% comprised manager/employers and mentors/supervisors. The remaining percentage was distributed across skills development providers/officers and assessors/facilitators.

3.3   Measuring instrument

A five-item monitoring and evaluation (ME) scale developed by Tshilongamulenzhe [35] was used for data collection. Responses were measured on a 6-point Likert scale ranging from (1) "Strongly agree"

to (6) "Strongly disagree." Sample items included "SETAs must monitor the performance quality of skills development providers" and "Regular contact by SETAs with employers and learners must occur."

3.4   Research procedure

The researcher obtained permission to conduct this research from five SETAs and an HR professional body. After receiving approval, the researcher started the process of planning for sampling and data collection with the respective organizations. The fieldwork took place in the provinces of Gauteng, North West and Mpumalanga in South Africa.

3.5   Statistical analysis

Statistical Package for Social Sciences (SPSS, Version 23) and Analysis of Moment Structures (AMOS, Version 23) [36] were used to analyse the data. Both the exploratory and confirmatory factor analyses were computed. Exploratory factor analysis (EFA) was executed using SPSS and AMOS was used to conduct structural equation modeling as part of confirmatory factor analysis (CFA). Exploratory factor analysis focuses on whether the covariance or correlations between a set of observed variables can be explained in terms of a smaller number of unobserved constructs known either as latent variables or common factors [37]. The analysis is often used to gather information about inter-relationship among the set variables. Tabachnick and Fidell [38] give the following advice regarding sample size for exploratory factor analysis: 50 is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or more is excellent. In the current research, a sample size comprising 557 cases was considered very adequate for factor analysis.

First, factor analysis was computed in the current study to test the suitability of data for further analysis. An inspection of a correlation for a coefficient above 0.30 was done followed by the calculation

of sample adequacy (Kaiser-Meyer-Olkin [KMO] and Bartlett's test for Sphericity). Suitability of KMO sampling is connected with the suitability of the correlations among the scale items. The values vary between 0 and 1, and values closer to 1 are better. The suggested minimum value that is acceptable for further analysis is 0.60 [38]. If KMO value is high, Bartlett's test becomes statistically significant. Second, the factorial structure of the ES scale was tested through the execution of principal components analysis. A principal components analysis is essentially a method of data reduction that aims to produce a small number of derived variables that can be used in place of the larger number of original variables to simplify subsequent analyses of data [37]. Construct validity was tested using structural equation modelling, a CFA technique that is applied to estimate, analyze and test models that specify relationships among variables [39]. A model is established at the beginning and thereafter tested to ascertain whether it is supported by the data obtained. The factor structure obtained during EFA was tested through CFA.

4   Results

As depicted in Table 1, the Kaiser-Meyer-Olkin (KMO) index of .779 indicates that the items in the monitoring and evaluation (ME) scale are very suitable for factor analysis [40], and therefore, the factorial structure to be obtained from the principal component analysis (PCA) will be acceptable. KMO is a measure of how much the items have in common. A KMO value closer to 1 indicates that the variables have a lot in common. The Bartlett's test of sphericity was also conducted to test the null hypothesis that "the correlation matrix is an identity matrix." An identity matrix is a matrix in which all the diagonal elements are 1 and off-diagonal elements are 0.

Table 1. KMO and Bartlett's Test values

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.779
Bartlett's Test of Sphericity	Approx. Chi-Square	789.831
	df	10
	Sig.	.000

As depicted in Table 1, the Bartlett's test of sphericity was statistically significant (Approx x^2 = 789.831; df. 10; p ≤ .000) and thus the null hypothesis that "the correlation matrix is an identity matrix" was rejected. The determinant of the correlation matrix between the factors was set to zero due to orthogonal rotation restriction, which imposes the condition that the factors cannot be correlated. Taken together, the results of these tests meet a minimum standard that should be passed before a PCA is conducted. A single-factor structure was obtained during the PCA (using varimax rotation) and the results are depicted in Table 2. All items of the ME scale yielded communality values above .66, which is by far higher than the .30 threshold suggested by Tabachnick and Fidell [38]. These results suggest that all items of the ME scale are working together in support of the monitoring and evaluation construct, and this proves the unidimensionality of the ME scale.

