Performance of exploratory structural equation model (ESEM) in detecting differential item functioning

Ayanwale, Musa Adekunle

PERFORMANCE OF EXPLORATORY STRUCTURAL EQUATION MODEL (ESEM) IN DETECTING DIFFERENTIAL ITEM FUNCTIONING

Musa Adekunle,Ayanwale
Department of Science and Technology Education
University of Johannesburg
Auckland Park, South Africa, PO Box 524, 2006
kunleayanwale@gmail.com ayanwalea@uj.ac.za

Abstract

The validity of a standardised test is questioned if an irrelevant construct is accounted for the performance of examinees, which is wrongly modeled as the ability in the construct (test items). A test must ensure precision in the examinee's ability irrespective of their sub-population in any demographic variables. This paper explored the potentials of gender and school location as major covariates on the West African Examinations Council (WAEC) mathematics items among examinees (N=2,866) using Exploratory Structural Equation Modeling (ESEM). The results remarked that the test is multidimensional (six-factors) with compliance fix indices of (χ^2 (940)=4882.024, p < 0.05, CFI=0.962, TLI=0.930, RMSEA=0.038, SRMR=0.030, 90 % CI=0.037-0.039, Akaike information criterion (AIC)=147290.577, Bayesian information criterion (BIC)=149585.436 and Sample-size adjusted BIC=148362.154) respectively. Also, there were 10 (20 %) significant DIF items in the WAEC to gender, while 3 (6 %) of the items indicated significant DIF to school location. Observed DIF items acquaint test developers; the existence of DIF may differentially affect the performance of examinees with the same ability level. The implications of the test are severe for the examinees. Hence, accurate and unbiased assessment should be the basic principles for any test item measurement, and test developers need to test the items to be free from biases psychometrically.

Keywords: Differential Item Functioning (DIF), Exploratory Structural Equation Modeling (ESEM), Exploratory Factor Analysis (EFA), Multidimensionality, West African Examinations Council (WAEC), Confirmatory Factor Analysis.

1. Introduction

Ensuring test fairness and equity among examinees is very important. Examinees should be given equal opportunity to display what they know and perform well in the tested area according to their demographic profile, such as race, gender, location, ethnicity, religion, color, linguistic background, etc. The measurement community has increasingly employed standardised testing as one of the significant tools, deployed in assessing examinees' outcomes and precise ability and examinee potential for future academic success. However, measurement experts over the years were apprehensive of the likelihood that test items, either cognitive or non-cognitive, might function differently (that is, favouring sub-group) across a group of examinees [1, 2]. Thus, the validity of a standardised test is questioned if an irrelevant construct is accounted for the performance of examinees, which is wrongly modeled as the ability in the construct (test items). A test must ensure precision in the examinee's ability irrespective of their sub-population in any demographic variables. The examinee's test score interpretation concerns the extent of its statistical independence across and among different groups of examinees in educational testing.

The West African Examinations Council (WAEC) is an examination board, established by law to determine the examinations, required in the public interest in the English-speaking West African countries, to conduct terminal examinations for grade 12 students, and to award certificates comparable to those of equivalent examining authorities internationally [3]. Since its establishment in 1952, the council has contributed to education in Anglophonic countries of West Africa (Ghana, Nigeria, Sierra Leone, Liberia, andtGambia), with the number of examinations they have coordinated and certificates they have issued. It is noteworthy, that this public examining body must avoid integrating items that do not establish a comparative opportunity to examinees' in their test. They need to assess all items in a given test to ensure that they provide all the examinees equal opportunity to demonstrate their innate ability (traits) regardless of their demographic variables.

Dimensionality is the number of traits, underlying a test that accounts for student test performance variation. The student's performance, in a higher set of n-dimensional latent space, is often described by a vector of ability scores as $(\theta_1, \theta_2, \theta3 \ldots \theta_n)$. Item response models that presume many latent traits, which account for students' test performance, are known as multidimensional item response models, whereas item response models that assume a single latent ability are known as unidimensional models [4–6]. Also, [7, 8] suggested the exploratory approach to bundle items. This approach involves the use of statistical procedures to identify distinct dimensions, and numerical test items, such as quantitative reasoning, mental ability achievement test, etc., were generally believed to measure more than one trait (multidimensional) [9–12].

A method of assessing the extent, to which each item in a given test gives an equal opportunity to all the examinees, is known as differential Item Function (DIF). Also, differential item function detection is the most potent tool to evaluate the statistical independence of cognitive and non-cognitive instruments across a sub-group of examinees. DIF is a phenomenon that arises when an examinee's probability of responding to an item correctly is independent of their actual ability but somewhat dependent on group membership. Exclusively, it occurs when examinees from various groups display the differing likelihood of success on the item after matching on their underlying ability that the item is purported to measure. This happens between two groups, a reference group, otherwise known as the majority, while a focal group is a minority (subject of investigation). The essence of DIF is to detect and estimate the interaction between item difficulties and various subgroups within the population of examinees [13–16].

There are two types of DIF that items can flag. These include uniform and non-uniform DIF. An item exhibits uniform DIF if there is no interaction between ability level and group membership; that is, the likelihood of endorsing an item correctly is more significant for one group, uniformly overall matched ability levels. Conversely, for an item with non-uniform DIF, the probability of answering an item correctly is not the same over all the matched ability. There are several methods of describing and detecting DIF. These include; Logistic regression, Item Response Theory Likelihood Ratio Test (IRT-LRT), Simultaneous Item Bias Test (SIBTEST), Mantel-Haenszel [17–22]. However, in this paper Exploratory Structural Equation Model (ESEM) of covariates was advanced. This, to the best of the researcher, had not been explored in the literature to investigate how mathematics items of WAEC behave across different sub-population of examinees. Though researchers have used variants of ESEM like Multiple Indicator Multiple Cause (MIMIC), Measurement Invariant (MI), and Multiple Group Analysis (MGA) to examine how a statistical property of a measurement that indicates the same underlying construct is being measured across groups or time in numerous studies [23–26].

Nevertheless, the question now is how tenable the performance of the exploratory structural equation model in detecting differential item function of the cognitive instrument is? That is the relationship between observed covariates and a latent variable. Although cross-loading of items is more visible and practicable with exploratory factor analysis (EFA), it is crystal clear that better techniques and approaches are more evident with confirmatory factor analysis (CFA) than EFA. Consequently, to combine the two methods, an alternative method was advanced [27, 28] called Exploratory Structural Equation Modeling (ESEM), which integrates advantages of EFA and CFA into its technique.

