British Journal of Education

Vol.6, No.1, pp.28-42, January 2018

Published by European Centre for Research Training and Development UK (www.eajournals.org)

DIFFERENTIAL ITEM FUNCTING (DIF) IN THE ANXIETY AND STRESS SCALE ON STUDENTS ACADEMIC ACHIEVEMENT IN CHEMISTRY USING STRUCTURAL EQUATION MODELING (SEM)

Iweka Fedelis (Ph.D)

Dept of Educational Psychology, Guidance and Counselling, University of Port Harcourt, Nigeria.

ABSTRACT: *This is a study that applied two different Structural Equation Modeling (SEM) methods for the detection of gender-and age-related item bias in the anxiety and stress scale on students' academic achievement in chemistry (AASS). Specifically, a multi-group SEM approach was used to investigate both uniform and non uniform item bias in each subscale of AASS separately, and a multidimensional SEM approach that enabled in both subscales of AASS and with regard to both gender and age simultaneously. Results from the multi group SEM approach and the multidimensional SEM approach with regard to the detection of uniform item bias were largely consistent and generally agreed with the results of the ordinal logistic regression, item response theory (IRT) and contingency tables methods reported by Cameron et al 2014; Ogidi and Iweka 2015; and Iweka 2017. Inspection of parameter estimate's of the final model showed that there was a significant positive association between Anxiety and stress (r=0.83, P<.001) indicating that symptom severity with respect to Anxiety goes together with symptom severity with respect to stress. There was a significant negative association between Age and Anxiety (r.024) P<.001), indicating that older students score lower on Anxiety than younger students.*

KEYWORDS: Differential Item Functing, Structural Equation Modeling, Academic Achievement, Anxiety, Stress Scale.

INTRODUCTION

The students academic achievement play an important role in producing the best quality graduates who will become great leaders and manpower for the country, thus, responsible for the country's economic and social development. Academic achievement is affected by numerous factors ranging from social, psychological, economic, environmental to personal factors (Iweka, 2017). Anxiety and stress are part of the psychological factors that can influence the academic achievement of students, especially in chemistry.

The role of anxiety and stress in predicting academic achievement can hardly be overstated. Academic achievement helps to diagnose students' current level of academic functioning as well as the extent to which they have acquired the knowledge of the same grade. It also helps in proper evaluation of students and identification of their strength and weakness (Iweka, 2017). Assessment of the anxiety and stress in academic achievement of students is becoming standard practice in Education. An accurate assessment of the effect of anxiety and stress in academic achievement of students may be invalidated by the occurrence of differential item functioning (DIF). DIF, also referred to as item bias occurs when two people with the same value on trait of interest have a different probability of giving a certain response on an item from a questionnaire or test that measures the trait of interest, due to differences on other variables such as age, gender, attitude, mood, etc. Mellenbergh (1989) gave a formal definition

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

of item bias: An item measuring trait T is unbiased with respect to another variable V, if and only if: $f_1(X/V = V, T = t) = f_2(X/T = t)$ where f_1 is the distribution of the item responses given the values v and t of variables V and T, and f_2 is the distribution of item responses given only the values t of variable T. Mellenbergh emphasized the generality of the definition, where the variables X,V and T may have nominal, ordinal or interval measurement scales.

In the presence of item bias, differences between two people on observed item scores may not reflect "true" differences on the trait variable. For example boys and girls may score differently on an item that measures wellbeing even though their well being does not differ. If a bias is uniform, it is constant for all levels of the latent trait, for example, the size of the bias is independent of the level of well being. When the bias is non uniform, it differs for different levels of the latent trait, for example, the difference may be larger for larger levels of well being. Statistical methods for the detection of item bias can be distinguished based on their operationalization of the trait variable T. One group of the methods use the summary of the observed item scores or scale scores to operationalise the trait variable, such as, log-linear models, contingency tables methods, logistic regression models, standardization method, and another group of methods operationlise an unobserved latent trait variable, such as, item response theory (IRT) analysis and structural equation modeling (SEM) methods (Millsap & Everson, 1993). It is important to distinguish between methods that can detect uniform item bias and methods that can also detect non uniform item bias. Ogidi and Iweka (2010) investigated the equivalent of three different bias detection methods for the detection of gender- and age-related bias in the items of the Academic Anxiety and Stress Scale (AASS): they applied ordinal logistic regression, IRT, and contingency tables methods to investigate item bias in the Anxiety and Stress subscales of AASS respectively. All three methods were used to detect uniform item bias only.

Although the researchers mentioned SEM methods as a fourth option that can be applied to investigate item bias, they did not incorporate SEM methods in their comparison. SEM methods may have several important advantages for the detection of item bias. The multi-group SEM approach can be applied to detect bias in observed item scores with respect to group membership such as gender or age category; and a continuous latent trait variable such as stress and anxiety. Advantages of the multi-group SEM approach are that it uses a latent trait operationalization, it enables the detection of both uniform and non uniform bias, and possible item bias can be taken into account to assess true differences between groups.

Purpose of the Study

1. To illustrate how to apply the multi-group SEM approach to investigate both uniform and non uniform gender-and-age related item bias in each subscale of the academic anxiety and stress scale (AASS).

