Investigation of Measurement Invariance According to Home Resources: TIMSS 2015 Mathematical Affective Characteristics Questionnaire

Derya Cakici Eser [iD][1,*]

[1]Ankara Music and Fine Arts University, Faculty of Music and Fine Arts Education, Department of Educational Sciences, Ankara, Turkey

Abstract: This study aimed to examine the measurement invariance of the mathematical affective characteristics model obtained from TIMSS 2015 4th grade Turkey administration according to home resources. For this purpose, firstly, the factor structure of the mathematical affective characteristics questionnaire was examined by explanatory factor analysis and Velicer's maximum average partial (MAP) test. It was revealed that the questionnaire had three factors. Then the structure was validated by confirmatory factor analysis. In the next stage, multi-group confirmatory factor analysis was employed with a purpose to examine whether the model displayed measurement invariance across the variables of home resources such as internet connection, heating system, cooling system, and dishwasher. The results showed that the strict measurement invariance of the mathematical affective characteristics model was achieved among the subgroups of each of the internet connection, heating system, cooling system, and dishwasher variables. Accordingly, means, variance, covariances, and item residual variances in the subgroups were found to be similar. According to the results of the study, the comparison of the mathematical affective characteristics model based on the home resources is found to be significant and comparisons made show that possible differences arise from the relevant home resource.

1. INTRODUCTION

Exams provide various information to education stakeholders depending on the purpose of exams being administered at all levels of education. Accordingly, based on the information obtained from the exams, information is acquired on such points as the current situation of students in terms of their relevant characteristics, their need for support, the efficiency of the education programs pursued, and whether the educational methods used meet the needs. In addition, curriculum improvements in education are determined with the comparisons made based on the exam results.

Since the late 1990s, education systems of countries have been compared while student achievements through exams have aimed at specific areas, targeting at specific audience with the participation of many countries. Program for International Student Assessment (PISA),

*CONTACT: Derya CAKICI ESER ✉ deryacakicieser@gmail.com ⌨ Ankara Music and Fine Arts University, Faculty of Music and Fine Arts Education, Department of Educational Sciences, Ankara, Turkey

Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) are the leading exams to such ends. PISA is a program administered by the OECD every three years and it focuses on 15-year-old students' reading skills, mathematics literacy, and science literacy. TIMSS and PIRLS are programs run by International Association for the Evaluation of Educational Achievement (IEA). PIRLS is an exam that has been held every 5 years since 2001 to measure the reading skills of 4th grade students. TIMSS is a student achievement research program administered every 4 years for 4th and 8th grade students. TIMSS makes it possible to determine the academic success of students, to direct the change over time, to compare one country's situation with those of other countries, and to monitor the results of the attempts in order to increase the level of success with questions prepared in the field of mathematics and science skills (Ministry of National Education [MNE], 2016).

In TIMSS, the education system is managed together with all its stakeholders and components. Accordingly, in practice, there are questionnaires for the home environment thought to have an effect on the upbringing of children and on their success in school, in the school environment as the determinant of efficiency in achieving educational goals, and in the classroom environment where most of the learning and teaching take place. Besides, since many studies in the literature reveal the relationship between student achievement and student attitude, IEA also makes use of questionnaires to determine attitudes towards mathematics and science in TIMSS administration (IEA, 2019).