Theoretical Framework
Exploratory and Confirmatory Factor Analysis
Confirmatory factor analysis (CFA), according to [29], has been the essential ingredient of psychometric research since its inception and quickly became a default, "go-to" method in psychometrics due to the methodological advances, associated with it (that is, goodness-of-fit, estimation of different models, and the inclusion of method factors) relative to exploratory factor analysis (EFA). Also, its performance compares to EFA in terms of allowing items to load on their respective factor, while constraining cross-loading on other factors to zero is one of the significant underlying characteristics of CFA. Conversely [30–32] submitted that EFA freely estimates all cross-loadings. The researchers do not impose a specific latent structure on the observed indicators

but rather allow the optimal number of factors to be determined based on several statistical and interpretability criteria. All these might indicate that EFA is less essential and inferior to CFA. Nevertheless, [33] argued that CFA's acceptance, acknowledgment, and usefulness could be seen as a motivation to create more parsimonious measurement models. More often than not, these models and items include a certain level of systematic measurement error in the form of cross-loadings.

[34] asserted that items are rarely pure indicators of their corresponding constructs; they are fallible; thus, at least some degree of construct-relevant association can be expected between items and the non-target, yet conceptually related constructs. When non-zero cross-loadings are present and unexpressed simultaneously, such restrictive constraints (that is, items can only load on one factor) could inflate the associations between the factors as the mis-specified cross-loadings could only be expressed through these factorial associations. Moreover, the goodness-of-fit of the models and the discriminant validity of the factors could also be undermined by these overly restrictive specifications [27]. To proffer solutions to these limitations, the Exploratory Structural Equation Modeling (ESEM) framework [23, 28] has been developed, which integrates the advantages of the less restrictive EFA (that is allowing cross-loadings) and the more advanced CFA (that is goodness-of-fit) at the same time, providing a collaboration that is the best of both features and can sufficiently account for complex measurement models.

Exploratory Structural Equation Model (ESEM)
For the exploratory structural equation model [23, 24, 26], the response variables $Y=(Y_1, ..., Y_n)$, the predictor variables $X=(X_1, ..., X_m)$ and k latent variables $\eta=(\eta_1, ..., \eta_k)$ are under the standard assumptions that the ε and ζ are normally distributed residuals with mean of zero and variance-covariance matrix θ and ψ respectively. Λ is a factor loading matrix, while C and Γ are matrices of regression coefficients that relate latent variables.

$$Y = v + \Lambda\eta + KX + \varepsilon,$$

$$\eta = \alpha + C\eta + \Gamma X + \xi.$$

However, all parameters can be identified with the maximum likelihood estimation method (MLE); the model is generally not identified unless additional constraints are imposed. In CFA analyses, the two typical approaches are to identify the metric of the latent variable by either fixing the variance of the latent variable to be 1.0 or by fixing one of the factor loadings for each factor typically to be 1.0. The ESEM approach differs from the typical CFA approach in that all factor loadings are estimated, subject to constraints, so that the model can be identified. More importantly, [30] suggested that the ESEM model's estimation comprises many steps. First, an SEM model is estimated using the ML estimator. The factor variance-covariance matrix is indicated as an identity matrix ($\psi=I$), giving k (k+1)/2 restrictions. The EFA loading matrix (Λ) has all entries above the main diagonal (that is for the first k rows and column in the upper right-hand corner of the factor loading matrix, Λ), fixed to 0, providing remaining k (k−1)/2 identifying restrictions. This initial, unrotated model provides starting values that can be rotated into an EFA model with k factors. The asymptotic distribution of all parameter estimates in this starting value model is also obtained. Consequently, the ESEM variance-covariance matrix is computed.

Researchers, such as [32, 35, 36], argued that ESEM was a better and efficient method to adjust for the cross-factor loading instead of latent variables analysis, which assesses a measurement model of constructs through exploratory factor analysis (EFA) in place of CFA. Generally, ESEM showed improved model fit results as well as deflated inter-factor correlations that, in turn, improve the discriminant validity of the factors as well as produce a more realistic representation of the data [31, 33, 34, 37]. Hence, the relative novelty of the ESEM method in literature is now adequately established from various studies within the context of behavioural socio-science research [38–42]. Numerous studies have remarked the impressive performance of ESEM compared to CFA in investigating the measurement structure of latent variables [27, 37, 43].

Based on the review above, it's expected that ESEM would perform better in detecting differential item function for the demographic profile of the examinees. Consequently, there is a likelihood that the predictors (gender and school location) have unique and distinct effects on the test items that their impact on the latent variables cannot fully explain. Succinctly, bias is an issue to be addressed since tests are used as gatekeepers for educational opportunities, and test items should be fair for every examinee. Their importance is justified only if the measures, used to produce valid outcome data for different sub-populations, are presented with the same test. Personal experience has shown that examinees, responding to WAEC mathematics items, were from various regions in Nigeria and taught by different mathematics teachers. The teachers from these regions employed different constructs and instructional pedagogies to teach the examinees. The areas of specialty, years of experience, and qualifications of these teachers differed; therefore, content exposure and constructs, exposed to students, differed across regions. It is expected, that these students should write the same examination. The question is that do the public examining bodies, such as WAEC, take cognisance of this situation when developing their test items? The answer is an absolute NO.

Consequently, this may have significant interference on the students' performance in the test due to observed differing content exposure and location. Also, this may pose a threat to equity and fairness and make decisions taken not reflect the students' actual ability. This could affect the predictive and construct validity of the test since the item biasness was not adequately handled at the level of test development (i. e., assessing the psychometric properties of the items).

Therefore, this study aimed to establish the performance of exploratory structural equation modeling to detect differential item functioning and provides answers to the following research questions. They include (i) what is the number of underlying factors in the WAEC mathematics items, (ii) is there significant effect of gender covariate on the mathematics items, and (iii) is there significant effect of school location covariate on the mathematics items, respectively.

2. Materials and methods

Design, Population, and Sample

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the University of Johannesburg (protocol code Sem 2-2021-164). Informed consent was obtained from all participants, involved in the study.

This paper employed a non-experimental design of instrumentation research type. The study population was prospective K-12 students, selected randomly from Education District 1 (Agege, Alimosho, and Ifako/Ijaye) of Lagos State, Nigeria. A sample of 2,866 participants, 1,233 (43 %) male, 1,633 (57 %) females, aged between 14 and 20 years, were obtained from 28 schools in the State.

Summary descriptive statistics were calculated (**Table 1**), followed by exploratory factor analysis using WLSMV estimator and Goemin rotation, and subsequently confirmatory factor analysis (CFA). After finding the best model, differential item functioning (DIF) was conducted using the exploratory structural equation model (ESEM) to determine if the examinee's responses to item differed by gender and school location after controlling for theta (θ).

