

Factor structure and measurement invariance of the TIMSS 2015 mathematics attitude questionnaire: Exploratory structural equation modelling approach

Seyma Uyar ^{1,*}

¹Burdur Mehmet Akif Ersoy University, Faculty of Education, Department of Educational Sciences, Burdur, Turkey

ARTICLE HISTORY

Received: Sep. 18, 2020

Revised: July 31, 2021

Accepted: Sep. 23, 2021

Keywords:

Exploratory structural equation model,
Measurement invariance,
TIMSS,
Multi-group confirmatory factor analysis,

Abstract: In the current study, the appropriateness of the Mathematics Attitude Questionnaire administered to middle school 8th grade students in the TIMSS 2015 application to the exploratory structural equation and confirmatory factor analysis models was examined. The study was conducted on 6079 students making up the sample of Turkey. In the TIMSS 2015 application, the attitude items are presented under four headings called students' interest in mathematics, students' views on engaging teaching in mathematics lessons, students' self-confidence in mathematics, and students' value mathematics. As a result of the investigation of the factor structure of these items, the attitude questionnaire with its 5 factors and 35 items was accepted to be suitable for the Exploratory Structural Equation Model (ESEM). Moreover, invariance of the TIMSS 8th grade mathematics attitude questionnaire depending on gender was investigated at six stages as configural, weak (metric), strong (scalar), strict, variance-covariance, and latent mean invariance through ESEM. It was concluded that the questionnaire satisfied all the invariance conditions.

1. INTRODUCTION

Comparative studies are thought to have a large share in shaping the education policies of countries. For this reason, it is seen that many countries take part in comparative studies, which include international measurement and evaluation studies such as PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study), and TALIS (The OECD Teaching and Learning International Survey). For example, TIMSS is a survey research conducted by the International Education Achievement Assessment Organization (IEA) for the comparative evaluation of the knowledge and skills acquired by the 4th and 8th grade students in the fields of mathematics and science at four-year intervals. In the TIMSS application, information about students' performances, education systems, curricula, student characteristics, characteristics of teachers, and schools is collected, and student achievement is compared with

*CONTACT: Şeyma Uyar ✉ syuksel@mehmetakif.edu.tr 📍 Burdur Mehmet Akif Ersoy University, Faculty of Education, Department of Educational Sciences, Burdur, Turkey

e-ISSN: 2148-7456 /© IJATE 2021

other countries or in different subgroups constructed based on gender and socioeconomic level in the same sample (Ministry of National Education [MNE], 2016). However, the biggest problem in such studies is whether the measurement tools applied to the compared groups are really equivalent in terms of the measured property. When a cognitive or behavioural feature is to be measured under different conditions (measurement time, test application methods or group), this construct may mean different for each condition (Bornstein, 1995). In this case, it is not easy to completely distinguish the difference between individuals from measurement time, measurement method or group membership (Horn & Mcardle, 1992). In order to be able to compare a construct correctly and appropriately in the given conditions, it is necessary to examine the invariance of the meaning of the construct under these conditions (Putnick & Bornstein, 2016). For example, when a factor structure shows similar conformity for data obtained at different times, it is necessary to talk about a measurement invariance within time (longitudinal) (Little, 2013) or if the factor structure remains the same as a result of administration of a test on the Internet environment or as a paper and pencil test, it is necessary to talk about the invariance in terms of the measurement method (Whitaker & McKinney, 2007). Similarly, when a factor structure remains the same in subgroups constructed on the basis of gender (male-female), country or socioeconomic level for different groups, it may indicate that measurement invariance between groups is ensured (Kline, 2005). In tests, when measurement invariance is achieved, it is possible to base the difference obtained in terms of the measured property on individual characteristics (Başusta & Gelbal, 2015).

Measurement invariance studies are generally carried out within the scope of SEM with multi-group confirmatory factor analysis (MGCFA) or item response theory (IRT) approaches (Chung et al., 2016). In fact, confirmatory factor analysis (CFA) is recommended instead of IRT in order to examine measurement invariance in measurement tools consisting of ordinal items (Stark et al., 2006). CFA is basically a factor analysis and one of the aims of factor analysis is to reveal the validity of the scores. Validity responds to the question of whether the measuring tool provides a score for the desired dimension. At the same time, it questions whether the items in one dimension really and only measure this dimension (Thompson, 2004). When answers are found to these questions, factor analysis can also be the evidence of content validity, construct validity, and even face validity. For this reason, factor analysis is considered to be the heart of psychological constructs (Nunnally, 1978).

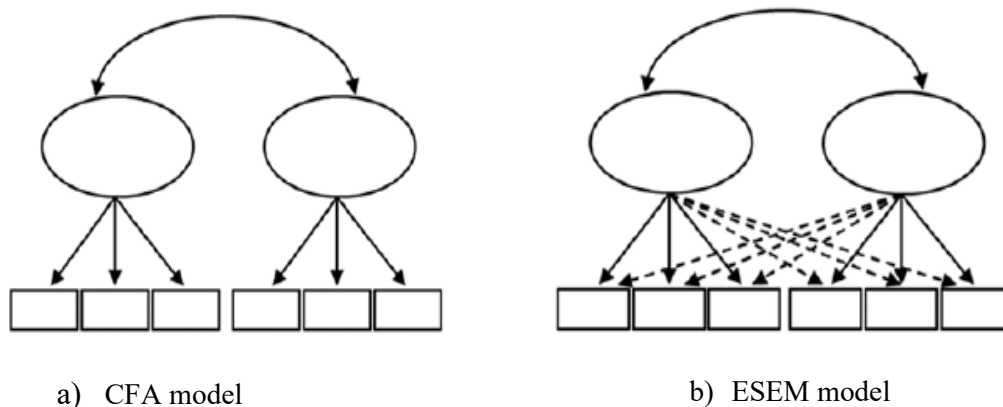
Factor analysis is a method originally developed to explain the characteristics called unobserved (latent) variables or factors underlying a performance related to observed variables. These factors cannot be observed directly, but instead the observed variables are considered to be indicators of latent constructs. Statistically, the purpose of factor analysis is to reveal the maximum variance by identifying new constructs (factors and dimensions) that may occur in fewer numbers through the relationships between observed variables (Brown, 2006; Büyüköztürk, 2002; Özdamar, 2004; Reykov & Marcoulides, 2008). Thus, the fact that factors are interpretable to a large extent on the basis of data and for which reasons observed variables have high levels of correlations with each other can be explained (Reykov & Marcoulides, 2008).