Each measurement tool is basically developed with the assumption that it measures the same feature in every group in which it is administered. However, in practice, the results might differ depending on the groups they are administered. Accordingly, results may not have equal/equivalent psychometric qualities and therefore, it would be inaccurate to generalize the results for groups (Başusta & Gelbal, 2015). For these reasons, measurement tools administered in different groups should measure the same construct in each subgroup. If it is shown that the factor loadings, inter-dimensions correlations, and error variances of a measurement model are the same in each group, it indicates that the measurement tool has the same structure in different groups (Jöreskog & Sörbom, 1993). In this context, measurement invariance can determine whether a measurement tool measures in the same way in different groups or not. With measurement invariance studies, researchers obtain evidence about whether or not scales measure the same construct in subgroups (Cheung & Lau, 2012; Millsap & Olivera-Ogilar, 2012; Cheung & Rensvold, 2002). Accordingly, it is stated that measurement tools that are not invariant across groups measure different characteristics in subgroups after the measurement invariance study. This is a validity problem for the measurement tool, and after such a measurement process, mathematical relations between the measurement tool variables will be different in each group. Interpretations regarding the results of group comparisons based on such a measurement tool would also be incorrect (Vandenberg & Lance, 2000). On the other hand, if it is shown that a measurement tool is invariant across groups, in other words, if it is shown that the mathematical relations between its variables are equal between groups, two types of validity proofs are obtained based on the measurement tool. These are (1) proof of construct validity in terms of showing that the measurement tools are used to measure the same structure in each group, (2) proof of the external validity in terms of statistically proving that the results of the comparison across groups can be generalized. In this respect, considering the vital importance of obtaining measurement invariance in interpreting the findings of a study, group comparisons made without demonstrating measurement invariance should be approached with suspicion. Hence, along with the definition of measurement invariance, its theoretical foundations, and how to test it need to be briefly explained.

1.1. Measurement Invariance

Measurement invariance is whether the measurement tool employed corresponds to the same meaning in individuals in different groups. The fact that individuals in different populations but in the same condition in terms of measured constructs get the same observed score in a test means that the measurement is invariant. If the individuals are identical in terms of the measured construct but their scores differ, the test violates the assumption of measurement invariance (Schmitt & Kuljanin, 2008). If measurement invariance is not demonstrated, the results of the comparisons across groups cannot be interpreted with certainty. It cannot be known whether the resulting difference can be attributed to a real attitudinal difference or to the difference in psychometric responses to scale items. Although this point is not instantly obvious, it is a very critical point (Cheung & Rensvold, 2002; Horn & McArdle, 1992). For these reasons, it is important to examine measurement invariance before comparing measurements obtained from two or more groups.

Confirmatory factor analysis (CFA) is one of the methods used to test measurement invariance (Kline, 2011; Schmitt & Kuljanin, 2008; Jöreskog & Sörbom, 1993). Under structural equation modelling, measurement invariance is tested using a series of tests through multi-group confirmatory factor analysis (MGCFA). By using MG-CFA in different ways with various constraints, measurement invariance is tested in four stages in a hierarchical manner (Vandenberg & Lance, 2000, Meredith, 1993).

Configural invariance comes first in the hierarchical order of measurement invariance. Configural invariance is that the construct in the measurement tool is the same across groups. If configural invariance is achieved, it can be concluded that the items in the measurement tool measure the same construct in the groups in which the invariance is investigated (Vandenberg & Lance, 2000). Configural invariance is also called as baseline model. This model reveals that the number of factors in each group and the variables that make up the factors are the same (Millsap & Olivera-Ogilar, 2012). If configural invariance is not achieved, measurement invariance will not be ensured at other stages (Kline, 2011).

When it is shown that configural measurement invariance is achieved, the metric invariance test can be conducted (Cheung & Rensvold, 2002). Metric invariance is also known as weak invariance (Meredith, 1993) or pattern invariance (Millsap, 2011). In this phase of invariance, the answer to the question of "Do common factors mean the same in all groups?" is sought (Gregorich, 2006). In the metric invariance analysis, the invariance of the factor structure of the model and the factor loadings of the items in the model in different groups are tested. While the factor variance of all groups is fixed to one in the configural model, the factor variance of the group selected as a reference in the metric model is fixed to one, and the factor variance restriction of other groups is removed (Millsap & Olivera-Ogilar, 2012). If metric invariance is achieved, the results obtained by quantitative group compares of factor variances and covariances are defensible (Gregorich, 2006).