Table 1

Summary Descriptive Statistics of WAEC Mathematics Test items

Variable	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
1	2	3	4	5	6	7
Gender	1.52	0.500	−0.077	0.046	−1.995	0.091
Location	1.59	0.492	−0.369	0.046	−1.865	0.091
V1	0.72	0.449	−0.983	0.046	−1.034	0.091
V2	0.71	0.456	−0.906	0.046	−1.180	0.091
V3	0.78	0.412	−1.381	0.046	−0.093	0.091

Continuation of Table 1

1	2	3	4	5	6	7
V4	0.93	0.256	–3.358	0.046	9.282	0.091
V5	0.63	0.484	–0.524	0.046	–1.727	0.091
V6	0.64	0.481	–0.574	0.046	–1.672	0.091
V7	0.73	0.442	–1.064	0.046	–0.867	0.091
V8	0.65	0.478	–0.613	0.046	–1.625	0.091
V9	0.76	0.426	–1.229	0.046	–0.490	0.091
V10	0.68	0.467	–0.769	0.046	–1.410	0.091
V11	0.73	0.446	–1.015	0.046	–0.971	0.091
V12	0.51	0.500	–0.038	0.046	–2.000	0.091
V13	0.69	0.462	–0.830	0.046	–1.312	0.091
V14	0.79	0.407	–1.432	0.046	0.051	0.091
V15	0.70	0.457	–0.895	0.046	–1.200	0.091
V16	0.90	0.306	–2.590	0.046	4.712	0.091
V17	0.74	0.438	–1.099	0.046	–0.792	0.091
V18	0.91	0.283	–2.912	0.046	6.483	0.091
V19	0.70	0.460	–0.857	0.046	–1.267	0.091
V20	0.63	0.482	–0.552	0.046	–1.697	0.091
V21	0.69	0.461	–0.844	0.046	–1.288	0.091
V22	0.68	0.465	–0.795	0.046	–1.369	0.091
V23	0.70	0.457	–0.895	0.046	–1.200	0.091
V24	0.66	0.475	–0.655	0.046	–1.572	0.091
V25	0.61	0.487	–0.458	0.046	–1.791	0.091
V26	0.65	0.477	–0.623	0.046	–1.613	0.091
V27	0.74	0.437	–1.112	0.046	–0.764	0.091
V28	0.52	0.500	–0.087	0.046	–1.994	0.091
V29	0.69	0.462	–0.832	0.046	–1.309	0.091
V30	0.71	0.452	–0.949	0.046	–1.100	0.091
V31	0.73	0.442	–1.054	0.046	–0.889	0.091
V32	0.57	0.495	–0.296	0.046	–1.913	0.091
V33	0.57	0.495	–0.296	0.046	–1.913	0.091
V34	0.75	0.434	–1.146	0.046	–0.688	0.091
V35	0.87	0.339	–2.168	0.046	2.702	0.091
V36	0.72	0.450	–0.966	0.046	–1.068	0.091
V37	0.65	0.478	–0.610	0.046	–1.629	0.091
V38	0.73	0.444	–1.032	0.046	–0.935	0.091
V39	0.69	0.462	–0.830	0.046	–1.312	0.091
V40	0.69	0.463	–0.812	0.046	–1.341	0.091
V41	0.71	0.455	–0.908	0.046	–1.177	0.091
V42	0.69	0.461	–0.844	0.046	–1.288	0.091
V43	0.69	0.463	–0.816	0.046	–1.335	0.091
V44	0.74	0.438	–1.097	0.046	–0.797	0.091
V45	0.79	0.405	–1.445	0.046	0.089	0.091
V46	0.73	0.445	–1.028	0.046	–0.943	0.091
V47	0.72	0.448	–0.993	0.046	–1.015	0.091
V48	0.69	0.462	–0.830	0.046	–1.312	0.091
V49	0.75	0.431	–1.180	0.046	–0.607	0.091
V50	0.63	0.483	–0.532	0.046	–1.718	0.091

Instrumentation

In Nigeria, WAEC is a high-stake examination. It is a standardised examination, prepared by experienced and seasoned item developers from various higher institutions with test and measurement experts. The 2020 mathematics instrument contained 50 items with four options (A-D). Students took a survey with 50 multiple-choice items, drawn across common areas of the math-

ematics syllabus by the examining body. They rest assured that their responses would be treated with utmost secrecy. The survey took a maximum of 1 hour 30 minutes to complete. The empirical reliability was calculated to be 0.85. It was administered to 3,000 participants, and the optical mark's reader (OMR) sheet was used to shade the correct answers, but 2,866 OMR sheets were returned and used in the analysis. Data obtained was analysed using Mplus 7.4 [44] and estimated with the robust maximum likelihood estimator (MLR), which provides standard errors and tests of model fit that are robust to the non-normality of the data.

3. Results

Participants' responses to dichotomous test items (mathematics achievement test) were subjected to EFA, implemented in Mplus to determine the test data's optimal factor. To achieve this, 1-factor was hypothesised that fit the test data. Also, 2-factors were hypothesised that underlie the test data, then the fit indices for 1-factor and 2-factors were compared. Hence, if 2-factors fit the data better than 1-factor, the data is further calibrated under the hypothesis that 3-factors fit the test data, and the fitness of 2-factors and 3-factors were compared in the same trend. This process persists until the optimal factor, underlying the test data, is achieved. **Tables 2, 3** presented the model fit summary information for the test data.

Table 2

Model fit summary information for the test data

Model	Number of Parameters	Chi-Square	df	p-value	CFI	TLI	SRMR	RMSEA	90 % C.I
1–factor	50	9797.73	1175	0.00	0.81	0.79	0.09	0.051	0.050–0.052
2–factor	99	7504.18	1126	0.00	0.86	0.84	0.08	0.044	0.044–0.045
3–factor	147	6579.56	1078	0.00	0.88	0.86	0.07	0.042	0.041–0.043
4–factor	194	5885.70	1031	0.00	0.89	0.87	0.06	0.041	0.040–0.042
5–factor	240	4928.98	985	0.00	0.91	0.89	0.05	0.037	0.036–0.038
6–factor	285	3905.60	940	0.00	0.96	0.94	0.04	0.033	0.032–0.034

Table 3

Comparison of Model Fit Information

Models Compared	Chi-Square	df	p-value	CFI	TLI	SRMR	RMSEA
1-factor against 2-factor	1365.6	49	0.00	0.86	0.84	0.08	0.044
2-factor against 3-factor	920.43	48	0.00	0.88	0.86	0.07	0.042
3-factor against 4-factor	749.9	47	0.00	0.89	0.87	0.06	0.041
4-factor against 5-factor	901.49	46	0.00	0.91	0.89	0.05	0.037
5-factor against 6-factor	873.5	45	0.00	0.96	0.94	0.04	0.033

Table 2 presents the number of hypothesized factors, underlying the 2020 WAEC mathematics test, while **Table 3** depicted the nested model fit when the hypothesised factors were compared. According to [45] as cited in [46], it is suggested, that a Comparative Fit index (CFI) of 0.90 indicates an acceptable level of fit and a value of 0.95 indicates good fit; CFI=1 indicates perfect fit, Tucker-Lewis index (TLI) of 0.90 as an acceptable threshold and 0.95 indicates a good fit; Standardised Root Mean Square Residual (SRMR) of 0.00 indicates perfect fit and value of 0.08 or less is considered to be an indicator of good model fit. The Root Mean Square Error of Approximation (RMSEA) is the benchmark for judging the overall fitness of a model with the threshold value of 0.05 or less as a sign of good model fit.