The most used types of factor analysis are Exploratory Factor Analysis (EFA; Jennrich & Sampson, 1966) and Confirmatory Factor Analysis (CFA; Joreskog, 1969) (Özdamar, 2004; Stapleton, 1997). EFA and CFA are basically similar in terms of explaining the observed relationships between indicators with fewer latent variables. However, they are different techniques in terms of the number and nature of the priori features and limitations of the model (Brown, 2006). EFA is thought of as a precursor of CFA used to describe and distinguish basic psychological constructs (Cudeck & MacCallum, 2007). In EFA, the researcher is not expected to determine a construct about the data in advance. Although the researcher has some expectations,

EFA may not provide a suitable model for these expectations, and the analysis process should not be affected by the expectations of the researcher (Brown, 2006; Thompson, 2004). EFA is used as an exploratory and descriptive technique in terms of which observed variables are reasonable indicators of latent dimensions. CFA does not aim to discover or reveal factors, but it is used to verify, test or quantify a hypothetical construct predicted among a set of measurements. In the area of interest of CFA lies the examination of the model for the relationship of factors with each other and with observed variables (Reykov & Marcoulides, 2008).

CFA is a type of structural equation modelling and deals with measurement models. CFA tries to model the relationships between test items, test scores, or the observation levels of behaviour called indicators or observed variables and latent variables or factors (Brown, 2006). In fact, CFA is hypothesis-based in nature. CFA tests a theory, while EFA generates a theory (Brown, 2006; Stapleton, 1997). Cross-loadings allowed between variables by EFA are forced to be zero in CFA. There is a limitation in CFA that argues that the scores are only related to the relevant factor; however, the imposition that a latent variable is only related to the target item and not to other dimensions gives rise to some difficulties for empirical validation. In reality, such a pure relationship is unlikely. Although a factor plays a dominant role in explaining the target observed variable, it is very difficult to say that this item is explained only by the relevant factor. The fact that there is such a restrictive requirement in CFA may cause inflated factor correlations and thus the structural relationships between factors to be damaged (Asparouhov & Muthen, 2009; Marsch et al., 2009). Therefore, it can be seen that a model that is thought to be well defined with EFA is not confirmed by CFA (Guo et al., 2019). In CFA, the suitability of the models is evaluated by looking at the model fit indices. When the indices are not within the acceptable limits, this often causes the absence of confirmation or doubts about the reliability and reproducibility of the models (Asparouhov & Muthen, 2009; Booth & Hughes, 2014). The Explanatory Structural Equation Model (ESEM) is a technique developed to overcome this limitation. Its most important advantage is that it allows different latent variables to cross-load with different items (Asparouhov & Muthen, 2009; Gomes et al., 2017). When the structure of ESEM model is examined in Figure 1 (b), it is seen that ESEM brings together the advantages of EFA and CFA (Marsh et al., 2014).

Figure 1. CFA and ESEM representations for a two-factor model (Booth & Hughes, 2014).



ESEM, like EFA, is flexible when testing measurement models and includes possible rotations for the construct matrix. It tests measurement models of latent variables using EFA instead of CFA. It is a more flexible type of structural equation model that models using an explanatory approach considering which of the factor rotations are appropriate (Schmitt, 2011). In CFA, all parameters are defined a priori by the researcher. It also presents different hypotheses for the relationships between observed and latent constructs. ESEM, on the other hand, requires only the number of factors as a priori knowledge and freely estimates all the other parameters (Booth

& Hughes, 2014). ESEM uses EFA while creating the measurement model that is part of a factor model and calculating the variance of residuals, rotated factor loadings, factor variances, and covariances (for example, regressions of latent factors on independent variables). Unlike CFA, an item does not necessarily load on a single factor. Small but statistically significant cross-loadings are not forced to be zero in the analysis with ESEM.

Existing research shows that ESEM model provides model data fit better than CFA in personality scales (Kristjansson et al., 2011). In addition, some studies show that better model fit coefficients are obtained compared to CFA (Rosellini & Brown, 2011; Mattson, 2012). ESEM models are also expanded to multi-group analysis, allowing factorial structures to be compared in terms of measurement invariance or differential item functioning (Marsch et al., 2009; Tomás, et al., 2014). ESEM is a preferred model to reveal the factor structure when it exhibits better model data fit than CFA (Marsh, et al., 2014). In a study comparing CFA and ESEM in terms of model-data fit, it was stated that the data in the subgroups of culture, socioeconomic level and social capital fit better with ESEM model in the PIRLS 2006 and PISA 2009 applications (Caro et al., 2014). In another study where the factor structure of the Academic Motivation Scale was compared with CFA and ESEM, it was stated that the ESEM approach fit better with the data and the pattern of factor correlations was ranked more appropriately to the theoretical framework (Guay et al., 2015). Joshanloo and Lamers (2016), in their study examining the construct of well-being with CFA and ESEM, revealed that the factor structure in CFA is not clear enough, and that the two dimensions are very well separated with ESEM. Marsh et al. (2011), on the other hand, evaluated the 11-factor construct addressed under two main headings called cognitive and affective related to the academic motivation and responsibility scale. They stated that although the number and pattern of factors obtained with CFA and ESEM were the same, better fit indices were obtained with the ESEM model.

In the literature, it is stated that testing the factor structure of a measurement tool with ESEM is more advantageous than CFA. On the other hand, Marsch et al. (2009) stated that when ESEM and CFA are tested together and the model fit is acceptable for both, it is more appropriate to continue the analysis by using CFA. For this reason, in the current study, it is aimed to determine which of ESEM and CFA models is more suitable for revealing the factor structure and construct validity of TIMSS 8th grade mathematics attitude questionnaire. In previous studies examining the TIMSS questionnaire, it is seen that CFA was used for model testing (Bofah & Hannula, 2015; Ertük & Erdiñç-Akan, 2018; Polat, 2019). In the current study, it is thought that the comparability of the results to be obtained with ESEM with previous studies will contribute to revealing the factor structure of the TIMSS mathematics attitude questionnaire. In the current study, it is also aimed to test the measurement invariance of the factor structure of the TIMSS attitude questionnaire depending on gender with the accepted model. Consistent with the TIMSS 2015, 2011 and 2007 results, it was stated that mathematics achievements of the students who have interest in mathematics, who value mathematics, who have self-confidence in mathematics and who love mathematics in the Turkish sample are high (Raport, 2015). In some studies, it has been seen that the gender factor is important in terms of beliefs about mathematics (Bofah & Hannula, 2015; Simpkins et al., 2005). Watt (2004) stated that girls' interest in mathematics is higher than that of boys, but that there is no difference between their mathematic performances. According to him, gender is more important in affecting students' self. The current study focuses on invariance by gender in order to be a reference for future research and to make more reliable interpretations on the relationship between the mathematics attitude questionnaire and achievement. In this connection, in the current study, answers to the following questions were sought: "Do the Turkish data obtained from the TIMSS 2015 8th grade mathematics attitude questionnaire fit the ESEM and CFA models? Does the questionnaire achieve measurement invariance in female and male student groups?"