Table 2. Item factor load

Item	Factor load
1. SETAs must keep records of the registered occupational learning program agreement	.753
2. SETAs must oversee the implementation of occupational learning programs	.802
3. SETAs must monitor the performance quality of skills development providers	.770
4. Regular contact by SETAs with employers and learners must occur	.662
5. SETAs have the overall responsibility to monitor and evaluate the impact of occupational learning programs	.685

Extraction Method: Principal Component Analysis
1 component extracted

The scree plot representation shown in Figure 1 supports a single-factor structure obtained for the ME scale. The first item accounts for 54.20% of the total variance explained for the ME scale (with 2.71 eigenvalue units).
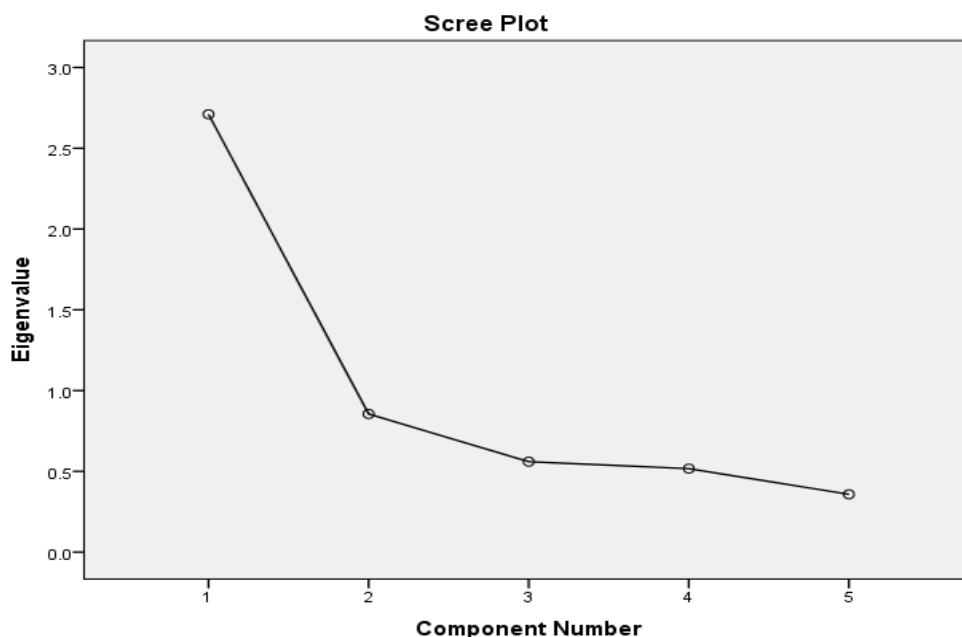

Figure 1. Scree plot for items of ME scale

4.1   Scale reliability

The ES scale was tested for its reliability and the results are depicted in Table 3. The item mean values ranged from 1.50 to 1.68 and that of the total scale was 8.04. The ES scale has yielded a good reliability coefficient of 0.78 as shown in Table 3.

Table 3. Item statistics and reliability coefficients for the ES scale (n = 557)

Items	Mean	Std. Dev	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correla-tion	Cronbach's Alpha if Item Deleted
1. SETAs must keep records of the registered occupational learning program agreement.	1.50	.726	6.54	6.954	.569	.429	.741
2. SETAs must oversee the imple-mentation of occupational learning programs.	1.67	.880	6.36	6.163	.623	.481	.718
3. SETAs must monitor the perfor-mance quality of skills develop-ment providers.	1.54	.785	6.50	6.617	.601	.368	.729
4. Regular contact by SETAs with employers and learners must occur.	1.68	.907	6.36	6.536	.496	.274	.763
5. SETAs have the overall responsi-bility to monitor and evaluate the impact of occupational learning programs.	1.64	.936	6.39	6.329	.521	.294	.756
Total scale (n = 5)	8.04	3.108					.782

4.2   Structural equation modeling

Structural equation modeling was computed to test multiple relationships in the current study as well as the construct validity of the ES scale. First, the standardized regression estimates were computed as depicted in Table 4. The standardized regression estimates for the model tested in this study ranged between .509 and .791, while the standard error coefficients ranged between .074 and .088 as depicted in Table 4. All five items were found to be signifi-cant predictors of the monitoring and evaluation construct (p ≤ .001).