The nested model in **Table 3** indicated that when the fitness of 1-factor model was compared to 2-factor model, 2 factor model fitted the data better with $\chi^2_{(49)}$=1365.6, $p<0.05$, CFI=0.86, TLI=0.84, RMSEA=0.04 and SRMR=0.08. When the fitness of 2-factor and 3-factor models was compared, the analysis showed that 3-factor fitted the data better than the 2-factor model with $\chi^2_{(48)}$=920.43, $p<0.05$, CFI=0.88, TLI=0.86, RMSEA=0.04 and SRMR=0.07. Also, when the fitness of 3-factor was compared with 4-factor, 4-factor fitted the data better than the 3-factor with $\chi^2_{(47)}$=749.9,

$p<0.05$, CFI=0.89, TLI=0.87, RMSEA=0.04 and SRMR=0.06. Moreover, the procedure continues till when the fitness of 5-factor and 6-factor models was compared, it was remarked, that 6-factor fitted the data better than the 5-factor model with $\chi^2_{(45)}$=873.5, $p<0.05$, CFI=0.96, TLI=0.94, RMSEA=0.03 and SRMR=0.04 respectively. Consequently, the result showed that 6-factors fitted the data better. It is evident, that there are 6 possible sub-factors that reflect the mathematics test item of WAEC.

More importantly, to establish the number and characteristics of the factors, this is interpretable, non-linear exploratory structural equation modeling was employed. **Fig. 1** presents the ESEM of the six factors, explaining the variance, observed in the performance of examinees' in the test.
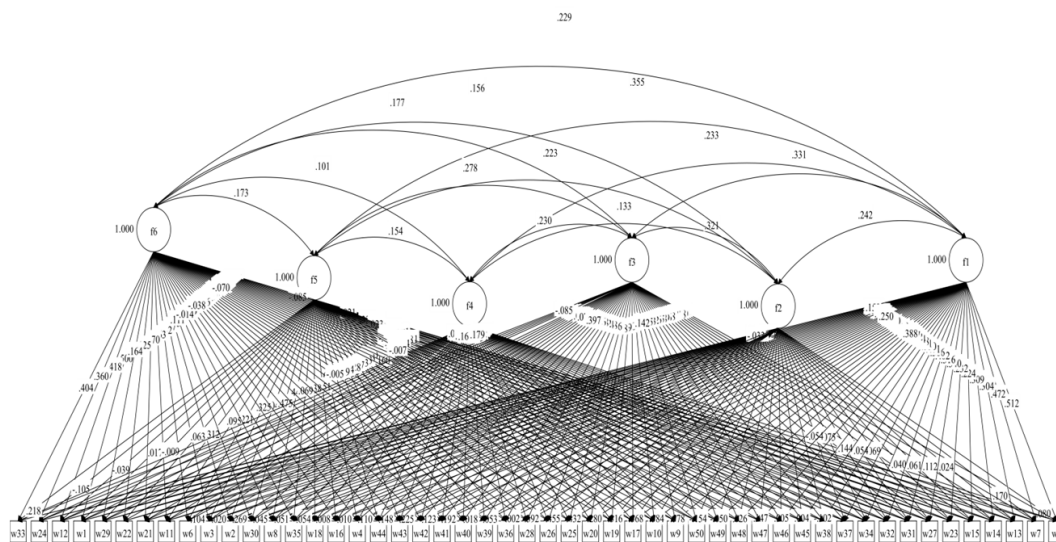

Fig. 1. Path Diagram of Exploratory Structural Equation Model (ESEM) for the 6-factor

Fig. 1 depicts the ESEM for the 6-factor model for the test. The model remarked that six salient factors, underlying the test, were interpretable with $\chi^2_{(940)}$=4882.024, $p<0.05$, CFI=0.962, TLI=0.930, RMSEA=0.038, SRMR=0.030, 90 % CI=0.037-0.039, Akaike information criterion (AIC)=147290.577, Bayesian information criterion (BIC)=149585.436 and Sample-size adjusted BIC=148362.154 respectively. Based on these fit indices, six factors are embedded in the test data, with each factor having more than three substantial item loadings of 0.32 and above. **Fig. 2** present the performance of ESEM in detecting gender biases in the mathematics test data.
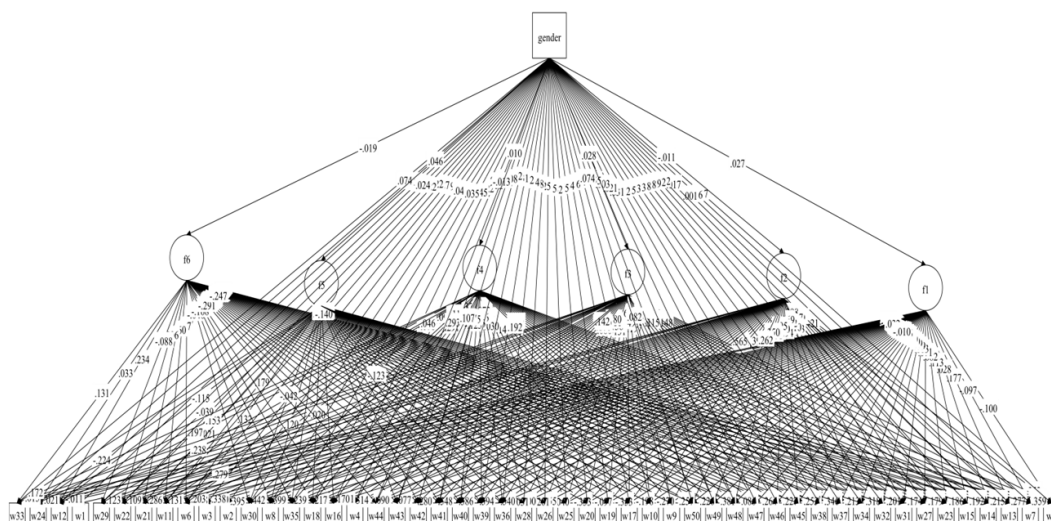

Fig. 2. ESEM with respect to Gender as Covariate

Fig. 2 depicts ESEM with gender as a covariate of the six factors that accounted for the student's performance in the test. This shows that the 6-factor model, describing the performance of students in mathematics test item, was viable and substantial with, $\chi^2_{(940)}=4938.197$, $p<0.05$, CFI=0.954, TLI=0.960, RMSEA=0.038, SRMR=0.029, 90 % CI=0.036–0.039. Consequently, the extent, to which gender affects the traits that underlie examinees' performance in mathematics tests, was evaluated. The results are presented in **Table 4** as follows:

Table 4
Model of Six Factors ESEM with Gender as Covariate

Item	Covariate	Estimate	Std. Error (S.E.)	Est./S.E.	p-value	Remarks
1	2	3	4	5	6	7
W1	ON					
	GENDER	0.019	0.022	0.842	0.400	No DIF
W2	ON					
	GENDER	0.050	0.064	0.791	0.429	No DIF
W3	ON					
	GENDER	0.145	0.050	2.923	0.003	DIF
W4	ON					
	GENDER	−0.033	0.012	−2.749	0.006	DIF
W5	ON					
	GENDER	0.012	0.020	0.613	0.540	No DIF
W6	ON					
	GENDER	0.138	0.047	2.963	0.003	DIF
W7	ON					
	GENDER	0.026	0.019	1.409	0.159	No DIF
W8	ON					
	GENDER	−0.025	0.020	−1.251	0.211	No DIF
W9	ON					
	GENDER	−0.002	0.019	−0.126	0.900	No DIF
W10	ON					
	GENDER	0.029	0.021	1.385	0.166	No DIF
W11	ON					
	GENDER	−0.362	0.060	−6.029	0.000	DIF
W12	ON					
	GENDER	−0.017	0.022	−0.755	0.450	No DIF
W13	ON					
	GENDER	−0.032	0.019	−1.743	0.081	No DIF
W14	ON					
	GENDER	−0.034	0.017	−2.063	0.039	DIF
W15	ON					
	GENDER	−0.013	0.018	−0.683	0.495	No DIF
W16	ON					
	GENDER	−0.015	0.014	−1.067	0.286	No DIF
W17	ON					
	GENDER	0.015	0.019	0.803	0.422	No DIF
W18	ON					
	GENDER	−0.044	0.012	−3.524	0.000	DIF
W19	ON					
	GENDER	−0.002	0.019	−0.093	0.926	No DIF
W20	ON					
	GENDER	−0.031	0.020	−1.572	0.116	No DIF
W21	ON					
	GENDER	0.188	0.071	2.647	0.008	DIF
W22	ON					
	GENDER	0.282	0.062	4.548	0.000	DIF

Continuation of Table 4

1	2	3	4	5	6	7
W23	ON					
	GENDER	0.031	0.017	1.835	0.066	No DIF
W24	ON					
	GENDER	−0.002	0.019	−0.087	0.931	No DIF
W25	ON					
	GENDER	−0.007	0.020	−0.333	0.739	No DIF
W26	ON					
	GENDER	0.036	0.018	2.078	0.038	DIF
W27	ON					
	GENDER	−0.003	0.016	−0.174	0.862	No DIF
W28	ON					
	GENDER	0.014	0.024	0.592	0.554	No DIF
W29	ON					
	GENDER	−0.044	0.052	−0.850	0.395	No DIF
W30	ON					
	GENDER	0.017	0.020	0.834	0.404	No DIF
W31	ON					
	GENDER	0.017	0.017	1.039	0.299	No DIF
W32	ON					
	GENDER	0.008	0.030	0.262	0.793	No DIF
W33	ON					
	GENDER	0.022	0.029	0.742	0.458	No DIF
W34	ON					
	GENDER	0.015	0.017	0.885	0.376	No DIF
W35	ON					
	GENDER	0.011	0.013	0.811	0.417	No DIF
W36	ON					
	GENDER	0.019	0.016	1.141	0.254	No DIF
W37	ON					
	GENDER	−0.004	0.019	−0.212	0.832	No DIF
W38	ON					
	GENDER	0.026	0.020	1.289	0.197	No DIF
W39	ON					
	GENDER	−0.027	0.022	−1.250	0.211	No DIF
W40	ON					
	GENDER	−0.035	0.022	−1.581	0.114	No DIF
W41	ON					
	GENDER	−0.001	0.023	−0.061	0.951	No DIF
W42	ON					
	GENDER	0.009	0.022	0.425	0.671	No DIF
W43	ON					
	GENDER	0.038	0.023	1.641	0.101	No DIF
W44	ON					
	GENDER	0.015	0.022	0.691	0.490	No DIF
W45	ON					
	GENDER	0.008	0.021	0.363	0.716	No DIF
W46	ON					
	GENDER	0.028	0.021	1.363	0.173	No DIF
W47	ON					
	GENDER	0.012	0.022	0.565	0.572	No DIF
W48	ON					
	GENDER	−0.014	0.020	−0.671	0.502	No DIF

Continuation of Table4

1	2	3	4	5	6	7
W49	ON					
	GENDER	0.036	0.020	1.806	0.071	No DIF
W50	ON					
	GENDER	0.056	0.021	2.701	0.007	DIF

Table 4 shows the DIF assessment of mathematics test items using ESEM with gender as covariate to model their direct effects on the factor's indicators. The **Table 4** remarked that the covariate has direct significant effects on 10 (20 %) out of 50 test items, namely; (item 3, p-value=0.003, item 4, p-value=0.006, item 6, p-value=0.003, item 11, p-value=0.000, item 14, p-value=0.039, item 18, p-value=0.000, item 21, p-value=0.008, item 22, p-value=0.000, item 26, p-value=0.038, and item 50, p-value=0.0037). This indicated that these items function differentially across gender of the students.

Fig. 3 shows the ESEM with the location as a covariate of the six factors that underlies the test's students' performance. This suggested that 6-factor model, explaining the performance of students in mathematics test item, was viable with, $\chi^2_{(940)}$=4937.553, p<0.05, CFI=0.968, TLI=0.977, RMSEA=0.038, SRMR=0.029, 90 % CI=0.037–0.039. Thus, the extent, to which school location impacts the characteristics that underlie students' performance in mathematics tests, was evaluated. The results are presented in **Table 5**.
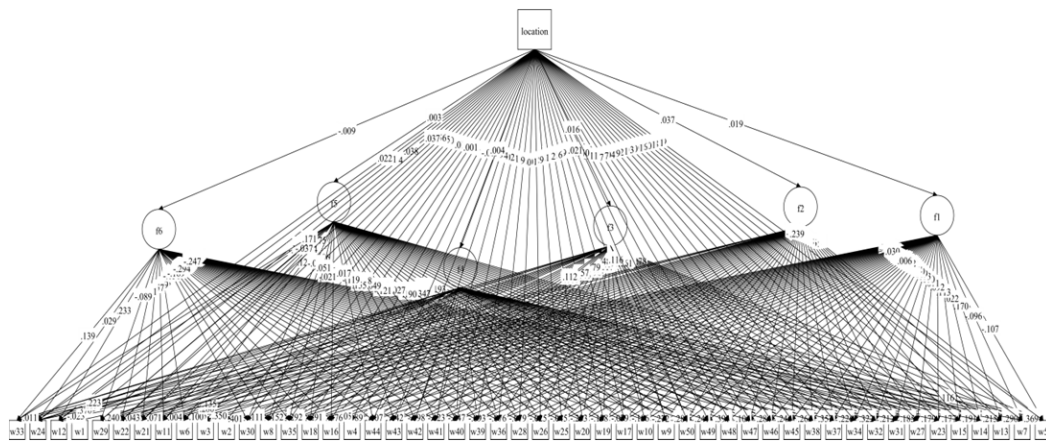