2. METHOD

In this study, factor structure of the TIMSS 8th grade mathematics attitude questionnaire and its measurement invariance by gender are examined. The current study is a descriptive research aimed at revealing the existing state.

2.1. Study Group

A total of 238 schools having 8th graders participated in the TIMSS 2015 study from Turkey. In the period when the TIMSS 2015 study was conducted, there was a total of 1,187,893 eighth grade students. The number of students included in the TIMSS sample was 6079. In the Turkish sample, 48.4% (2943) were females and 51.6% (3136) were males. The current study was conducted by using all the data obtained from 6079 students.

2.2. Data Collection Tool

In the TIMSS application, science and mathematics achievement tests and student, teacher, school, and parent questionnaires were used to determine the knowledge and skill levels of the 8th grade students. In the current study, an attitude questionnaire applied to the 8th grade students was used. The questionnaire was administered under four titles “Students’ Interest in Mathematics”, “Students’ Views on Engaging in Mathematics Lessons”, “Students’ Self-Confidence in Mathematics” and “Students’ Value Mathematics”. There is a total of 37 items in the questionnaire including 9 items (2 are negative) to measure students interest in mathematics, 10 items to measure their views on engaging teaching in mathematics lessons, 9 items for their self-confidence in mathematics (5 are negative), and 9 items to measure their value of mathematics. The scale items are in the form of 4-point Likert scale with the following response options: ‘1’ strongly disagree, ‘2’ a little disagree, ‘3’ a little agree, and ‘4’ strongly agree.

2.3. Data Analysis

Before starting the analyses, missing data and assumptions were examined and the data were made ready for analysis. In the TIMSS 8th grade data, there was missing data less than 1% in each variable. In all the data, the rate of missing data was 1.02%. This rate is lower than 5% (Çokluk et al., 2010; Raykov & Marcoulides, 2008). However, it is necessary to investigate whether the missing data are completely random or random and which technique will be applied to the missing data should be determined. In studies, it is stated that such data should be approached with caution regardless of the method (Allison, 2003; Graham, 2012; Little & Robin, 1987). The distribution of the missing data can be determined with the Little Missing Completely at Random (MCAR) test. In the current study, the MCAR test was found to be significant, meaning that the data were not completely random. In this case, assigning data can be seen as a more reliable method than deleting the data. For this reason, missing data were assigned with Expectation Maximization (EM) (Allison, 2003).

Since CFA and ESEM, two types of the structural equation modelling as multivariate statistical methods, are used in the current study, it is necessary to check whether the data can satisfy the multivariate normality assumption or not. Therefore, the data were evaluated in terms of univariate and multivariate outliers, univariate and multivariate normality, linearity, covariance, and multicollinearity. In order to determine the univariate outliers, the z values of the variables were calculated and those outside the ± 3 range were examined. It was seen that there were not any univariate outliers in the data. For multivariate outliers Mahalanobis distances were calculated and it was found that Mahalanobis values did not exceed the critical chi-square value at $p < 0.001$ ($\chi^2_{36,0.001} = 66.62$). For multicollinearity, the following conditions were taken into consideration: a confidence interval-CI value lower than 30, variance inflation factor (VIF) lower than 10, and the tolerance values greater than .20 (Tabachnick & Fidell, 2007). No mul-

ticollinearity was found in the data. However, based on the Levene test results, it was determined that some items did not provide covariance between groups. On the other hand, since most items had skewness and kurtosis values outside the range ± 1 , it was accepted that the normality assumption was not satisfied in the current study. For this reason, Robust Maximum Likelihood (MLR) estimation method was used in the Mplus 7.0 program.

In order to examine the factor structure of the TIMSS mathematics attitude questionnaire, the models analysed by CFA and ESEM were compared. The TIMSS questionnaire items were presented to the students under four headings. Therefore, in the current study, the questionnaire items were tested as 3-factor, 4-factor and 5-factor ESEM and 4-factor and 5-factor CFA and the results were compared. In order to decide the model having the best fit, Bayesian Information Criterion (BIC) values (Kuha, 2004), differences between adjusted chi-square (*adjusted χ^2*) values (Asparauhov & Muthen, 2006), fit indices, and factor loadings were examined. In addition, the level of correlation between the factors was also taken into account. In order to evaluate model fit, the following goodness-of-fit indices can be used: χ^2 , χ^2 /degree of freedom (χ^2/sd), root-mean-square error of approximation (RMSEA), the comparative fit index (CFI), goodness of fit index (GFI), adjusted goodness fit index (AGFI), the Tucker–Lewis index (TLI), and standard root mean square (SRMR). Although the use of such a variety of indices (especially when they take different values) creates a conflict about the fit of the model with the observed data, it can be decided about model fit by considering some suggested value ranges (Schermelleh-Engel et al., 2003). In addition, some studies indicate that CFI, TLI, and RMSEA indices are independent of the sample size (Hu & Bentler, 1995; Marsch et al., 2005). In the current study, it was decided that the model would be acceptable if the value of RMSEA and SRMR was smaller than 0.05 and CFI value was greater than 0.90, and TLI value was greater than 0.90 (Browne & Cudeck, 1993; Schermelleh-Engel & Moosbrugger, 2003). It was also accepted that the one with a lower BIC value of the two compared models would be the one with a better fit (Krueger et al., 2007). Moreover, it was decided that the model in which the correlations between the factors are smaller than 0.70 would be accepted to have a better fit with the data (Marsh et al., 2011; Guay et al., 2015).

After the selection of the model, the invariance of the attitude questionnaire was examined in the male and female groups. The females were taken as the reference group. By imposing restrictions in parameters in the males, the modal invariance was investigated. At this stage, the limited model and the less limited model were compared in terms of fit. The stages of the measurement invariance were hierarchically investigated as configural, weak, strong, strict, variance-covariance, and latent mean invariance (Guay et al., 2015; Marsch et al., 2010; Meradith, 1993; Morin & Maïano, 2011; Steenkamp & Baumgartner, 1998).