Table 4. Regression weights for the items of ME scale

			Estimate	S.E.	Std. Regres-sion Estimate	C.R.	P
monitor	<---	MonitoringEvaluation	1.012	.074	.677	13.731	***
contact	<---	MonitoringEvaluation	.879	.083	.509	10.581	***
evaluate	<---	MonitoringEvaluation	.967	.086	.542	11.235	***
implementation	<---	MonitoringEvaluation	1.325	.088	.791	15.102	***
keep_records	<---	MonitoringEvaluation	1.000		.723		

A single-factor five-item model depicting the ME scale as obtained from the PCA was developed for CFA as shown in Figure 2 with calculated item-factor correlations. Path coefficients in this model ranged from .88 to 1.32. Kline [41] indicated that path coefficients with absolute values less than .10 could indicate a "small" effect, values around .30 could suggest a "typical" or "medium" effect, and a "large" effect could be indicated by coefficients with absolute values ≥ .50. In the current study, all the values were higher than .80, thus supporting large effect.
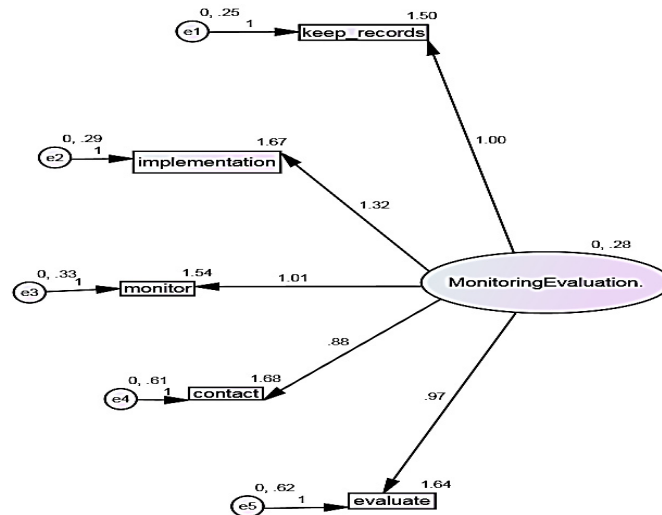
Figure 2. Initial hypothesised ME model

Several fit indices were computed in the current study to test the structural equation model against the data. The most widely used indices are the absolute fit indices (e.g., chi-square (x^2), standardized root mean square residual (SRMR)), relative fit indices (e.g., normed fit index (NFI), Tucker Lewis index (TLI), and incremental fit index (IFI)) and non-centrality-based indices (e.g., comparative fit index (CFI) and root mean square error of approximation (RMSEA)). The current study tested the data against all these fit indices and the results are depicted in Table 5.

Table 5. Fit indices for the initial structural equation model

Model	X^2	X^2/df	TLI	IFI	NFI	CFI	SRMR	RMSEA
Criteria for a good fit	>.05	≤ 3	≥ .95	≥ .95	≥ .95	≥ .95	≤ .08	≤ .06
Initial model	70.61	14.12	.83	.91	.91	.91	.06	.15

The relative chi-square (X^2/df) was found to be above the ≤ .3 value recommended by Kline [41]. The TLI was found to be very low (.83) and the other relative fit indices (IFI and NFI) were at an acceptable level, but below the ≥ .95 threshold recommended by Hu and Bentler [42]. The CFI was also found to be at an acceptable level, but also below the ≥ .95 threshold recommended by Hu and Bentler [42]. The SRMR value was at an acceptable level (.06) relative to ≤ .08 threshold recommended by Hu and Bentler [42] and Schreiber et al [43]. The RMSEA value was found to be extremely higher (.15) than the ≤ .06 threshold for a good fit as recommended by Hu and Bentler [42]. In view of the findings that the initial structural equation model for the ME scale does not adequately fit the data, the researcher examined the standardized regression weights, squared multiple correlations and modification indices in order to explore the sources of lack-of-fit so as to establish how the model can be respecified. In a confirmatory factor analysis, an initial model can be respecified in order to improve its goodness of fit, parsimony and interpretability [44]. Model respecification is based on modification indices (empirical evidence) and substantive justification (theoretical evidence). Trimming off variables with low loadings and correlating errors of variables are two general ways of model respecification [41]. Correlated errors are specified when some of the covariance across two variables is not explained by the latent construct [44]. Although correlated error can be specified according to modification indices, they need to be supported by a theoretical rationale.