Fig. 3. ESEM with respect to school location as a covariate

Table 5

Model of six factors ESEM with school location as a covariate

Item	Covariate	Estimate	Std. Error (S.E.)	Est./S.E.	p-value	Remarks
1	2	3	4	5	6	7
W1	ON					
	LOCATION	−0.018	0.022	−0.811	0.417	NO DIF
W2	ON					
	LOCATION	0.021	0.017	1.242	0.214	NO DIF
W3	ON					
	LOCATION	0.032	0.017	1.885	0.059	NO DIF
W4	ON					
	LOCATION	−0.005	0.012	−0.396	0.692	NO DIF
W5	ON					
	LOCATION	0.077	0.020	3.833	0.000	DIF

Continuation of Table 5

1	2	3	4	5	6	7
W6	ON					
	LOCATION	−0.004	0.021	−0.203	0.839	NO DIF
W7	ON					
	LOCATION	0.000	0.018	0.014	0.989	NO DIF
W8	ON					
	LOCATION	0.017	0.020	0.833	0.405	NO DIF
W9	ON					
	LOCATION	0.013	0.019	0.673	0.501	NO DIF
W10	ON					
	LOCATION	0.009	0.019	0.490	0.624	NO DIF
W11	ON					
	LOCATION	−0.011	0.020	−0.556	0.578	NO DIF
W12	ON					
	LOCATION	−0.054	0.021	−2.610	0.009	DIF
W13	ON					
	LOCATION	0.009	0.018	0.471	0.638	NO DIF
W14	ON					
	LOCATION	−0.005	0.016	−0.281	0.779	NO DIF
W15	ON					
	LOCATION	−0.015	0.018	−0.831	0.406	NO DIF
W16	ON					
	LOCATION	0.004	0.013	0.349	0.727	NO DIF
W17	ON					
	LOCATION	0.009	0.018	0.520	0.603	NO DIF
W18	ON					
	LOCATION	−0.012	0.012	−0.999	0.318	NO DIF
W19	ON					
	LOCATION	−0.016	0.018	−0.893	0.372	NO DIF
W20	ON					
	LOCATION	0.000	0.018	0.006	0.996	NO DIF
W21	ON					
	LOCATION	−0.032	0.019	−1.740	0.082	NO DIF
W22	ON					
	LOCATION	−0.066	0.018	−3.718	0.000	DIF
W23	ON					
	LOCATION	−0.027	0.016	−1.655	0.098	NO DIF
W24	ON					
	LOCATION	0.066	0.018	3.592	0.000	DIF
W25	ON					
	LOCATION	0.027	0.019	1.434	0.151	NO DIF
W26	ON					
	LOCATION	0.022	0.016	1.334	0.182	NO DIF
W27	ON					
	LOCATION	−0.012	0.016	−0.767	0.443	NO DIF
W28	ON					
	LOCATION	−0.002	0.021	−0.116	0.908	NO DIF
W29	ON					
	LOCATION	0.022	0.018	1.222	0.222	NO DIF
W30	ON					
	LOCATION	−0.011	0.018	−0.593	0.553	NO DIF
W31	ON					
	LOCATION	0.015	0.016	0.929	0.353	NO DIF

Continuation of Table 5

1	2	3	4	5	6	7
W32	ON					
	LOCATION	−0.018	0.025	−0.717	0.473	NO DIF
W33	ON					
	LOCATION	−0.011	0.024	−0.444	0.657	NO DIF
W34	ON					
	LOCATION	0.002	0.016	0.123	0.902	NO DIF
W35	ON					
	LOCATION	−0.007	0.013	−0.555	0.579	NO DIF
W36	ON					
	LOCATION	0.024	0.020	1.202	0.229	NO DIF
W37	ON					
	LOCATION	0.026	0.019	1.406	0.160	NO DIF
W38	ON					
	LOCATION	−0.022	0.020	−1.117	0.264	NO DIF
W39	ON					
	LOCATION	−0.020	0.022	−0.913	0.361	NO DIF
W40	ON					
	LOCATION	0.027	0.022	1.239	0.215	NO DIF
W41	ON					
	LOCATION	−0.031	0.023	−1.351	0.177	NO DIF
W42	ON					
	LOCATION	0.004	0.021	0.178	0.859	NO DIF
W43	ON					
	LOCATION	−0.030	0.021	−1.421	0.155	NO DIF
W44	ON					
	LOCATION	−0.033	0.021	−1.549	0.121	NO DIF
W45	ON					
	LOCATION	−0.019	0.020	−0.933	0.351	NO DIF
W46	ON					
	LOCATION	−0.004	0.020	−0.200	0.842	NO DIF
W47	ON					
	LOCATION	0.005	0.021	0.212	0.832	NO DIF
W48	ON					
	LOCATION	−0.007	0.019	−0.368	0.713	NO DIF
W49	ON					
	LOCATION	0.032	0.019	1.715	0.086	NO DIF
W50	ON					
	LOCATION	0.013	0.020	0.652	0.514	NO DIF

Table 5 shows the DIF assessment of mathematics test items using ESEM with school location as a covariate to model their direct effects on the factor's indicators. The school location showed significant direct effects on 3 (6 %) out of 50 test items, namely; (item 5, p-value=0.000, item 12, p-value=0.009, and item 22, p-value=0.000). The implication is that those items function differentially across the school location of the students.

4. Discussions

The number of factors that underlie the WAEC mathematics test items was established using exploratory factor analysis, resulting in all items having statistically significant loadings on their intended factor. The model fit information was compared, and the 6-factor model demonstrated excellent fit. The six factors were labeled: Number and Numeration, Algebraic Process, Introduction to Calculus, Statistics and Probability, Mensuration, and Trigonometry, respectively. Thus, the instrument is multidimensional, and more than one trait explained the observed variance in

students' performance to the test items. This submission conforms with the findings of [10–12, 47], who posited that standardised instruments, developed for selections and placement and scholarship awards, might not be unidimensional, especially when test items were from various areas. For instance, the National Benchmark Test in South Africa consists of Academic Literacy, Quantitative Literacy, and Mathematics; Graduate Management Admission Test consists of mathematics, verbal reasoning, quantitative reasoning, and English language; Joint Admission and Matriculation Board consist of Mathematics, English, Physics, Chemistry, Biology and so on. Also, this result laid credence to the findings of [5, 6], which argued that there was no evidence of unidimensionality in 2018 Osun State unified multiple-choice mathematics achievement test items. However, this result was in dissonance with the findings of [4, 48] that the unidimensionality of the test was met when comparing three methods for evaluating dimensionality.

Also, ESEM was used to confirm the appropriateness of isolated six factors for its viability; the fit indices were adequately acceptable, making the factors interpretable. The ESEM was a potent tool for determining and identifying items that function differently across a sub-group of students. Few items were flagged to operating differentially, concerning covariate gender and school location, respectively. The findings align with the results of [21, 22, 25, 34, 37, 49] that personal attributes such as gender and school location systematically affected examinees' performance on an item, though the method differed from ESEM.