In the configural invariance stage, it is examined whether the factor model is equal for groups or not. In other words, factor loadings, factor means (intercept), and error variances are set free in both groups, latent variances are equated to 1, and latent means are equated to 0 in the reference group. In the weak invariance stage, as different from the previous model, restriction of equality between groups is imposed on factor loadings and cross-loadings. In the strong invariance stage, along with factor loadings, factor means are also restricted and forced to be equal between the groups compared. In the strict invariance stage, restriction of equality between groups is imposed on measuring errors at item level. In fact, the invariances ensured up to this stage prove that the properties of a measurement tool are the same between groups. In the current study, the model continued to be restricted and variance/covariance invariance was examined by restricting the variance/covariance matrix to be equal to 1 in all groups. In the last stage, the latent mean invariance was equated to 0 and latent mean invariance test was conducted (Morin & Maïano, 2011).

In order to find evidence for invariance, the differences between χ^2 , CFI, TLI, and RMSEA values obtained from hierarchical models can be used. Since the MLR was used to estimate the parameters in the current study, Satorra-Bentler χ^2 ($S - B_{\chi^2}$) value was obtained. For this reason, in order to calculate the difference between χ^2 values, it is necessary to calculate TR_d values by making adjustments. When the obtained TR_d value is greater than the critical value at the relevant degree of freedom and 0.05 significance level, the null hypothesis is rejected and it is interpreted that the models are different from each other (Asparouhov & Muthen, 2010; Bryant & Satorra, 2012; Satorra & Bentler, 2010). The formula in Equation 1 is used in calculating the TRd value.

$$c_d = [(d_0 * c_0) - (d_1 * c_1)] / (d_0 - d_1) \tag{1}$$

$$TR_d = [(T_0 * c_0) - (T_1 * c_1)] / c_d$$

d_0 : degree of freedom obtained for the restricted model, c_0 : scaling factor of the restricted model, d_1 : degree of freedom of the compared model, and c_1 : scaling factor of the compared model. T_0 : $S - B_{\chi^2}$ value of the restricted model; T_1 : $S - B_{\chi^2}$ value of the compared model.

It is stated in the literature that CFI and RMSEA values are more reliable than χ^2 because they are not sensitive to the sample size. For this reason, it is appropriate to evaluate other fit indices together with χ^2 in model comparisons. Chen (2007) states that measurement invariance can be achieved when the decrease in the CFI value is .01 or less, or the increase in the RMSEA value is .015 or less. These values are suggested for the Maximum Likelihood (ML) estimation; however, Sas et al. (2014) showed that CFI and RMSEA values gave similar results to ML in MLR method. Therefore, in model comparisons, besides χ^2 test, ΔCFI and $\Delta RMSEA$ values were examined and thus decision was made about the invariance (Guay et al., 2015; Jung, 2019).

3. FINDINGS

3.1. Model Fit of the TIMSS 8th Grade Attitude Questionnaire

The TIMSS 8th grade attitude questionnaire items were presented to the students under four different headings. Therefore, Table 1 presents the goodness-of-fit indices of the models tested as four-factor models and of the models constructed as alternatives.

Table 1. Goodness-of-fit indices obtained for the alternative models and information criteria

	χ^2/sd	CFI	TLI	RMSEA	BIC	SRMR	<i>sf</i>
3-factor ESEM	14119.9/558	0.853	0.825	0.063	515966.04	0.042	1.318
4-factor ESEM	8075.98/524	0.918	0.896	0.049	508445.10	0.027	1.337
5-factor ESEM	5246.81/491	0.948	0.930	0.040	504986.23	0.020	1.343
4-factor CFA	14636.47/623	0.848	0.838	0.061	516245.98	0.072	1.329
5-factor CFA	10046.53/619	0.898	0.890	0.05	510174.40	0.059	1.329

sf: scaling factor

As can be seen in Table 1, the model having the highest goodness-of-fit indices is the 5-factor ESEM model ($\Delta\chi^2 = 4944,33$; $\Delta df = 128, p < .05$; RMSEA = 0.040, CFI =0.948, TLI =0.930, and SRMR = 0.020). At the same time, the lowest BIC value (BIC = 504986.23) was obtained in this model. After this model, the model having the best fit is the 5-factor CFA model (RMSEA = .05, CFI = .898, TLI = .890 and SRMR = .059). As the TR_d value indicating the $S - B_{\chi^2}$ difference between two models was calculated to be 4944.33, it can be argued that

there is a significant difference between the models ($\Delta\chi^2(128) = 155.40, p < .05$). In order to decide on the final situation, factor loadings obtained by the 5-factor ESEM and 5-factor CFA and correlations between factors were compared and the analysis results are given in Table 2 and Table 3.

Table 2. Factor loadings and cross loadings of the mathematics attitude questionnaire obtained with the 5-factor ESEM and 5-factor CFA.

Factors	Item	F1	F2	F3	F4	F5	CFA
Students' interest in mathematics	1	0.721	0.036	-0.017	0.066	0.037	0.805
	2	-0.527	0.010	0.361	0.167	-0.042	-0.561
	3	-0.622	-0.017	0.374	0.133	-0.006	-0.679
	4	0.463	0.111	0.086	0.039	0.084	0.548
	5	0.849	-0.020	-0.046	0.037	0.011	0.880
	6	0.674	0.012	0.079	0.089	0.019	0.713
	7	0.662	-0.030	-0.010	0.215	0.004	0.811
	8	0.700	0.061	-0.009	0.130	-0.026	0.810
	9	0.696	0.011	-0.106	0.177	-0.016	0.868
Students' views on engaging in mathematics les- sons	1	-0.041	0.465	-0.013	0.206	0.043	0.549
	2	0.025	0.724	-0.033	0.030	-0.023	0.741
	3	0.179	0.477	-0.002	0.077	0.090	0.655
	4	0.095	0.337	0.155	0.227	-0.011	0.466
	5	0.009	0.757	-0.034	0.026	0.013	0.779
	6	0.051	0.770	-0.033	-0.041	-0.011	0.773
	7	-0.016	0.705	0.039	0.014	0.003	0.693
	8	0.085	0.649	0.050	-0.007	-0.008	0.682
	9	-0.053	0.746	-0.010	-0.001	0.023	0.719
	10	-0.020	0.752	-0.009	-0.036	0.028	0.732
Students' mathe- matics anxiety	1	0.110	0.016	0.651	-0.195	-0.010	0.685
	2	-0.086	0.005	0.609	-0.208	-0.033	0.766
	3	0.059	-0.088	0.616	0.031	0.020	0.543
	4	-0.045	0.026	0.684	-0.153	0.019	0.785
	5	-0.125	-0.032	0.684	-0.100	0.016	0.800
Students' self- confidence in mathematics	1	0.174	0.043	-0.074	0.605	0.036	0.820
	2	0.103	0.075	-0.020	0.652	0.050	0.798
	3	0.064	-0.023	-0.009	0.752	0.020	0.773
	4	-0.002	0.151	-0.037	0.694	0.003	0.761
Students' value mathematics	1	0.296	0.168	0.058	0.000	0.295	0.587
	2	0.223	0.068	0.114	0.030	0.440	0.630
	3	-0.052	-0.035	-0.023	0.081	0.733	0.695
	4	-0.055	-0.035	-0.026	0.089	0.767	0.733
	5	0.214	-0.064	-0.008	0.303	0.354	0.627
	6	0.140	0.002	-0.006	-0.032	0.714	0.786
	7	0.061	0.022	-0.012	-0.039	0.720	0.740
	8	-0.111	0.100	0.038	0.028	0.521	0.493
	9	0.048	0.126	-0.032	-0.071	0.631	0.679
Cross Loadings		$\overline{ X } = .085$			$\overline{SD} = .109$		