2. To illustrate how to apply the multi-dimensional SEM approach to both subscales of AASS and investigate uniform gender-and age-related item bias simultaneously.

Research Questions:

1. What is the extent of the uniform and non uniform gender-and -age- related item bias in each subscale of AASS when multi-group SEM approach was applied.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

2. What is the extent of the uniform gender-and age-related item bias when multi-dimensional SEM approach is applied to both subscales of AASS.

METHODS

A total of 200 senior secondary chemistry students were given the Academic Anxiety and Stress Scale (AASS) to complete or fill. The AASS is a 14-item self report instrument that consists of an Anxiety (AASS-A) seven items and stress (AASS-S) seven items subscale where higher scores represent greater symptom severity. All items are answered on an ordinal response scale with four response categories (0-3). The sample consisted of 100 male students and 200 female students with ages ranging from 18-24 years (mean age = 21.5, SD = 18). Mean anxiety scores (AASS, - A) were 7.7 with a standard deviation of 4.7 and mean stress scales (AASS – S) were 4.9 with a standard deviation of 4.2.

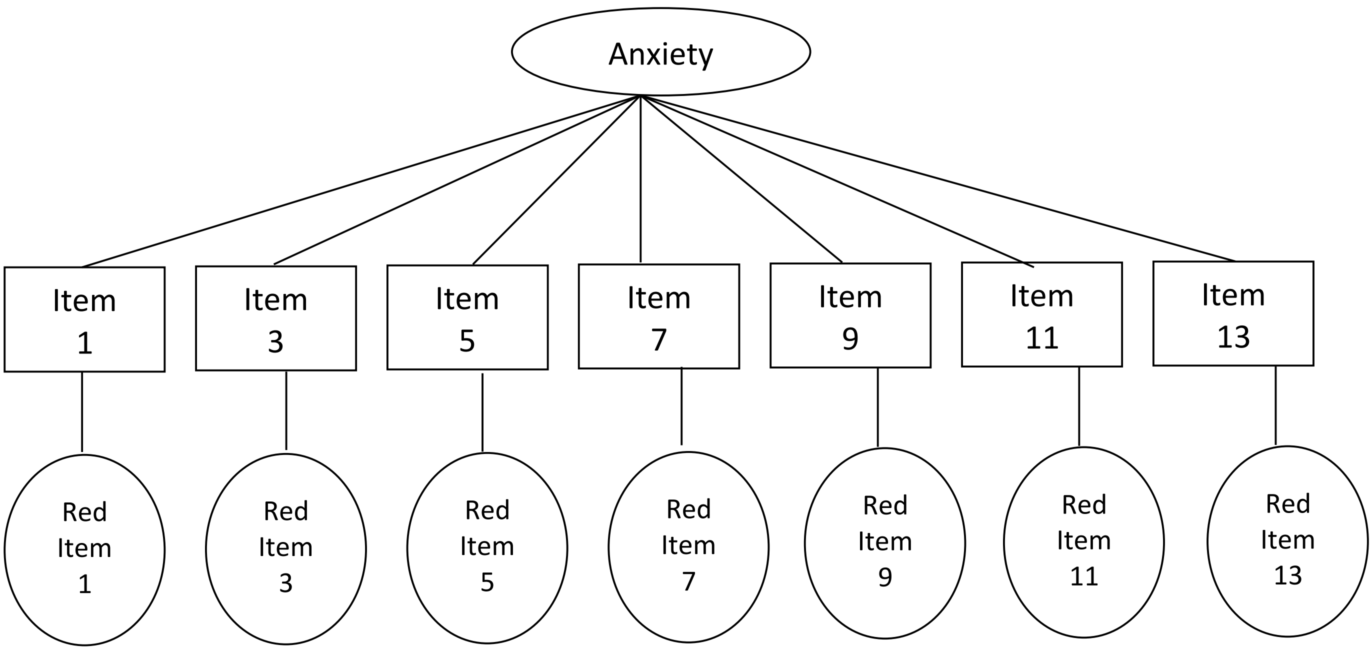
Multi-group SEM Procedure

Gender –and age – related item bias was investigated for the anxiety and stress subscales for the academic anxiety and stress scales (AADS). Separately by comparing a criterion and experimental group. For age, there were 200 participants in the criterion group (21.5 years) and 100 for experimental group (21.5 years). For gender, there were 100 participants in the criterion group (girls) and 100 in the experimental group (boys). The categorization of age and the separate analysis of the subscales of the AASS were chosen in order to enable comparison of the (SEM) Structural Equation Modeling results with the results from the other detection methods as reported by Cameron, Scott, Adler & Reid (2014).