This study has implications for public examining bodies, test developers, and practitioners' on the existence of DIF, which may inappropriately and differentially affect the performance of examinees with the same ability level in an examination. The implications of the test are severe for the examinees. Hence, stakeholders in educational assessment need to test the items to be free from biases psychometrically. Also, the findings of this study would serve as a scientific basis for drawing inferences, from which conclusions would be deduced, leading to recommendations for better improvements in the process of test development. The study's limitation is that the results might not be generalised because of the scope of this study.

Further study can establish item biasness of other demographic profiles, such as linguistic background, race, ethnicity, etc., using different methods aside from ESEM. The scope can also be expanciated to other regions in the Country, such as south-south, south-east, north-west, north-central, and north-east respectively to establish how these items function. Also, another limitation is that ESEM is a new technique, employed by few researchers to test item biasness and invariance of a measurement instrument across sub-populations (e. g., [30, 39, 42]). Nevertheless, given the peculiarity of the approach and few studies that used it in the past, some caution has to be taken. A recent computational simulation [50] suggests that ESEM has problems with convergence (e. g., the algorithm does not run), especially if the sample sizes are smaller (less than 200 or the ratio of variables to cases may be too small). ESEM is apt when there are considerable cross-loadings of items [50]. In instances where cross-loadings are close to zero and the factor structure is clear (high loadings of items on the relevant factors), ESEM may not be necessary. Hence, ESEM might be an appealing method if a researcher has large samples, and substantive cross-loadings in the model cannot be ignored.

5. Conclusion

It's crucial for item commission writers and test developers of this public examining body to ensure that the test items are valid, reliable, and free from bias. Factors that increase the validity of these test scores need to be improved. Those variables that lower the validity of scores interpretation from the test should be correctly deleted. These unwanted constructs, embedded in the test that affect the decision, should be removed from the scores. Decisions were made for the students based on the outcomes from this test. However, it is possible for items in a test to function inappropriately for a different subpopulation. It is another way to reason if the observed difference is embedded in the construct being assessed or a source of test interpretation bias. Therefore, differential item functioning should be performed on the items and ascertain that it functions equally across a sub-group of students before administering it. This would bring sanity and increase the credibility of the award certificate by the examining board.

Data Availability Statement

The dataset, presented in this study, are available on request. The data are not publicly available due to privacy reasons.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

The study was performed without financial support.

References

[1] Finch, W. H. (2015). Detection of Differential Item Functioning for More Than Two Groups: A Monte Carlo Comparison of Methods. Applied Measurement in Education, 29 (1), 30–45. doi: http://doi.org/10.1080/08957347.2015.1102916

[2] Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. Language Testing, 24 (1), 7–36. doi: http://doi.org/10.1177/0265532207071510

[3] Hemker, B. T., Sijtsma, K., Molenaar, I. W. (1995). Selection of Unidimensional Scales From a Multidimensional Item Bank in the Polytomous Mokken I RT Model. Applied Psychological Measurement, 19 (4), 337–352. doi: http://doi.org/10.1177/014662169501900404

[4] Yu, C. H., Popp, S. O., Digangi, S., Jannasch-Pennell, A. (2007). Assessing unidimensionality: A comparison of Rasch modeling, Parallel analysis, and TETRAD. Practical Assessment, Research and Evaluation, 12 (14), 1–19.

[5] Ayanwale, M. A., Isaac-oloniyo, F. O., Abayomi, F. R. (2020). Dimensionality assessment of binary response test items: A non-parametric approach of bayesian item response theory measurement. International Journal of Evaluation and Research in Education (IJERE), 9 (2), 385–392. doi: http://doi.org/10.11591/ijere.v9i2.20454

[6] Bichi, A. A., Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. International Journal of Evaluation and Research in Education (IJERE), 7 (2), 142. doi: http://doi.org/10.11591/ijere.v7i2.12900

[7] Bolt, D., Stout, W. (1996). Differential Item Functioning: its Multidimensional Model and Resulting Sibtest Detection Procedure. Behaviormetrika, 23 (1), 67–95. doi: http://doi.org/10.2333/bhmk.23.67

[8] Douglas, J. A., Roussos, L. A., Stout, W. (1996). Item-Bundle DIF Hypothesis Testing: Identifying Suspect Bundles and Assessing Their Differential Functioning. Journal of Educational Measurement, 33 (4), 465–484. doi: https://doi.org/10.1111/j.1745-3984.1996.tb00502.x

[9] Nenty, H. (2003). Possible Factors that Influence Student's Pattern of Responses to Mathematics Examination Items. J. Educ. Assess., 17 (4), 47–58.

[10] Tate, R. (2003). A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to Test Items. Applied Psychological Measurement, 27 (3), 159–203. doi: http://doi.org/10.1177/0146621603027003001

[11] Li, Y., Jiao, H., Lissitz, R. (2012). Applying Multidimensional Item Response Theory Models in Validating Test Dimensionality: An Example of K-12 Large-Scale Science Assessment. Journal of Applied Testing Technology, 13 (2), 220–239.

[12] Comrey, A. L., Lee, H. B. (2013). A First Course in Factor Analysis. Hillsdale: Erlbaum, 442. doi: http://doi.org/10.4324/9781315827506

[13] Linn, R. L., Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. Journal of Educational Measurement, 18 (2), 109–118. doi: http://doi.org/10.1111/j.1745-3984.1981.tb00846.x

[14] Karami, H. (2012). An Introduction to Differential Item Functioning. International Journal of Educational and Psychological Assessment, 11 (2), 59–76.

[15] Nunally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

[16] Santelices, M. V., Wilson, M. (2011). On the relationship between differential item functioning and item difficulty: An issue of methods? Item response theory approach to differential item functioning. Educational and Psychological Measurement, 72 (1), 5–36. doi: http://doi.org/10.1177/0013164411412943

[17] Bastug, Ö. (20160. A Comparison of Four Differential Item Functioning Procedures in the Presence of Multidimensionality. Educational Research and Reviews, 11 (13), 1251–1261.

[18] Atalay Kabasakal, K., Arsan, N., Gök, B., Kelecİoğlu, H. (2014). Comparing performances (Type I error and Power) of IRT likelihood ratio SIBTEST and Mantel-Haenszel methods in the determination of differential item functioning. Educational Sciences: Theory & Practice, 14 (6), 2186–2193. doi: http://doi.org/10.12738/estp.2014.6.2165

[19] Magis, D., Béland, S., Tuerlinckx, F., De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. Behavior Research Methods, 42 (3), 847–862. doi: http://doi.org/10.3758/brm.42.3.847

[20] Acar, T. (2012). Determination of a Differential Item Functioning Procedure Using the Hierarchical Generalized Linear Model: A comparison study with logistic regression and likelihood ratio procedure. SAGE Open, 2 (1). doi: http://doi.org/10.1177/2158244012436760

[21] Pedrajita, J. (2009). Using logistic regression to detect biased test items. Internationals Journal of Educational and Psychological Assessment, 2, 54–73.