F1: Factor 1, F2: Factor2, F3: Factor 3, F4: Factor4, F5: Factor5

When the factor loadings obtained with ESEM in Table 2 are examined, it is seen that the factor loadings of the items in the first factor called “students’ interest in mathematics” vary between .463 and .849. The cross-loadings of these items on the other factors are close to zero. In other words, these items do not exhibit high loading values in the other factors. The factor loadings of the items in the second factor called “students’ views on engaging in mathematics lessons” were found to be ranging from 0.337 to 0.752. The cross-loadings of these items are close to zero. There are five items in the third factor called “mathematics anxiety” and the factor loadings of these items were found to be ranging from 0.609 to 0.684. These items exhibit factor loadings close to zero in the other factors. According to the results of ESEM, in the fourth dimension called “students’ self-confidence in mathematics”, there are four items and the factor loadings of these items were found to be ranging from 0.605 to 0.752. Finally, in the fifth factor called “students’ value mathematics”, there are 9 items and the factor loadings of the items were found to be ranging from 0.295 to 0.767. However, two items in this factor were found to be exhibiting high factor loadings in the first factor called “students’ interest in mathematics” (main factor loading value is 0.295, cross-loading value is 0.296). The fifth item in the factor called “students’ value mathematics” exhibits a similar loading value in the factor called “students self-confidence in mathematics” and shows a high cross-loading value (main factor loading is 0.354, cross loading value is 0.303). This situation casts doubt on the validity of both of the items. According to Marsch et al. (2011) cross-loadings of items should be as close to zero as possible. The discriminant validity is considered to be poor if the cross-loading of the item moves away from zero and gives a high loading on a factor other than its own (Hair et al., 2010). It may be appropriate to remove such items from the measurement tool. At this point, although ESEM follows a strict path in eliminating weak items, it actually wants to increase the validity of the factors. The mean cross-loading of factor loadings obtained by ESEM as absolute value is .085, while the standard deviation is .109. When the factor loadings obtained with CFA are examined, it is seen that the lowest factor loading is .548, while the highest factor loading is .880. The factor loadings obtained with CFA are generally higher than the factor loadings obtained with ESEM. Correlations between the factors in the model are given in Table 3.

Table 3. Correlations between the factors obtained with ESEM and CFA.

	Factor	Factor 2	Factor 3	Factor 4	Factor 5
ESEM	1	0.472	-0.332	0.660	0.518
	2		-0.043	0.328	0.450
	3			-0.410	-0.097
	4				0.430
CFA	1	0.531	0.810	-0.565	0.640
	2		0.482	-0.174	0.552
	3			-0.611	0.597
	4				-0.262

When the correlation values given in Table 3 are examined, it is seen that the correlation values between the first and fourth factors (-.565), between the second and fourth factors (-.174), and between the fourth and fifth factors (-.262) in the CFA model are lower compared to those of ESEM. The correlation found between the first and third factors in the CFA is .810, while the same correlation was calculated to be -.332 in ESEM. The high correlation between the two factors in the CFA model creates doubts about whether they measure similar features and whether the model has five or four factors. For this reason, after examining factor loadings and correlation values, the ESEM model was accepted, but two items, "mathematics will help me"

and "job involving mathematics", were excluded from the analysis due to their high cross-loadings. The fit indices obtained after the items were removed are shown in Table 4.

Table 4. Fit indices obtained for the five-factor and 35-item TIMSS 8th grade attitude questionnaire with ESEM

	CFI	TLI	RMSEA	BIC	SRMR	sf.
ESEM (35 items)	0.953	0.935	0.039	474776.122	0.019	1.350

When the fit indices given in Table 4 are examined, it can be said that ESEM is consistent with the data (RMSEA = 0.039, CFI = 0.953, TLI = 0.935 and SRMR = 0.019). Therefore, it was accepted that the TIMSS 8th grade mathematics attitude questionnaire conforms to ESEM with 35 items and 5 factors and the measurement invariance of the questionnaire between male and female groups was examined with ESEM. The results obtained are given in Table 5.

3.2. Measurement Invariance in terms of Gender

As can be seen in Table 5, when the fitting of the ESEM model in male and female groups is examined, it can be said that the model is acceptable (for males: RMSEA = 0.097, CFI = 0.959, TLI = 0.943, SRMR = 0.018; for females: RMSEA = 0.043, CFI = 0.947, TLI = 0.926, SRMR = 0.02). Since the model was confirmed separately in each group, the invariance phase was initiated.

Table 5. Fit indices obtained for gender groups and invariance stages with ESEM.