A graphical representation of the multi-group model for item bias detection is presented as follows:

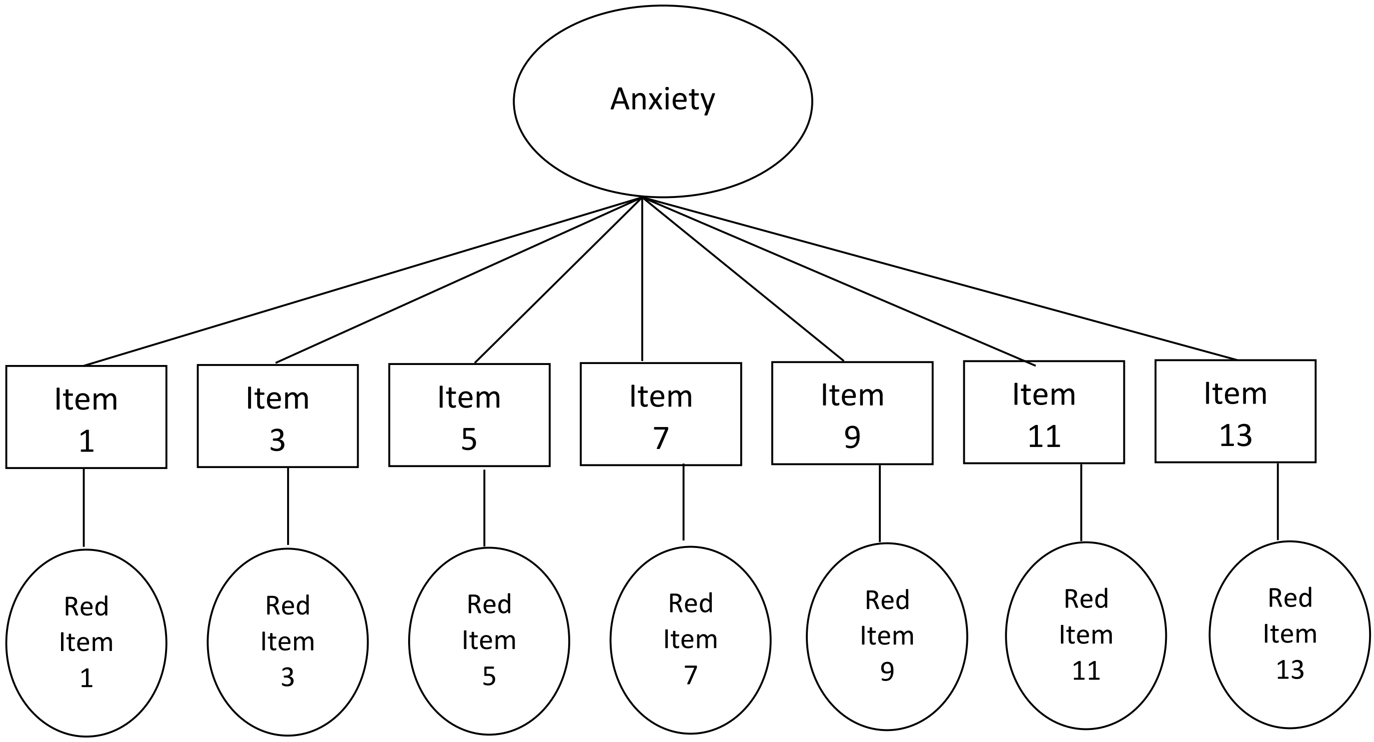
The graph Representation

Reference group: Female students



ISSN 2055-0111(Print), ISSN 2055-012X(Online)

Experimental Group: Male Students



The graphical representation shows two-group measurement model for gender –related item bias detection in the anxiety sub-scale of the AASS. Similar models have been used for the detection of age-related item bias in the anxiety subscale of the AASS, and for the detection of gender-and age-related item bias in the stress subscale of the AASS.

The squares represent the underlying continuous variables associated with the observed item responses of item 1 to item 13. The circle at the top is the underlying common factor anxiety, which represents every thing that item 1 to 3 have in common. Each item is associated with a residual factor which represents everything that is specific to the corresponding item. Item bias is operationalized as across-grouping differences in intercepts or uniform and factor loadings (non uniform).

The procedure for the multi-group model for item bias detection using Structural Equation Modeling (SEM) is as follows:

In stage 1, the model of underlying continuous variables that represent the observed discrete variables was used to estimate thresholds and Polychromic correlations under the assumption of vicariate normality in both group was evaluated using the root mean square error of approximation. When the hypothesis of vicariate normality under equal threshold holds for all pairs of variables, the estimated polychromic correlations, variances and means of the underlying continuous variables can be used in subsequent analyses of stage 2. When the hypothesis of bivariate normality does not hold, then this indicates that the assumption of multivariate normality under equal thresholds is not tenable. A possible solution for this problem is to eliminate the offending variables.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

In stage 2, step 1, the estimates from the underlying variables from stage 1 were used to estimate a multi-group common factor model, such as a one factor model for "Anxiety", with seven indicator items, for both boys representation of the multi-group model for the item bias detection. The measurement model has no across-group constraints. The appropriateness of the measurement model was evaluated using overall goodness of fit.

The chi-square test can be used to evaluate exact goodness of fit, where a significant chi-square value indicates a significant difference between data and model. In step 2, the No item Bias model was fitted to the data, where all measurement parameters were constrained to be equal across groups. Item bias was operationalised as across-group differences between common factor loadings, that is, non uniform items bias; across-group differences in the extent to which an item measures the latent trait variable to test for the presence of item bias, the No item Bias Model can be compared to the measurement model. The chi-square difference test was used to test the difference in exact fit, where a significant chi-square difference indicates that the No item Bias model has significantly were fit as compared to the measurement model.