Once metric invariance is achieved, the scalar invariance test follows it. Scalar invariance test consists of a combination of metric invariance and item intercepts invariance (Millsap & Olivera-Ogilar, 2012). Meredith (1993) called scalar invariance strong factorial invariance. At this stage of invariance, an answer to the question "Is it reasonable to compare group means? is sought (Gregorich, 2006). In scalar invariance, the factor means of the reference group are set to zero. The means of the other groups are not constrained (Millsap & Olivera-Ogilar, 2012). If it is proven that factor loadings and item interceptions are invariant in groups, in other words, if scalar invariance is achieved, group differences estimated based on factor means are neutral. Also, group differences between observed scores are directly related to factor means (Gregorich, 2006).

Scalar invariance is followed by testing strict invariance or strict factorial invariance (Meredith, 1993). At this stage, the aim is to prove the invariance of item residual variances in addition to those whose invariance was proven in previous stages (Gregorich, 2006). Only factor means and factor covariance matrices are released in the analysis (Millsap & Olivera-Ogilar, 2012). By demonstrating that of strict invariance, which is a difficult invariance stage in practice, measurement invariance is fully ensured.

Comparisons across groups based on large-scale exams will only be reasonable when all four stages of measurement invariance given above are met. TIMSS is an exam that makes sure to obtain a very large data set and also to enable longitudinal evaluations, since it is conducted at two levels of education (4th and 8th grades) and repeated every four years. In order to show that the obtained findings are unbiased and accurate, research studies are needed to ensure the measurement invariance across the groups. In this particular study, the aim was to examine the measurement invariance of the mathematical affective characteristics questionnaire in the TIMSS 2015 4th grade assessment according to home resources.

Home resources have been defined as one of the indicators of socio-economic status (SES), pointing to facilities such as books, computers, study rooms, and educational resources (Sirin, 2005). SES refers to the position of an individual or a family in a hierarchy according to access to welfare, power, and social status (Gustafsson, Nilsen & Hansen, 2018). Parental income, parental education, parental occupation, and home resources are four indicators of SES (Sirin, 2005). White (1982) analysed approximately 200 studies investigating the relationship between SES and academic achievement in his meta-synthesis study. In his study White (1982) reported that the relationship between SES and academic achievement points to a weak relationship (r=0.22) contrary to expectations. Sirin (2005) replicated the study of White (1982) about 23 years later. Sirin (1982) conducted a meta-analysis on the studies on SES and academic achievement between the years of 1990 and 2000. According to the results, contrary to White (1982), studies conducted in the following years showed that the intensity of the relationship between SES and academic achievement grew and the value of the correlation changed from medium to high. When the meta-analysis studies are evaluated together, it can be said that the relationship between SES and academic achievement has become stronger in the following years. There are many studies in the literature that examine the relationships between home resources which are under the scope of SES and cognitive and affective characteristics (Yıldırım, 2019; Acar Güvendir, 2017; Bofah & Hannula, 2017; Caponera & Losito, 2016; Bouhlila, 2014; Walzeburg, 2014; Azina & Halimah, 2012; Shen, 2005). However, these past studies do not contain evidence that measurement invariance between groups is achieved. In the invariance studies in the literature, invariance was examined across genders, regions, cultures, and testing language (Kıbrıslıoğlu, 2015; Polat, 2015; Uyar & Doğan, 2014; Segeritz & Pant, 2013; Marsh et al., 2006; Erikan & Koh, 2005). Reviews based on home resources are not available. As in all group-based difference studies, researchers should test measurement invariances before performing studies based on SES variables and demonstrate that test invariance is ensured. Apart from these, measurement invariance studies based on large-scale tests in the literature have mainly investigated the invariance of student achievement/performance (Ölçülüoğlu & Çetin, 2016; Aliverinini, 2011; Teo, 2010; Wu, Li, & Zumbo, 2007), but studies investigating the invariance of affective characteristics are relatively few (Ertürk & Erdinç-Akan, 2018; Polat, 2019). The present study is important because it focuses on the invariance of the affective characteristics of students towards mathematics and also investigates invariance according to home resources, which is related to education but has not been addressed before.

2. METHOD

In this section, the population and sample of the study are defined; data collection and data analysis are discussed.