After analyzing the initial model results, the researcher examined the standardized regression weights and squared multiple correlations of ob-

served variables in order to establish the source of inadequate fit. As can be seen in Figure 2, no variable was found with low standardized regression weight (less than .50) and/or low squared multiple correlations (less than .15) as suggested by Liu [45]. The item loadings (standardized regression estimates) for all variables in the initial model were statistically significant and higher than .50 (p < .001). Similarly, the correlation estimates depicted in Figure 2 for all the observed variables ranged from .25 to .62 and were also statistically significant. Thus, an examination of both the standardized regression weights and squared multiple correlations offered no clear suggestions for model respecification. The next avenue was to explore the modification indices. The results of modification

indices are depicted in Table 6. Modification indices are numerical estimates calculated in AMOS that indicate places in the model where paths could be added to improve model fit [46]. Specifically, by allowing for error terms with similar sources of measurement error to covary, the resulting observed scores are closer to respondents' "true" scores on those measures, and this will often improve model fit. Larger modification indices values (i.e., those greater than 4.0) are considered to be an indication that covariance paths should be considered; however, additional paths were specified only when it made sense theoretically [46]. As depicted in Table 6, priority had to be given to error covariance that is theoretically justifiable and that could yield a better parameter change.

Table 6. Modification indices

Error terms			M.I.	Par Change
e2	<-->	e1	17.291	.061
e5	<-->	e1	13.263	-.070
e4	<-->	e1	6.412	-.048
e4	<-->	e2	11.914	-.075
e4	<-->	e5	37.034	.170
e3	<-->	e4	7.718	.059

Consequently, e4 and e5 were covaried first (with marginal model improvement), followed by e1 and e2 (with substantial model improvement) as

depicted in Fig. 3. The path coefficients for the revised model ranged from 1.08 to 1.36.
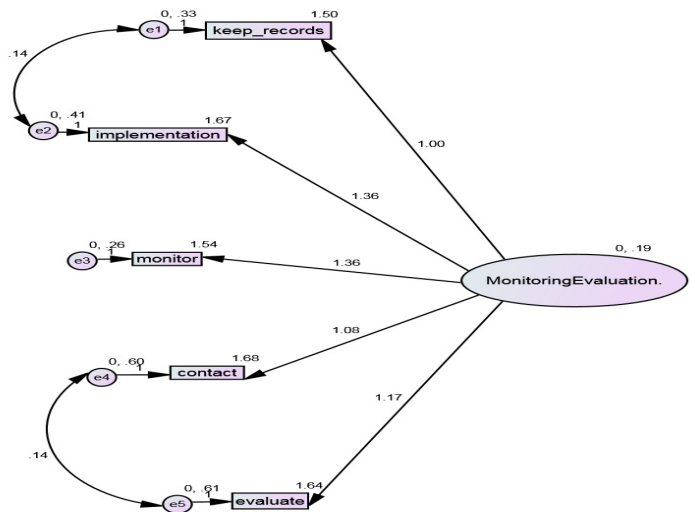

Figure 3. Revised hypothesized ME model

The results of the fit indices for the respecified model are depicted in Table 7. It is clear that the covariance allowed on error terms of applicable observed variables as suggested by the modification indices has yielded the best model fit.

Table 7. Fit indices for the revised structural equation model

Model	X^2	X^2/df	TLI	IFI	NFI	CFI	SRMR	RMSEA
Criteria for a good fit	>.05	≤ 3	≥ .95	≥ .95	≥ .95	≥ .95	≤ .08	≤ .06
Initial model	4.75	1.58	.99	.99	.99	.99	.03	.01

As shown in Table 7, the relative chi-square is below the threshold of ≤ 3 as suggested by Kline (2005). All other indices have shown an excellent model fit (TLI = .99; IFI = .99; NFI = .99; CFI = .99; SRMR = .03; RMSEA = .01).

5   Discussion

The purpose of this study was to develop and validate a measure of monitoring and evaluation for occupational learning programs in South Africa. Specifically, the construct of monitoring and evaluation was treated as a latent variable while its attendant items were manifest variables. A single-factor model suggested by Tshilongamulenzhe [35] was tested in the current study using principal components analysis (PCA) and rotated using varimax technique. The findings of this study support a single-factor structure for monitoring and evaluation (ME) scale that was developed by Tshilongamulenzhe [35].