If the invariance restrictions of the No item Bias Model led to a significant deterioration in model fit, this indicated the presence of item bias.

As the chi-square test statistic is very sensitive to large sample sizes, to guard against false positives, the researcher considered P values <.001 to indicate statistical significance.

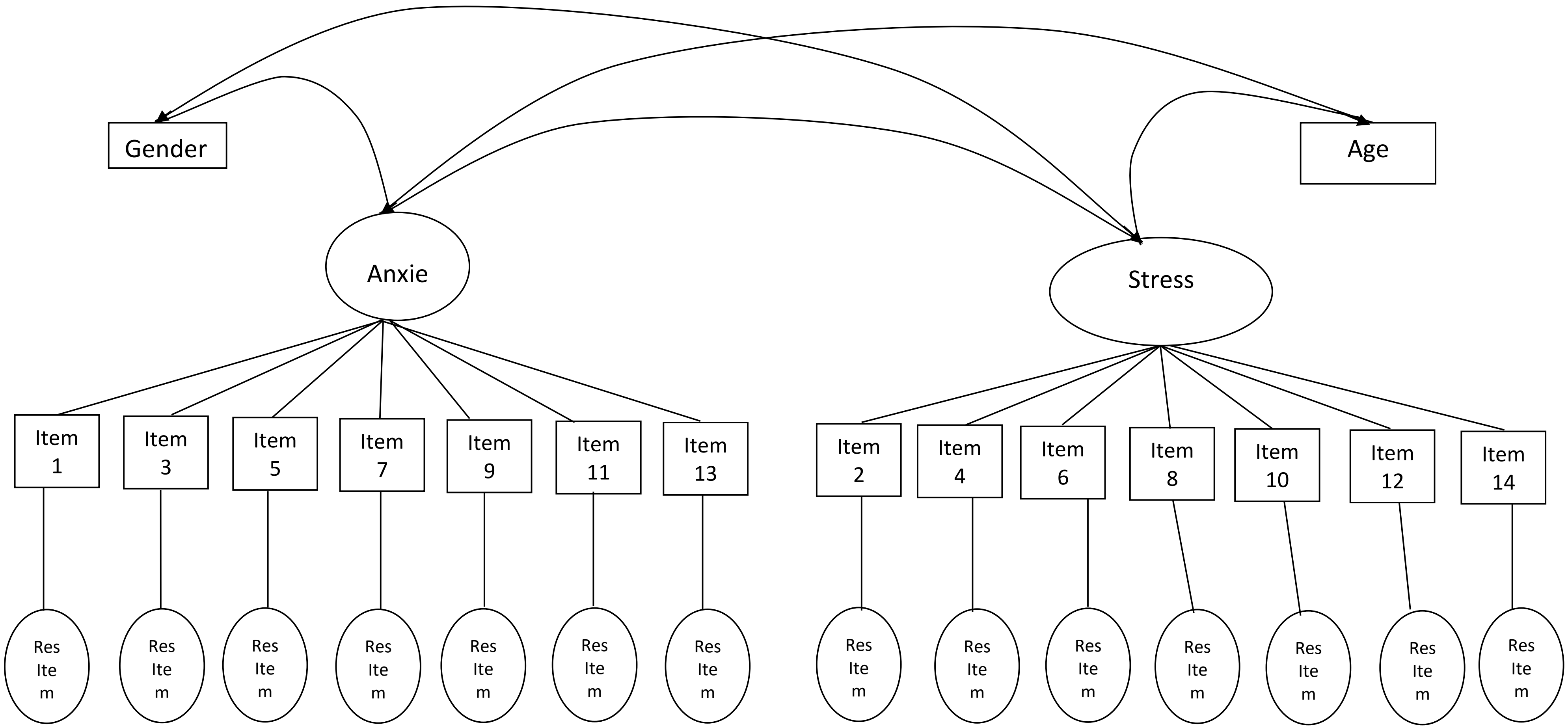
In step 3, a step by step modification of the No item Bias model was used to arrive at the final model in which all items that showed item bias were taken into account. The identification of item bias was guided by an iterative procedure, where each across-group constraint was set free one at a time, and the freely estimated parameter that led to the largest improvement in fit was included in the model. Each indication of bias was tested by evaluating the improvement in model fit using the chi-square difference test to evaluate differences in exact fit. To guard against false positives, the researcher considered P-values >.001 to indicate statistical significance. The trial model was compared to the measurement model to test equivalence of exact fit as an indication that all apparent item bias was taken into account. To give an indication of the size of the detected item bias, the researcher calculated cohsen's d' effect size indices for the impact of both uniform and non uniform item bias on the differences between the item means across groups, where values of 0.2, 0.5 and 0.8 indicate small, mechum and large effects.

In step 4, the estimates of common factor means of the trial model in which all apparent item bias was taken into account, was used to assess true differences between the groups. Cohen's d effect size was calculated to give an indication of the size of the difference. Also, the overall impact of item bias on the assessment of true differences can be evaluated through the comparison of effect size indices before and after taking possible item bias into account.