	$S - B_{\chi^2}$	df	sf	CFI	TLI	RMSEA	SRMR	ΔCFI	$\Delta RMSEA$
Males	2241.46	430	1.361	0.959	0.943	0.037	0.018	-	-
Females	2750.71	430	1.304	0.947	0.926	0.043	0.020	-	-
M1	4981.13	860	1.332	0.953	0.935	0.040	0.019	-	-
M2	5418.74	895	1.319	0.948	0.931	0.041	0.025	-0.004	0.001
M3	5610.14	1040	1.317	0.948	0.940	0.038	0.025	0	-0.003
M4	5887.86	1075	1.337	0.945	0.939	0.038	0.028	-0.003	0
M5	6100.86	1090	1.337	0.943	0.937	0.039	0.048	-0.002	0.001
M6	6233.90	1095	1.335	0.941	0.936	0.039	0.046	-0.002	0

df: degree of freedom M1: Configural invariance, M2: Weak invariance, M3: Strong invariance, M4: Strict invariance, M5: Variance/covariance invariance, M6: Latent mean invariance

When the first stage (M1); configural invariance is examined, it can be said that the model is confirmed and the configural invariance is accepted (RMSEA = .040, CFI = .953, TLI = .935, SRMR = .019). It can be said that in the weak invariance stage (M2), fit indices are within the acceptable ranges (RMSEA = .041, CFI = .948, TLI = .931, SRMR = .025). The TR_d value for the chi-square difference between the two models was calculated to be 512.67. Accordingly, it can be said that there is a significant difference between the two models ($\Delta\chi^2(35) = 49.80, p < .05$). However, the difference between fit indices obtained in configural and weak invariance stages is acceptable ($\Delta CFI = -.004, \Delta RMSEA = .001$). This shows that weak invariance is achieved. When M3 (strong invariance) is examined for the third stage, it can be said that fit indices of the model are at an acceptable level (RMSEA = .038, CFI = .948, TLI = .940, SRMR = .025). TR_d value between M3 and M2 models was calculated to be 184.90. Accordingly, there is a significant difference between the two models ($\Delta\chi^2(145) = 174.1, p < .05$). When the change of fit indices is examined in addition to χ^2 difference value, it can be

said that the difference between the models is not significant ($\Delta CFI = 0$, $\Delta RMSEA = -.003$). In this case, it can be concluded that the strong invariance of the TIMSS attitude items is tenable across male and female student groups. In the fourth stage, after the restriction of equality of error variances between the groups is imposed, it can be stated that the fit indices of the strict invariance (M4) model remain within acceptable ranges (RMSEA = .038, CFI = .945, TLI = .939, SRMR = .028). The TR_d value between M4 and M3 was calculated to be 250.36 and a significant difference was found between the models as ($\Delta\chi^2(35) = 49.80$). However, the difference of the fit indices between M4 and M3 provides support for strict invariance across gender ($\Delta CFI = -.003$, $\Delta RMSEA = 0$). When the fit indices obtained for M5 are examined at the next stage, it can be said that the model produces acceptable fit (RMSEA = .039, CFI = .943, TLI = .937, SRMR = .048). TR_d value calculated between M5 and M4 was found to be 213.0. Although, there is a significant difference between the two models ($\Delta\chi^2(15) = 24.99$), the change of fit indices are within the acceptable range ($\Delta CFI = -.002$, $\Delta RMSEA = .001$). The fit indices belonging to M6 model constructed to test whether latent means are invariant or not can be said to be in acceptable ranges (RMSEA = .039, CFI = .941, TLI = .936, SRMR = .046). When M6 and M5 are compared, it can be said that the change in M6 is not significant compared to M5. Thus, latent mean invariance is achieved: ($\Delta CFI = -.002$, $\Delta RMSEA = 0$).

4. DISCUSSION and CONCLUSION

In the current study, the fit of TIMSS 8th grade mathematics attitude questionnaire to the ESEM and CFA models was examined in the first stage and it was seen that the data fit the 5-factor ESEM model the best. ESEM is a model that allows cross-loadings. In CFA, an item's factor loading (even very small values) is forced to be zero in factors other than its own factor. This may cause high correlation values between factors and poor model fit (Marsh et al., 2009). Stromeier et al. (2015) stated that cross-loadings allow for better model estimation based on such indicative information rather than causing pollution to the structure. As a matter of fact, in the current study, the 5-factor ESEM model exhibited fit better than the 5-factor CFA and 4-factor CFA models. In addition, the correlation between the 2nd and 3rd factors in the CFA model was found to be high as 0.81. This might be because of the restriction on cross-loadings. Failure to detect very small cross-loadings by CFA may cause bias in correlation values between factors (Jung, 2019). On the other hand, in the 5-factor ESEM model, the cross-loading values of 2 items move slightly away from 0 and give rise to doubts on the discriminant validity of these items. After these two items were removed from the analysis, it was concluded that ESEM was compatible with the data. Polat (2019) stated that the TIMSS 8th grade mathematics affective feature model fits CFA with 4 factors and 34 items. However, in order to be able to decide this in his study, he first applied EFA to the data and decided to exclude 3 items from the analysis according to EFA results. He stated that the model was confirmed with 4 factors and 34 items by applying CFA after the items were removed. In the model he established, he reported the correlation between the dimensions of "students like learning mathematics" and "students are confident in mathematics" as .85. Although the removal of items contributes to the model, the level of correlation between dimensions remains high. This situation can be considered as an indication that the factor loadings are affected by the cross-loadings being forced to zero. Guay et al. (2015) stated that the correlations between factors in the motivation scale are lower than they are in CFA. In the TIMSS 8th grade science attitude questionnaire, which Jung (2019) examined as 3 factors, the correlation values between the dimensions of 'students' like learning science', 'students' confidence in science' and 'students' value science' remained at low levels compared to CFA (the highest correlation with ESEM is .762; the highest correlation is .823 with CFA). The correlation results obtained in the current study are consistent with the related studies in the literature. On the other hand, when cross-loadings were examined with the ESEM model, it was concluded that the cross-loadings of the items "mathematics will help me" and

"job involving mathematics" in the factor called students value mathematics were high. Therefore, these two items were excluded from the analysis. The item "mathematics will help me" was excluded from the analysis because it was in a different dimension according to EFA results in the study conducted by Polat (2019). In addition, Polat (2019) concluded that two items in the dimension of "self-confidence in mathematics" should be removed from the analysis as a result of EFA. In the current study conducted with ESEM, the dimension of "students' self-confidence in mathematics" was divided into two different dimensions according to ESEM results and was named as "students' self-confidence in mathematics" and "students' mathematics anxiety". For this reason, different from the study of Polat (2019), there was no need for item exclusion.