As stated by Tshilongamulenzhe [35] and Tshilongamulenzhe, Coetzee and Masenge [47], the development of the ME scale complied with the established scientific conventions suggested by Benson and Clark [48], DeVellis [49], and Worrington and Whittaker [50]. The psychometric attributes of the scale were examined in accordance with the recommended practices as suggested by Gerbing and Anderson [51], and these included assessment of measures of scale validity and reliability. Scale reliability is defined as the proportion of variance in participants' scores on an instrument due to true differences in their scores [52]. Reliability reflects the consistency of items over time, tests, and groups [41, 53]. The internal consistency method uses vari-ous algorithms to estimate the reliability of a measure from measure administration at one point in time [54]. It assesses the equivalence, homogeneity and inter-correlation of the items used in a measure. This means that the items of a measure should hang together as a set and should be capable of independently measuring the same construct. Nunnally [55] states that newly developed measures can be accepted with ≥ .60; otherwise ≥ .70 should be the threshold [41, 52]. The findings of this study show a good reliability coefficient of .78 for the ME scale and this is above the .70 threshold. Therefore, the items of the ME scale are measuring the same construct, thus making the measure reliable.

The factorial structure of the ME scale was also subjected to structural equation modeling (SEM) as part of confirmatory factor analysis. SEM estimates indirect effects as well as direct effects among latent variables that allow for the estimation of the total effect. The path diagram in the SEM helps to clearly present the direction of each effect and the covariances among all variables in one complete picture [56; 57]. The structural equation model specifies the relationships among the latent variables, and describes the causal effects and amount of unexplained variance [58]. The findings of this study show that the model suggested for ME scale fits the data perfectly (x^2/df = 1.58; TLI = .99; IFI = .99; NFI = .99; CFI = .99; SRMR = .03; RMSEA = .01). A good model fit would provide an insignificant x^2 result at p > .05 threshold [59; 60]. A general rule for acceptable model fit is that the ratio of the x^2 to df (x^2/df) should be ≤ 2 [38] or ≤ 3 [41]. The chi-square test findings (x^2 = 4.75; x^2/df = 1.58) in the current study show that the data provides a good model fit.

A standardized root mean square residual (SRMR) was tested in this study and the findings show a value of .03, thus denoting the best model-fit. An SRMR is the square root of the difference between the residuals of the sample covariance matrix and the hypothesized covariance model [60]. Values of the SRMR range from zero to 1.0 with well-fitting models obtaining values less than .05 [61; 62]. The smaller the SRMR, the better fit of the model. Values as high as .08 are deemed acceptable [42; 43]. An SRMR of 0 indicates perfect fit.

The following relative fit indices were examined to determine model fitness in this study: normed fit index (NFI), incremental fit index (IFI), and Tucker-Lewis index (TLI). The findings show a perfect model fit with a value of .99 for NFI, IFI, and TLI respectively. Bentler and Bonnet [63] recommend that values greater than .90 for NFI indicate a good fit. According to Schumacker and Lomax [64], by convention, NFI values above .95 are good, between .90 and .95 are acceptable, and below .90 indicate a need to respecify the model. A suggestion by Hu and Bentler [42] is that the cut-off criterion should be NFI ≥ .95. Regarding IFI, a suggestion by Schreiber *et al.* [43] is that the cut-off criterion should be IFI ≥ .95. For TLI, Hu and Bentler [42] have suggested ≥ .95 as the cutoff for a good model fit and this is widely accepted. TLI values below .90 indicate a need to respecify the model. All the relative fit indices in the current study are by far in excess of the set thresholds and therefore support a perfect model fit.

Lastly, this study examined the two noncentrality-based indices, that is, the comparative fit index (CFI) and root mean square error of approximation (RMSEA). The CFI assumes that all latent variables are uncorrelated (null/independence model) and compares the sample covariance matrix with this null model [65]. A cut-off criterion of CFI ≥ .90 was initially advanced [60]. However, studies have shown that a value greater than 0.90 is needed in order to ensure that misspecified models are not accepted [42]. Schreiber et al. [43] have suggested CFI ≥ .95 as the cutoff for a good model fit. The CFI is the most popular index in the recent period to be reported and included in structural equation modeling, because it is one of the measures least affected by sample size [66]. The findings of this study show a CFI value of .99 and this is by far in excess of the suggested threshold for a good model-fit.