Multi-dimensional SEM Procedure

Gender-and age-related item bias was investigated for the academic anxiety and stress scale (AASS) simultaneously, by including both age and gender as exogenous variables in the multi-dimensional model. The graphical representation of the multi-dimensional model for item bias detection is as follows

ISSN 2055-0111(Print), ISSN 2055-012X(Online)



ISSN 2055-0111(Print), ISSN 2055-012X(Online)

This graphical representation of the multi-dimensional model for item bias detection is a multi-dimensional "no item bias" model for gender-and age-related item bias detection in the anxiety and stress sub-scales of the (AASS). The squares represent the underlying continuous variables associated with the observed item responses of item 1 to item 14. The circles at the top are the underlying common factors of Anxiety and stress. Anxiety represents everything that item 2 to item 14 have in common. Each item is associated with a residual factor, which represents everything that is specific to the corresponding item. The multi-dimensional model includes two exogenous variables: Gender and Age.

Uniform item bias is operationalised as significant direct effects of the exogenous variables on the indicator variables, that is, item 1 to item 14.

The multi-dimensional structural equation modeling procedure that is used in this study is also known as the restricted factor analysis (RFA) procedure. It yields equivalent results as multiple-indicator multiple-cause (MIMIC).

The procedure for item bias detection using the multi-dimensional approach was largely similar to the procedure for item bias detection using the multi-group approach which has been explained. The researcher will therefore only describe the differences in the procedures as follows:

In stage 1, correlations between all variables in the model were estimated. In step 1 of stage 2, the estimates from the underlying variables from stage 1 were used to establish a multi-dimensional measurement model that included the common factors "Anxiety" and "stress" each with seven indicator variables. In step 2 stage 2, the multi-dimensional measurement model was extended to include the variables "Age" and "Gender". These variables were allowed to correlate with the common factors, but all direct effects of Age and Gender on the items were constrained to zero.

This model is referred to as the No item Bias Model. The overall model fit of this model was used to give an indication of the presence of item bias, where the Root mean Square Error of Approximation (RMSEA) value <.08 was taken as a global indication that there was no presence of item bias. In step 3, an iterative procedure was used, where each constrained direct effect of the exogenous variables age and gender was set free to be estimated one at a time, and the freely estimated parameter that led to the largest improvement in fit according to the chi-square difference test was included in the model, where P<.001 was taken to indicate statistical significance.

When freeing additional parameters did not lead to a significant improvement in model fit, this was taken as an indication that all apparent bias was taken into account. The important criterion for item bias was evaluated using the standard direct effects, which can be interpreted as effect size r, with values of 0.1, 0.3 and 0.5 indicating small, medium, and large effect sizes (Cohen, 1988). In step 4, the correlations between the exogenous variables age and gender and the common factors of the final model, in which all apparent bias has been taken into account were used to assess true differences between the gender, and true associations with age. The overall impact of item bias on the assessment of true differences between the genders and true associations with age can be evaluated through the comparison of correlations before and after taking possible item bias into account.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

RESULTS

Model fit results of the item bias detection producer are presented in table 1. An overview of the items that were identified as having bias by using structural equation modeling approach are given in table 2 on first report and results of the multi-group SEM approach and the those of the multidimensional SEM approach.

Multi-group SEM Approach Results.

Results of stage 1 indicated that the hypothesis of bivariate normality under equal threshold was tenable for all items pairs, for both subscales and both gender and age groups. Estimated polychoric correlations, variances, and means were used in subsequent analysis of stage 2. The report of the results of gender-and-age-related item bias for each subscale of the AASS was done separately.

Anxiety-related item bias Result:

Results of stage 2 indicated that the measurement model showed close approximate fit (Model 1a, Table 1). Imposition of equality constraints on measurement parameters across groups yielded the No item Bias Model (Model 1b). The No Item Bias Model showed a significant deterioration in model fit as compared to the measurement model, indicating the presence of gender-related item bias of the AASS-A, as shown in table 1.

Indications of uniform bias were detected for item 9 (CHISQ diff (I)-14.54, p<.001) and for item 11 (CHISQ diff (I) = 30.57, p<.001).

The final model, in which both biases were incorporated in the model showed close approximates fit (model ic, Table 1).

Age-related item bias-the measurement model showed close approximate fit (model 3a, Table. 1). Age-related item bias detection model showed close approximate fit (model fit, indicating the presence of age-related item bias of the AASS-A. Two items with uniform bias and one item with non-uniform bias were identified. The final model that incorporated these three biases (model 3c) showed equivalent fit compared to the measurement model (see tabl5656+45+6e 1).

445Uniform bias was detected for item 1 (CHISQ diff (I) = 18.36, p<.001) and for item 13 (CHISQ diff (I) = 50.78, p<.001), where as non uniform bias was detected for item 3 (CHISQ diff (I) = 12.31, p<.001).

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

Table 1: Goodness of overall model fit and difference in model fit of the model for gender-and age-related item bias detection models in stage 2; for both the multi group structural equation modeling approach, and the multidimensional structural equation modeling approach.