In the current study, since it was decided that ESEM was the more suitable model, measurement invariance was carried out with ESEM. The TIMSS 8th grade mathematics attitude questionnaire achieved configural, weak, strong, strict, variance/covariance, and latent mean invariance. According to the results obtained, it can be said that the factor structure, factor loadings, factor means, errors, variance/covariance matrix, and latent means of the questionnaire are similar in male and female groups, so it can be used safely in studies related to gender (Guay et al., 2015). TIMSS questionnaires have been the subject of many measurement invariance studies. Polat (2019) examined the TIMSS 8th grade mathematics affective model in the Turkish data as 4 factors and 34 items with MGCFA. He stated that all the invariance stages were satisfied between the genders in the study. Jung (2019) examined the invariance of the TIMSS 2015 8th grade science attitude questionnaire between genders with three-factor ESEM in American data. In the study, configural, weak, and strong invariance was examined and it was stated that these three stages were satisfied. Ertürk and Erdinç-Akan (2018) examined three of the affective characteristics affecting mathematics separately in the TIMSS 2015 4th grade Turkish data; namely, "like learning mathematics", "interest in mathematics" and "self-confidence in mathematics", and the invariance of each variable depending on gender. As a result of the study, strict invariance condition was met for only the variable of "like learning mathematics". It was observed that the variables of "interest in mathematics" and "self-confidence in mathematics" achieved configural invariance. Bofah and Hannula (2015) stated that the TIMSS 2011 scale, which consisted of items such as "like learning mathematics", "value mathematics", "self-confidence in mathematics", "teacher responses", and "parent participation" achieved configural, weak and strong invariance in ten different countries in male and female groups. Marsh et al. (2013) examined the factor structure of the TIMSS 2007 mathematics and science motivation scale in the Arab countries of Saudi Arabia, Oman, Egypt, and Jordan, England, Scotland, Australia, and the U.S. with CFA. They also examined the invariance of the questionnaire between genders for each country and stated that the scale fulfils the requirement of complete invariance. On the other hand, it was stated that in the Turkish data in PISA 2012, another international exam, the learning model fulfilled all the conditions of invariance (Kıbrıslıoğlu, 2015). Unlike these results, according to PISA 2015 Turkish data, Güngör and Atalay Kabasakal (2019) stated that the science motivation and self-efficacy model and Uyar and Kaya-Uyanık (2019) stated that the science learning model did not achieve invariance in relation to gender. In the studies conducted, it is stated that although it is not clear whether the questionnaires used in international studies achieve the invariance in terms of gender, the measurement invariance for gender in mathematics questionnaires has been achieved in general. It is seen that the results obtained in the current study are consistent with the studies in the related literature.

Failure to achieve measurement invariance for a measurement tool may prevent a healthy comparison of scores. For this reason, researchers are recommended to examine the invariance of the measurement tools used between the groups they will compare. In addition, it may be suggested that they take into consideration the advantages of ESEM and use ESEM in studies where EFA and CFA will be used together. In the current study, invariance only in terms of

gender was examined. In future studies, invariance in sub-groups such as country, socio-economic level or geographical region can be examined with the ESEM model. This study focused solely on the questionnaire applied to 8th grade students. Since TIMSS is an evaluation study that is also applied to 4th graders, it may be suggested to examine the invariance of questionnaires related to mathematics and science among younger students with ESEM. Researchers need to apply EFA followed by CFA in scale development or adaptation studies. ESEM can be recommended to researchers as it can provide information about the model at once instead of analyzing it twice.

Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

ORCID

Seyma UYAR  <https://orcid.org/0000-0002-8315-2637>

5. REFERENCES

- Asparouhov, T., & Muthén, B. (2006). Robust chi square difference testing with mean and variance adjusted test statistics. *Mplus Web Notes*, 10.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438. <https://doi.org/10.1080/10705510903008204>
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 545. <https://doi.org/10.1037/0021-843X.112.4.545>
- Asparouhov, T; Muthén, B. (2010). Computing the strictly positive Satorra-Bentler chi-square test in Mplus. *Mplus Web Notes*, 12, 1-12.
- Başusta, N. B., & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği [Examination of measurement invariance at groups' comparisons: A study on PISA student questionnaire]. *Hacettepe University Journal of Education*, 30(4), 80-90. <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/1773-published.pdf>
- Bofah, E. A. T., & Hannula, M. S. (2015). TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties. *Large-Scale Assessments in Education*, 3(1), 1-36. <http://dx.doi.org/10.1186/s40536-015-0014-y>
- Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment*, 21(3), 260-271. <https://doi.org/10.1177/1073191114528029>
- Bornstein, M. H. (1995). Form and function: Implications for studies of culture and human development. *Culture & Psychology*, 1(1), 123-137. <https://doi.org/10.1177/1354067X9511009>
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage.
- Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 372-398. <https://doi.org/10.1080/10705511.2012.687671>
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı [Factor analysis: Basic concepts and using to development scale]. *Educational Administration in Theory and Practice*, 8(4), 470-483. <https://dergipark.org.tr/tr/pub/kuey/issue/10365/126871>
- Caro, D. H., Sandoval-Hernández, A., & Lüdtke, O. (2014). Cultural, social, and economic

- capital constructs in international assessments: An evaluation using exploratory structural equation modeling. *School Effectiveness and School Improvement*, 25(3), 433-450. <https://doi.org/10.1080/09243453.2013.812568>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Chung, H., Kim, J., Park, R., Bamer, A. M., Bocell, F. D., & Amtmann, D. (2016). Testing the measurement invariance of the University of Washington Self-Efficacy Scale short form across four diagnostic subgroups. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 25(10), 2559–2564. <https://doi.org/10.1007/s11136-016-1300-z>
- Cudeck, R., & MacCallum, R. C. (Eds.). (2007). *Factor analysis at 100: Historical developments and future directions*. Lawrence Erlbaum
- Çokluk, Ö., Şekercioglu, G. & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Pegem Akademi.
- Ertürk, Z., & Erdiñ-Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı değişkenlerin cinsiyete göre ölçme değişmezliğinin incelenmesi [The investigation of measurement invariance of the variables related to TIMSS 2015 mathematics achievement in terms of gender]. *Journal of Theoretical Educational Science*, 204-226. <https://dergipark.org.tr/pub/akukeg/issue/40520/412604>
- Gomes, C., Almeida, L., & Nunez, J. (2017). Rationale and Applicability of Exploratory Structural Equation Modeling (ESEM) in psychoeducational contexts. *Psicothema*, 29(3), 396-401.
- Graham, J. W. (2012). *Missing data: Analysis and design*. Springer Science & Business Media.
- Guay, F., Morin, A., Litalien, D., Valois, P., & Vallerand, R. (2015). Application of Exploratory Structural Equation Modeling to Evaluate the Academic Motivation Scale. *The Journal of Experimental Education*, 83(1), 51-82. <https://doi.org/10.1080/00220973.2013.876231>
- Guo, J. M. (2019). A Systematic evaluation and comparison between exploratory structural equation modeling and bayesian structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 529-556. <https://doi.org/10.1080/10705511.2018.1554999>
- Guo, J., Parker, H., Dicke, P., Lüdtke, T., & Diallo, T. (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. <https://doi.org/10.1080/10705511.2018.1554999>
- Güngör, M & Atalay Kabasakal, K. (2020). Investigation of measurement invariance of science motivation and self-efficacy model: PISA 2015 Turkey sample. *International Journal of Assessment Tools in Education*, 7(2), 207-222. <https://doi.org/10.21449/ijate.730481>
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117-144. <https://doi.org/10.1080/03610739208253916>
- Hu, L., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Sage.
- Jennrich, R. I. and Sampson, P. F. (1966). Rotation to simple loadings. *Psychometrika*, 31(3), 313–323. <https://link.springer.com/article/10.1007/BF02289465>
- Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Joshanloo, M., & Lamers, S. M. (2016). Reinvestigation of the factor structure of the MHC-SF in the Netherlands: Contributions of exploratory structural equation modeling. *Personality and Individual Differences*, 97, 8-12. <https://doi.org/10.1016/j.paid.2016.02>