The RMSEA tells of how well the model, with unknown but optimally chosen parameter estimates would fit the population covariance matrix [61]. In recent times, this statistic has become one of the most informative fit indices [62] due to its sensitivity to the number of estimated parameters in the model. Recommendations for RMSEA cut-off points have been reduced considerably. Up until the early 1990s, an RMSEA in the range of .05 and .10 was considered an indication of fair fit and values above .10 indicated poor fit [67]. It was then thought that a RMSEA value of between .08 and .10 provides a mediocre fit and below .08 shows a good fit [67]. However, in recent times, a cut-off value of .60 or less seems to be the general consensus among authorities in this area [60, 43]. The findings of this study show an RMSEA value of .01, which is extremely lower than the suggested threshold, thus suggesting a perfect model-fit.

Overall, the findings of this study support a valid and reliable ME scale for use in the South African occupational learning context. The scale fits perfectly to the data and its items are working together to support one construct (i.e., monitoring and evaluation), thus supporting unidimensionality and construct validity.

A conclusion that can be drawn from the findings is that the ME scale is unidimensional, valid and reliable and could be used with confidence to examine the effectiveness of monitoring and evaluation activities by stakeholders in the South African occupational learning context.

A limitation of this study is that the sample was drawn from two types of learning programs, that is, learner ship and apprenticeship. Therefore, the findings must be interpreted within the context of these two learning programs and nothing else.

It is recommended that a validation study be conducted on a different population and/or sample in South Africa and a careful consideration be made to test the structural invariance of the ME scale across population subgroups.

6   References

[1]   Rasool, F., Botha, C.J. - *The nature, extent and effect of skills shortages on skills migration in South Africa*. [in] South African Journal of Human Resource Management, Vol. 9, No. 1, 2011, pp. 1-12.

[2]   Powell L., Lolwana P. - *From reconstruction to deconstruction – The role of research in South African skills development,* NORRAG, 2012, http://www.norrag.org (accessed on 2014.05.20).

[3]   RSA. - *Skills Development Act, Act No. 97 of 1998.* Government Printers, Pretoria 1998.

[4]   Carnevale A. - *Workplace basics: The skills employers want* [in] Training and Development Journal, Vol. 42, No. 10, 1988, pp. 22-30.

[5]   Schultz K. - *Training for basic skills or educating workers? Changing conceptions of workplace education programmes*. National Centre for Research in Vocational Education, Berkeley, CA 1992.

[6]   Cloete F. - *Evidence-based policy analysis in South Africa: Critical assessment of the emerging government-wide monitoring and evaluation*. [in] Journal of Public Administration, Vol. 44, No. 2, 2009, pp. 293 – 311.

[7]   Wolman H. - *Policy transfer among local governments: An information-theory approach*. [in] *Governance: An International Journal of Policy and Administration*, Vol. 15, No. 4, 1992, pp. 577-601.

[8]   Visser M., Kruss G. - *Learnerships and skills development in South Africa: A shift to prioritize the young unemployed*. [in] Journal of Vocational Education and Training, Vol. 61, No. 3, 2009, pp. 357-374.

[9]   Kraak A. - *A critical review of the national skills development strategy in South Africa*. [in] Journal of Vocational Education and Training, Vol. 60, No. 1, 2008, pp. 1-18.

[10]   Fester R. - *A study on academic staff's perceptions of learnership programme delivery at a Further Education and Training Institution*. Magister Technologiae mini-dissertation. University of Johannesburg, Johannesburg 2006.

[11]   Powell L. - *Learnerships in South Africa. Unpublished report submitted to the Human Sciences Research Council.* HSRC, Pretoria 2007.

[12]   Kraak A. - *Human resources development and the skills crisis in South Africa: the need for a multi-prolonged strategy*. [in] Journal of Education and Work, Vol. 18, No. 1, 2005, pp. 57-83.

[13]   Karlsson J., Berger M. - *Well-grounded and prepared new teachers: Reflection on the promise of learnerships*. [in] Perspectives in Education, Vol. 24, No. 1, 2006, pp. 53-64.

[14]   Smith M.J., Jennings R., Solanki G. - *Perspectives on learnerships: A critique of South Africa's transformation of apprenticeships*. [in] Journal of Vocational Education and Training, Vol. 57, No. 4, 2005, pp. 537–561.

[15]   Cracknel, B. - *Evaluating development aid: Issues, problems, and solutions*. Sage, London 2000.

[16]   World Bank. - *Designing project monitoring and evaluation: Lessons and practices,* Vol. 8. Operations Evaluation Department, World Bank 1996.