	Model	Df	CHISQ P-Value	RMSEA (90C1%)	Compared to Diff	CHISQ	P Value
	\ulti-group gender-related -as detection Anxiety						
1a	Measurement model	28	50.64.005	0.039 (0.02; 0.056)			
1b	No item bias model	40	126.4 <.001	0.064 (0.051; 0.076)	Model 1a 12	75.76	<.001
1c	Final model stress subscale	38	81.29 <.001	0.040 (0.032; 0.060)	Model 1a 10	30.65	<.001
2a	Measurement model	28	46.26.016	0.035 (0.015; 0.052)			
2b	No itme bias model	40	120.9 < 00	0.062 (0.049; 0.074)	Model 29 12	74.63	<.001
2c	Final model multi-group age-related item bias detection Anxiety subscale	37	70.02 <001	0.041 (0.026'C0.055)	Model 29 9	23.76	.005
3a	Measurement model	28	61.02 <.001	0.047 (0.031; -.063)			
3b	No item bias	40	163.2 <.001	0.48 (0.04; 0.088)			
3c	Final model stress subscale	37	8.71 <.001	0.048 (0.04; 0.062)			
4a	Measurement model	28	42.59.038	0.031 (0.008; 0.049)			
4b	No item bias model	40	357.2< 0.001	0.122 (0.111; 0.134)			
4c	Final model	76	83.24 <.001	0.052 (0.038; 0.066)			
	Multidimensional gender-and age-related item bias detection anxiety and stress subscales						
5a	Measurement model	100	485.05<.001	0.07 (0.065; 0.077)			
5b	No item Bias mode	88	1029.8<.001	0.093 (088; 0.098)			
5c	inal model		455.71 <.001	0.063 (0.057; 0.068)			

N=1068. Overall model fit and difference in fit was evaluated using WLS Chi-Square values that are provided in the standard LISREL output (denoted C2 NNT).

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

True Difference Between the Groups:

Inspection of common factor means showed that male students' score significantly lower on the anxiety factor as compared to female (d=-0.30, p<.001) and that students older than 22 years scored significantly lower on the anxiety factor compared to students younger than 22years (d=-0.76, p<.001). If item bias would not have been taken into account the true differences between gender and age groups would have been estimated to be similar (d)=-0.26, p<.001; and d=-0.73, p<.001, respectively).

Stress Subscale

Gender-related item bias: The Measurement Model indicated close approximate fit (model 2a Table 1). Comparison of the No item Bias Model with the Measurement model indicated the presence of gender-related item bias the AASS-3. Step by step modification of the No Item Bias Model yielded the final model in which all bias was taken into account (Model 2c, Table 1). For item 4, both uniform bias (CHISQ diff (1) = 16.5, p<.001) and non uniform bias were detected (CHISQ diff (I) = 14.47, p<.001). In addition, non uniform bias was detected for item 10 (CHISQ diff (I) = 18.85, p<.001).

The final model showed equivalent fit as compared to the measurement model (See table 1).

Age-related item bias – The measurement model showed close approximate fit (Model 4a, Table 1), Bias Model indicated the presence of age-related item bias of the AASS-S (See table 1).

Uniform bias was detected in four items, and nonuniform bias was detected in three items, where one item showed both uniform and nonuniform bias. The final model which included all apparent bias, showed close approximate fit (model 4c). Although the final model did not yield equivalent fit as compared to the measurement model (see table 1). Freeing additional parameters did not significantly improve model fit. Uniform bias was detected for item 4 (CHISQ diff (1) = 63.68, p<.001), item 6 (CHISQ diff (1) = 30.57 p<.oo1).

Uniform bias was detected for item 4 (CHISQ diff (1) = 16.06, p<001) and item 12 (CHISQ diff (1) = 20.51, p<001).

Result of gender-and age-related item bias detection in the anxiety and stress scales of the AASS table 2. Questionnaire using the multi group structural equation modeling (SEM-Ma) and multidimensional structural equation modeling (SEM-MD) approaches.

Result of gender-and age-related item bias detection in the anxiety and stress scales of the AASS table 2. Questionnaire using the multi group structural equation modeling (SEM-Ma)

	Item	Gender-related item bias					Age-related item bias				
		LOGR[1]	IRT[2]	CONT[3]	SEMMG[4]	SEMMD[5]	LOGR[1]	IRT[2]	CONT[3]	SEMMG[4]	SEMMD[5]
	AASS-A										
1	I feel tense or would out	-	-	-	-	**-0.77[a]**	**-0.61[a]**	**-3.78[a]**	**-0.22[a]**	**-0.22[a]**	**0.09[a]**
3	I get a feeling as if	-	-	-	-	-	-	-	-	**0.09[b]**	

and multidimensional structural equation modeling (SEM-MD) approaches.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