089

- Jung, J. Y. (2019): *A Comparison of CFA and ESEM approaches using TIMSS science attitudes items: evidence from factor structure and measurement invariance*. [Master's Thesis, Purdue University]. Purdue University Graduate School, Department of Educational Studies, <https://doi.org/10.25394/PGS.7995890.v1>
- Kıbrıslıoğlu, N. (2015). *The investigation of measurement invariance PISA 2012 mathematics learning model according to culture and gender: Turkey-China (Shanghai)-Indonesia* [Master's Thesis]. Hacettepe University.
- Kline, R. B. (2005). *Methodology in the social sciences. Principles and practice of structural equation modeling* (2nd ed.). Guilford Press.
- Kristjansson, S. D., Pergadia, M. L., Agrawal, A., Lessov-Schlaggar, C. N., McCarthy, D. M., Piasecki, T. M. & Heath, A. C. (2011). Smoking outcome expectancies in young adult female smokers: Individual differences and associations with nicotine dependence in a genetically informative sample. *Drug and Alcohol Dependence*, 116, 37-44. <https://doi.org/10.1016/j.drugalcdep.2010.11.017>
- Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, 116(4), 645. <https://doi.org/10.1037/0021-843X.116.4.645>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford press.
- Little, R. J., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factorial, convergent, and discriminant validity of timss math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108–128. <https://doi.org/10.1037/a0029907>
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of Fit in Structural Equation Models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Multivariate applications book series. Contemporary psychometrics: A festschrift for Roderick P. McDonald* (p. 275–340). Lawrence Erlbaum Associates Publishers.
- Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J., & 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. <https://doi.org/10.1177/0734282911406657>
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471. <https://doi.org/10.1037/a0019227>
- Marsh, H. W., Morin, A. J., 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, 85-110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (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. <https://doi.org/10.1080/10705510903008220>
- Mattsson, M. (2012). Investigating the factorial invariance of the 28-item DBQ across genders and age groups: an exploratory structural equation modeling study. *Accident Analysis &*

- Prevention*, 48, 379-396. <https://doi.org/10.1016/j.aap.2012.02.009>
- Ministry of National Education (2016). TIMSS 2015 ulusal matematik ve fen ön raporu: 4. ve 8. sınıflar [TIMSS 2015 national mathematics and sciences preliminary report 4th and 8th grades]. https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_06/23161945_timss_2015_on_raporu.pdf
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://link.springer.com/article/10.1007/BF02294825>
- Morin, A. J. S., & Maïano, 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, 540–554. <https://doi.org/10.1016/j.psychsport.2011.04.003>
- Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill
- Özdamar, K. (2004). *Paket Programlar ile istatistiksel veri analizi* (Çok değişkenli analizler). Kaan Kitabevi.
- Polat, M. (2019). *The investigation of measurement invariance of TIMSS-2015 mathematics and science affective characteristics models according to culture, gender and statistical region* [Master's Thesis], Hacettepe University.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rencher, A. (2002). *Methods of multivariate analysis*. John Wiley & Sons, Inc.
- Raykov, T., & Marcoulides, G. (2008). *Introduction to applied multivariate analysis*. Routledge Taylor & Francis Group.
- Rosellini, A. J., & Brown, T. A. (2011). The NEO Five-Factor Inventory: Latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment*, 18(1), 27-38. <http://dx.doi.org/10.1177/1073191110382848>
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data with a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, 21, 167-180. <https://doi.org/10.1080/10705511.2014.882658>
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75(2), 243–248. <http://dx.doi.org/10.1007/s11336-009-9135-y>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23-74. <https://psycnet.apa.org/record/2003-08119-003>
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304-321. <https://doi.org/10.1177/0734282911406653>
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2005). Parents' socializing behavior and children's participation in math, science, and computer out-of-school activities. *Applied Developmental Science*, 9(1), 14-30. https://doi.org/10.1207/s1532480xads0901_3
- Stark, S., Chernshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292–1306.
- Stapleton, C. (1997). *Basic concepts and procedures of confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association January 23-25. Austin.
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90.

- Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*, 41(2), 491-520. <https://doi.org/10.1177/0149206314551962>
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481-498). Pearson.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
- Tomás, I., Marsh, H. W., González-Romá, V., Valls, V., & Nagengast, B. (2014). Testing measurement invariance across Spanish and English versions of the Physical Self-Description Questionnaire: An application of exploratory structural equation modeling. *Journal of Sport and Exercise Psychology*, 36(2), 179-188. <https://doi.org/10.1123/jsep.2013-0070>
- Uyar, Ş, Kaya Uyanık, G. (2019). Fen bilimlerine yönelik öğrenme modelinin ölçme değişmezliğinin incelenmesi: PISA 2015 örneği [Investigating measurement invariance of learning model towards science: PISA 2015 sample] *Kastamonu Education Journal*, 27(2), 497-507. <https://doi.org/10.24106/kefdergi.2570>
- Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade Australian students. *Child Development*, 75(5), 1556-1574.
- Whitaker, B. G., & McKinney, J. L. (2007). Assessing the measurement invariance of latent job satisfaction ratings across survey administration modes for respondent subgroups: A MIMIC modelling approach. *Behavior Research Methods*, 39(3), 502-509. <https://doi.org/10.3758/BF03193019>