[17]   European Union - *Evaluating and monitoring apprenticeship and traineeship schemes,* Seminar report. European Commission, Brussels 2014.

[18]   Lange L., Luescher T.M. - *A monitoring and evaluation system for South African higher education: Conceptual, methodological and practical concerns*. [in] South African Journal for Higher Education, Vol. 17, No. 3, 2003, pp. 82 – 89.

[19]   OECD/DAC. - *Key terms in evaluation and results based management*. OECD/DAC, Paris 2002.

[20]   United Nations Children's Fund. - *A UNICEF guide for monitoring and evaluation: Making a difference?* 2005, (accessed 2015.05.21) http://www.unicef.org/reseval/mander.html,

[21]   Naidoo P. - *Measuring indicators of quality in physical science education: Case studies of high performing schools*. PhD Thesis, University of Durban Westville, Durban 2001.

[22] Annecke W. - *Monitoring and evaluation of energy for development: The good, the bad and the questionable in M&E practice* [in] Energy Policy, Vol. 36, 2008, pp. 2839-2845.

[23] De Jong B.H.J., Tipper R., Taylor J. - *A framework for monitoring and evaluating carbon mitigation by farm forestry projects: Example of a demonstration project in Chiapas, Mexico* [in] Mitigation and Adaption Strategies for Global *Change*, Vol. 2, 1997, pp. 231-246.

[24] Hockings M., Stolton S., Dudley N., Parrish J. - *The enhancing our heritage toolkit: Evaluating effectiveness training manual.* University of Queensland, Gatton, Queensland 2001.

[25] Sawhill J., Williamson D. - *Measuring what matters in non-profits* [in] The McKinsey Quarterly, Vol. 2, 2001, pp. 98 – 107.

[26] Hatry, H.P. - *Performance measurement: Getting results.* Urban Institute Press, Washington D.C 1999.

[27] Blann, K., Light, S.S. - *The path of last resort: Adaptive environmental assessment and management.* Adaptive Management Practitioners' Network, Minneapolis, Minnesota 2000.

[28] Rigby D., Howlett D., Woodhouse P. - *Sustainability indicators for natural resource management and policy: A review of indicators of agricultural and rural livelihood sustainability.* University of Manchester, Manchester, United Kingdom 2000.

[29] Mark M.M., Henry G.T., Julnes G. - *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs.* Jossey-Bass, San Francisco 2000.

[30] Stem C., Margolius R., Slafsky N., Brown M. - *Monitoring and evaluation in conservation: A review of trends and approaches* [in] Conservation Biology, Vol. 19, No. 2, 2005, pp. 295-309.

[31] NQI. - *Canadian framework for business excellence: Overview document.* National Quality Institute, Canada 2007

[32] Chee S. S. - *Training through apprenticeship – The Singapore system* [at] The Network Conference of the Asia Pacific Economic Cooperation – Human Resource Development for Industrial Technology on "Quality Workforce through On-the-Job Training", Singapore 1992, pp. 24-36.

[33] DHET. - *Framework for the national skills development strategy 2011/12 – 2015/16: First Draft for Consultation, 29th April.* Department of Higher Education and Training, Pretoria 2010.

[34] DHET. - *Learning Programme Regulations.* Department of Higher Education and Training, Pretoria 2012.

[35] Tshilongamulenzhe, M.C. - *An Integrated Learning Programme Management and Evaluation Model for the South African Skills Development Context.* Unpublished Doctorate Thesis, University of South Africa, Pretoria 2012.

[36] IBM. - *IBM SPSS Statistics 23 Brief Guide.* IBM Corporation, New York 2015.

[37] Landau S., Everitt B.S. - *A handbook of statistical analyses using SPSS.* Chapman & Hall 2004.

[38] Tabachnick B.G., Fidell L.S. - *Using multivariate statistics.* Allyn & Bacon, New York 2007.

[39] Bruce P. - *Structural equation modeling: Applications in ecological and evolutionary biology.* Cambridge University Press, West Nyack, NY 2003.

[40] Kline R.B. - *An easy guide to factor analysis.* Routledge, London 1994.

[41] Kline R.B. - *Principles and practice of structural equation modeling.* The Guilford Press, New York, NY 2005.

[42] Hu L.T., Bentler P.M. - *Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.* [in] Structural Equation Modeling, Vol. 6, No. 1, 1999, pp. 1-55.