	something anful…										
5	Waving thoughts go through my mind	-	-	-	-	-	-	-	-	-	-
7	I can sit at ease and feel relaxed	-	-	-	-	-	-	-	-	-	-
9	I get a--- feeling like “butterflies in the stomach	-0.49[a]	-	**3.64[a]**	0.16[a]	**-0.13[a]**	-	-	-	-	-
11	I feel restless as if I hays to be on the sudden move	0.58[a]	**-0.62a**	**4.70[a]**	**0.12[a]**	-	-	-	-	-	-
13	I get sudden feelings of panic	-	-	-	-	-	-	-	-	0.22[a]	0.07[a]
	AASS-S										
2	I still contribute to class discussions	-	-	-	-	0.07[a]	-	-	-	-	-
4	I can write essays and assignments	-		0.01[b]	-	-	-	-	-	0.01[b]	**0.13[a]**
6	I feel happy to attend lecture	-	-	-	-	**0.11[a]**	**0.11[a]**	**-0.77[a]**	**-5.16[a]**	**-0.56[a]**	-0.23[a]
8	I feel as if I can do oral presentation	-	-	-	-	**-**	**0.92[a]**	**1.03[a]**	**6.72[a]**	**-**	**0.14[a]**
10	I have lost interest due to my continuous poor performance	-	-	-	-0.01[b]	-	00.60[a]	**-0.52[a]**	**-3.66[a]**	**-0.34[a]**	-0.14[a]
12	I look forward to sitting for exams	-	-	-	-	-	-	**-**	**-**	**-0.01**	-
14	I can enjot studing for exams	-	-	-	-	**0.12[a]**	-	**-**	**-**	**-0.29[a]**	**-0.18[a]**

Results are compared to the item bias detection results from the ordinal logistic regression methods (LOGR), the item response theory method (IRT), and the contingency table method (CONT.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

Uniform item bias

Non-uniform item bias. Results meeting the criteria for important item bias are marked in bold, results meeting only the significance criterion are marked in italics. Number are given bias detection results that were considered statistically significant 1log odds ratios are presented, where items were regarded as having important bias if the absolute magnitude of the log odds ratio was greater than 0.64 and p<.001.2 contracts with absolute value greater than 0.50 and p<0.05 were taken as an indication of important item bias. [3]Standard Liu-Agreti cumulative common log odds ratios (LOR Z) are presented where absolute values <3 and p<.001 are considered important item bias.

[4]Effect size indices are presented difference in intercept parameters between the groups divided by the pooled standard deviation. For non-uniform item bias these refer to the difference in factor loading parameter multiplied with the difference in common factor means between the groups divided by the pooled standard deviation. Effect sizes larger than 120 and P<.001 are indicative of important item bias. [5]Effect size indices are presented which are the standardized direct effect of Gender/Age on the specific item. Effect sizes larger 20 and p<.001 are indicative of important item bias.

True Differences Between the Groups.

There were no significant differences between the male and female.

Students (d = 0.03, P=64) or between the age groups (d=0.03, p=0.70) with respect to their scores on the underlying stress factor. Before taking into account item bias true differences between the male and female students, were estimated to be similar cd = - 0.34, P<.001). Thus, if item bias would not have been taken into account, the difference in stress severity between the age groups would have been overestimated .

Multidimensional SEM approach results of stage 1 indicated that the hypothesis of bivariate normality under equal threshold was tenable for all combinations of items and exagerious, variables. The estimated (polychoric) correlations, various, and means of all variables were used for subsequent analysis in stage 2. In stage 2, the measurement model that included both AASS subscales showed reasonable approximate fit (model 5a, Table). The No item Bias model that include the variables Age and Gender did not show acceptable fit (model 5b), indicating the presence of item bias (see Table 1). Uniform bias was detected in four items of the AASS-A, and six items of the AASS-S the final model, which included all apparent bias, showed reasonable approximate fit (model 5c, Table 1).

The Anxiety Subscale

Gender related bias of the AASS-A was detected for item 9 (CHISQ diff(1)=24.2, P<.001) and item 11(CHISQ diff (1) = 97.9, P<.001). age-related bias of the AASS-A was detected for item 1 (CHISQ diff (1) = 64.0, P<.001) and item 13 (CHISQ diff 91) = 104.8 P<.001).

The Stress Subscale

Gender-related bias of the AASS-S was detected for item 2 (CHISQ diff (1) = 22.9, P<.001), item 6 (CHISQ diff 91) = 28.2, P<.001), and item 14 (CHISQ diff 91) = 28.9, P<.001). age-related bias of the AASS-S was detected for item 4 (CHISQ diff (1) 25.9, P<.001), 8 (CHISQ

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

diff (1) 20.8, P<.001), item 10 (CHISQ diff (1) = 37.6, P<.001) and item 14 (CHISQ diff 91) = 52.5, P<.001).

True Differences and Association

Inspection of parameter estimates of the final model showed that there was a significant positive association between Anxiety and stress (r=0.83, P<.001) indicating that symptom severity with respect to stress.

There was a significant negative association between Age and Anxiety (r = -0.24, P.001), indicating that older students scored lower on Anxiety than younger students. There was also a significant negative association between gender and Anxiety (r = -0.216, P<.001), indicating that male students scored lower on Anxiety than female students. The association between gender and stress was negative, and between Age and stress was positive, but neither was significant (r=-0.04, p=.19, and r=0.01, p=.83, respectively). Lastly, there was a significant positive association between Age and Gender (r=0.11, P<.001), indicating that male students were on average significantly older than the female. If item bias would not have been taken into account, the Patten and size of true differences and associations would have been estimated to be similar, with the exception of the association between Age and stress. Without taking into account item bias this association was estimated to be negative and significant (r=-0.10, P<.001).