[22] Abedlaziz, N., Ismail, W., Hussin, Z. (2011). Detecting a Gender-Related DIF Using Logistic Regression and Transformed Item Difficulty. Online Submiss, 5, 734–744.

[23] Asparouhov, T., Muthén, B. (2009). Exploratory Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 16 (3), 397–438. doi: http://doi.org/10.1080/10705510903008204

[24] Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., Craven, R. G. (2019). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal Balance Between Goodness of Fit and Parsimony. Multivariate Behavioral Research, 55 (1), 102–119. doi: http://doi.org/10.1080/00273171.2019.1602503

[25] Vaingankar, J. A., Abdin, E., Chong, S. A., Sambasivam, R., Jeyagurunathan, A., Seow, E. et. al. (2016). Psychometric properties of the positive mental health instrument among people with mental disorders: a cross-sectional study. Health and Quality of Life Outcomes, 14 (1). doi: http://doi.org/10.1186/s12955-016-0424-8

[26] Booth, T., Hughes, D. J. (2014). Exploratory Structural Equation Modeling of Personality Data. Assessment, 21 (3), 260–271. doi: http://doi.org/10.1177/1073191114528029

[27] Gomes, C. M. A., Gjikuria, E. (2017). Comparing the ESEM and CFA approaches to analyze the Big Five factors. Revista Avaliação Psicológica, 16 (3), 261–267. doi: http://doi.org/10.15689/ap.2017.1603.12118

[28] Marsh, H. W., Morin, A. J. S., Parker, P. D., Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis. Annual Review of Clinical Psychology, 10 (1), 85–110. doi: http://doi.org/10.1146/annurev-clinpsy-032813-153700

[29] Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 34 (2), 183–202. doi: http://doi.org/10.1007/bf02289343

[30] Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., Trautwein, U. et. al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Structural Equation Modeling: A Multidisciplinary Journal, 16 (3), 439–476. doi: http://doi.org/10.1080/10705510903008220

[31] Bandalos, S. J., Finney, D. L.; Hancock, G. R., Mueller, R. O. (Eds.) (2010). Factor analysis: Exploratory and confi rmatory. New York: Routledge, 93–114.

[32] Morin B., Marsh, A. J. S., Nagengast, H. W.; Hancock, G. R., Mueller, R. O. (Eds.) (2013). Exploratory Structural Equation Modeling: An Introduction. Structural equation modeling: A second course. Greenwich: IAP.

[33] Tóth-Király, I., Bõthe, B., Rigó, A., Orosz, G. (2017). An illustration of the Exploratory structural equation modeling (ESEM) framework on the passion scale. Frontiers in Psychology, 8. doi: http://doi.org/10.3389/fpsyg.2017.01968

[34] Morin, A. J. S., Arens, A. K., Marsh, H. W. (2015). A Bifactor Exploratory Structural Equation Modeling Framework for the Identification of Distinct Sources of Construct-Relevant Psychometric Multidimensionality. Structural Equation Modeling: A Multidisciplinary Journal, 23 (1), 116–139. doi: http://doi.org/10.1080/10705511.2014.961800

[35] Sass, D. A. (2011). Testing Measurement Invariance and Comparing Latent Factor Means Within a Confirmatory Factor Analysis Framework. Journal of Psychoeducational Assessment, 29 (4), 347–363. doi: http://doi.org/10.1177/0734282911406661

[36] Schmitt, T. A. (2011). Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. Journal of Psychoeducational Assessment, 29 (4), 304–321. doi: http://doi.org/10.1177/0734282911406653

[37] Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. Journal of Psychoeducational Assessment, 29 (4), 322–346. doi: http://doi.org/10.1177/0734282911406657

[38] Bauer, D. J., Belzak, W. C. M., Cole, V. T. (2019). Simplifying the Assessment of Measurement Invariance over Multiple Background Variables: Using Regularized Moderated Nonlinear Factor Analysis to Detect Differential Item Functioning. Structural Equation Modeling: A Multidisciplinary Journal, 27 (1), 43–55. doi: http://doi.org/10.1080/10705511.2019.1642754

[39] Vazsonyi, A. T., Ksinan, A., Mikuška, J., Jiskrova, G. (2015). The Big Five and adolescent adjustment: An empirical test across six cultures. Personality and Individual Differences, 83, 234–244. doi: http://doi.org/10.1016/j.paid.2015.03.049

[40] Lin, C.-Y., Griffiths, M. D., Pakpour, A. H. (2018). Psychometric evaluation of Persian Nomophobia Questionnaire: Differential item functioning and measurement invariance across gender. Journal of Behavioral Addictions, 7 (1), 100–108. doi: http://doi.org/10.1556/2006.7.2018.11

[41] Guo, J., Marsh, H. W., Parker, P. D., Dicke, T., Lüdtke, O., Diallo, T. M. O. (2019). A Systematic Evaluation and Comparison Between Exploratory Structural Equation Modeling and Bayesian Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 26 (4), 529–556. doi: http://doi.org/10.1080/10705511.2018.1554999

[42] Fischer, R., Karl, J. A. (2019). A Primer to (Cross-Cultural) Multi-Group Invariance Testing Possibilities in R. Frontiers in Psychology, 10. doi: http://doi.org/10.3389/fpsyg.2019.01507

[43] Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. Psychological Assessment, 22 (3), 471–491. doi: http://doi.org/10.1037/a0019227

[44] Muthén, L., Muthén, B. (2012). Mplus Version 7 user's guide. Los Angeles.

[45] Kline, R. B. (2005). Principles and practice of structural equation modelling. New York: The Guilford Press.

[46] Oladipupo-Abodunwa, M. A., Adeleke, T. O., Ayanwale, J. O. (2019). Student Mathematics Engagement: Development and Validation of a Measurement Instrument. African Journal of Behavioral and Scale Development Research, 1 (2), 17–23.

[47] Stout, W., Habing, B., Douglas, J., Hae Rim Kim, Roussos, L., Jinming Zhang. (1996). Conditional Covariance-Based Nonparametric Multidimensionality Assessment. Applied Psychological Measurement, 20 (4), 331–354. doi: http://doi.org/10.1177/014662169602000403

[48] Nandakumar, R. (1994). Assessing Dimensionality of a Set of Item Responses-Comparison of Different Approaches. Journal of Educational Measurement, 31 (1), 17–35. doi: http://doi.org/10.1111/j.1745-3984.1994.tb00432.x

[49] Morin, A. J. S., Maiano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). Psychology of Sport and Exercise, 12 (5), 540–554. doi: http://doi.org/10.1016/j.psychsport.2011.04.003

[50] Mai, Y., Zhang, Z., Wen, Z. (2018). Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables. Structural Equation Modeling: A Multidisciplinary Journal, 25 (5), 737–749. doi: http://doi.org/10.1080/10705511.2018.1444993