[43] Schreiber J. B., Stage F.K., King J., Nora A., Barlow E.A. - *Reporting structural equation modeling and confirmatory factor analysis results: a review.* [in] The Journal of Educational Research, Vol. 99, No. 6, 2006, pp. 323-337.

[44] Brown T.A. - *Confirmatory factor analysis for applied research.* The Guilford Press, New York, NY 2006.

[45] Liu X. - *Assessing measurement invariance of the Teachers' Perception of Grading Practices Scale across Cultures*. [in] International Journal of Humanities and Social Science, Vol. 1, No. 20, 2011, pp. 70-80.

[46] Fletcher A.C., Walls J.K., Eanes A.Y., Troutman D.R. - *Maternal Management of Social Relationships as a Correlate of Children's School-Based Experiences*. [in] The School Community Journal, Vol. 20, No. 2, 2010, pp. 159 – 181.'

[47] Tshilongamulenzhe M.C., Coetzee M., Masenge, A. - *Development of the Learning Programme Management and Evaluation Scale for the South African Skills Development Context* [in] South African Journal of Industrial Psychology, Vol. 39, No. 1, 2013, pp. 1-14.

[48] Benson J., Clark F. - *A Guide for Instrument Development and Validation* [in] The American Journal of Occupational Therapy, Vol. 36, 1982, pp. 789-800.

[49] DeVellis R.F. - *Scale Development*. Sage Publications, Thousand Oaks, CA 2012.

[50] Worthington R.L., Whittaker T.A. - *Scale Development Research: A Content Analysis and Recommendations for Best Practices* [in] The Counselling Psychologist, Vol. 34, No. 6, 2006, pp. 806-838.

[51] Gerbing D.W., Anderson J.C. - *An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment* [in] Journal of Marketing Research, Vol. 25, 1988, pp. 186-192.

[52] Polit D.F., Beck C.T. - *Nursing research: Principles and methods*. Williams & Wilkins, Philadelphia, Lippincott 2004.

[53] Nunnally J.C., Bernstein I. - *Psychometric theory*. McGraw Hill, New York 1994.

[54] Forza C. - *Survey research in operations management: A process-based perspective* [in] International Journal of Operations & Production Management, Vol. 22, No. 2, 2002, pp. 152-194.

[55] Nunnally J. C. - *Psychometric theory*. McGraw-Hill, New York, NY 1978.

[56] Hair J.F.Jr., Anderson R.E., Tatham R.L., Black W.C. - *Multivariate data analysis*. Prentice-Hall, New Jersey 1998.

[57] Kline R.B. - *Principles and practice of structural equation modeling*. Guilford Press, New York 1998.

[58] Chavance M., Escolano S., Romon M., Basdevant A., de Lauzon-Guillain B., Charles M.A. - *Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology*. [in] BMC Medical Research Methodology, Vol. 10, 2010, p. 37.

[59] Barrett P. - *Structural equation modeling: Adjudging model fit*. [in] Personality and Individual Differences, Vol. 42, No. 5, 2007, pp. 815-24.

[60] Hooper D., Coughlan J., Mullen M. - *Structural Equation Modeling: guidelines for Determining Model Fit* [in] Electronic Journal of Business Research Methods, Vol. 6, No. 1, 2008, pp. 53-60.

[61] Byrne B.M. - *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications and programming*. Lawrence Erlbaum Associates, Mahwah, New Jersey 1998.

[62] Diamantopoulos A., Siguaw J.A. - *Introducing LISREL*. Sage Publications, London 2000.

[63] Bentler P.M., Bonnet D.C. - *Significance tests and goodness of fit in the analysis of covariance structures* [in] Psychological Bulletin, Vol. 88, No. 3, 1980, pp. 588-606.

[64] Schumacker R.E., Lomax R.G. - *A beginner's guide to structural equation modeling*. Lawrence Erlbaum Associates, Inc, Mahwah, NJ 2004

[65] Bentler P.M. - *Comparative fit indexes in structural models* [in] Psychological Bulletin, Vol. 107, No. 2, 1990, pp. 238-46.

[66] Fan X., Thompson B., Wang L. - *Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes*. [in] Structural Equation Modeling, Vol. 6, No. 1, 1999, pp. 56-83.

[67] MacCallum R.C., Browne M.W., Sugawara H.M. - *Power analysis and determination of sample size for covariance structure modeling* [in] Psychological Methods, Vol. 1, No. 2, 1996, pp. 130-49.