DISCUSSION

The study illustrated how to apply two different SEM methods for the detection of gender-and age-related item bias in the anxiety and stress subscales of the AASS, to account for item bias, and to more validity evaluate students' anxiety and stress. Specifically, a multi group SEM approach was used to investigate both uniform and non-uniform item bias in each subscale of the AASS separately, and a multidimensional SEM approach that enable the investigation of uniform item bias in both subscales of the AASS and with regard to both gender and age simultaneously. The multi-group SEM method identified a total of 10 items with bias, of which 8 items showed uniform bias and 4 items showed non uniform bias. The multidimensional SEM method was used to detect only uniform bias and identified a total of 10 items as biased. These indications of bias may invalidate the comparison of item scores of male and female students, and subjects with different ages or from different age groups. However, the overall effect of detected item biases on the assessment of true differences in and associations with anxiety and stress severity was generally small. Only for the stress subscale of the AASS the detected item bias would have led to an overestimation of the differences between age groups (multi-group SEM) or between students with different ages (multidimensional SEM). Without taking into account item bias older students would have been estimated to be less stressed than younger students, whereas after taking into account item bias, this difference was no longer significant. The detected item biases indicated that younger students experience more stress symptoms as compared to the older students, relative to the level of stress.

In contrast, the gender-and age-related biases that were detected in the items of the anxiety subscale of the AASS did not lead to different conclusions at the subscale. A possible explanation for these results is that the detected item biases canceled each other out at the subscale level.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

The results from the present study support valid comparisons between female and male students on both the anxiety and stress subscales of the AASS, whereas a valid comparison between students of different ages is only supported for the stress subscale. In the present research, non-uniform bias was only investigated with the multi-group SEM approach but not with the multidimensional SEM approach. Although it has been shown that investigation of non-uniform item bias is possible by including interaction terms between the underlying trait of interest and the other exogenous variable (Iweka, 2017).

Both SEM approaches identified the same items with uniform gender and/or age-related bias in the anxiety subscales of the AASS. In addition, the detection of uniform age-related bias in the stress subscale of the AASS was largely consistent across SEM approaches with agreement on 4 items, although less so with regard to the detection of gender-related bias.

Results from the multi-group SEM approach and the multidimensional SEM approach with regard to the detection of uniform item bias were largely consistent and generally agreed with the results of the ordinal logistic regression, item response theory (IRT(and contingency tables methods reported by Cameron et al 2014; Ogidi and Iweka 2015; and Iweka, 2017.

CONCLUSION

SEM provides a flexible tool for the investigation of item bias on the anxiety and stress scale in students' academic achievement in chemistry. Both the multi-group SEM approach and multidimensional SEM approach can be applied to detect bias in observed item scores. Advantages of the multigroup SEM approach are that it uses a latent trait operationalization, it can detect both uniform and non-uniform bias, and possible item bias can be taken into account to assess true differences between groups.

In addition, the extension to multidimensional models enables the investigation of item bias.

REFERENCES

Cameron, I. M, Scott, N. W., Adler, M. & Reid, I. C. (2014). A comparison of three methods of assessing differential item functing (DIF) in the Hospital Anxiety Depression Scale: Ordnial logistic regression, Rasch analysis and Mandel Chi-Square procedure. Quality of life Research, 23, 2883-2888.

Cohen, J. (1988). Statistical Power analysis for behavioural sciences (2nd ed.). Hillsdale, N.J: Lawrence Erlbatum.

Iweka, F. O. E. & Wokoma, T. Abbott (2017). Attitudes of teachers towards application of item Response theory in technical colleges in Rivers State. British Journal of Education. 5, (6), 39-356.

Iweka, F. O. E. (2005). Using structural equation modeling to detect response shifts and true change. Quality of life Research, 14, 857-598.

Iweka, Fidelis & Tei-Firstman R. I. (2017). Comparison of cut off scores using two setting methods. International Journal of Quantitative and Qualitative Research methods. 5 (2) 42-6.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)

Iweka, Fidelis (2017). Learning Environment as correlates of students academic achievement in junior secondary school integrated science. International Journal of Education, learning and Development, 5 (4), 48-54.

Mellenbergh, G. J. (1989). Item bias and item response theory. International Journal of Educational Research, 13, 127-142.

Millsap, R. E. & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. Applied psychological measurement, 17, 297-334. Doc:10, 1177/0/4662 169301700401.

Ogidi, R. & Iweka, F. O. E. (1998). Simulation study of item bias detection with restricted factor analysis. Structural Equation Modeling. 5, 107-124.

Spielberger, C. A. (1980). Statistically based tests for the number of common factors. In paper presented at the annual meeting of the psychometric society. Lowa city .I.A.

ISSN 2055-0111(Print), ISSN 2055-012X(Online)