2.1. Population and Sample

In TIMSS, the population consists of the participating country's 4th and 8th grade students, and the sample consists of the students who took the exam. Students who take the exam are determined in two stages. Accordingly, in the first stage, the schools are selected by the stratified random sampling method, and in the second stage, the classes that will participate in TIMSS are selected by the random sampling. Since the data obtained from the last administration of TIMSS in 2019 had not been released yet at the time of this particular study, this study was based on the 2015 administration and was limited to the 4th grade level. Within this scope, 6.456 students participated in TIMSS 2015 4th grade from Turkey. Because of multivariable statistical analysis based on assumptions, in this research a data screening and cleaning phase was carried out. At the end of this phase 331 cases were cleaned and the sample of this study consisted of the remaining 6.125 students.

2.2. Data Source

The data were obtained from the database at https://timssandpirls.bc.edu/. In the TIMSS administration, student, teacher, school, and house questionnaires are included in addition to mathematics and science achievement tests. In the student questionnaires, students are asked for information such as gender, date of birth, place of birth of their parents, and home resources. In addition, the student questionnaire includes items that examine affective characteristics regarding mathematics and science.

In the TIMSS 2015 4th grade administration, questions regarding 11 home resources in the form of yes/no answers were asked to students. While 7 of the 11 items predict the same home resources in all countries, 4 of them are constructed according to the structure of each country as a country-specific indicator of wealth. Accordingly, the first seven items consist of questions such as whether students have their own room, desk, and PC/tablet. In the next four items, the existence of financial opportunities such as having a piano at home, having a swimming pool, or having water running from the tap is investigated according to the welfare level of the country. The heating system, cooling system, washing machine, and dishwasher facilities were asked as welfare indicators in the 2015 Turkey administration. In the study, it was observed that the number of students who do not have a washing machine (n=295) was significantly smaller than those who have a washing machine (n=6,073). For this reason, the washing machine, which is one of the country-specific indicators, was not included in the study. Also, students in all participating countries were asked whether they have an internet connection at home. Internet connection at home is considered important in accessing educational technologies and educational resources, so it was decided to be included in the study. As a result, the study was carried out based on the four home resources included in the TIMSS 2015 4th grade Turkey administration. The names and definitions of the variables included in the study are given in Table 1.

Table 1. *Names and definitions of home resources variables.*

Variable name	Variable definition
ASBG05E	Internet connection
ASBG05H	Heating System
ASBG05I	Cooling System
ASBG05K	Dishwasher

In TIMSS 2015 4th grade administration, there are 28 items as scored based on 5-point Likert type related to affective characteristics towards mathematics. These items are organized under three question themes in a test form; namely, mathematics lesson, learning mathematics, and mathematics. Home resources used in the study and affective characteristics data regarding mathematics are included in the file named ASGTURM6.

2.3. Data Analysis

The analyses of the research were carried out in various stages. Accordingly, in the first stage, the data were examined in order to test the assumptions. Individuals whose relevant home resources responses were missing were excluded from the study. The missing data in the responses to the affective items were analysed with missing data analysis, and it was observed that the values obtained were less than 5% and were randomly distributed. Missing data were completed by the item means method. For determining multivariate outliers Mahalanobis distances were examined. Accordingly, it was seen that there was no Mahalanobis value exceeding the critical chi-square value at $p < 0.001$. Descriptive statistics' examination showed the variables normally distributed. Skewness and kurtosis values were in expected range. Tavşancıl (2005) stated Bartlett's test of sphericity can be used for normality, and it was found that chi-square was 27293.564 and $p < 0.00$. This value shows that the data have a multivariate normal distribution. Tolerance, VIF, and condition index (CI) were examined for multicollinearity. Accordingly, the tolerance was found to be $=1.00$, VIF <5 and CI <30, and it was observed that there was no multi-collinearity problem in the data set (Tabachnick & Fidell, 2007). All these results show that factor analysis is applicable to the data.

After examining the assumptions, exploratory factor analysis was performed. Accordingly, 28 items asked in relation to mathematical affective characteristics were analysed. As a result of the analysis, KMO value was obtained as 0.930. This value is interpreted as perfect and means that the sample size is sufficient for factorability (Tabachnick & Fidell, 2007). According to the results obtained as a result of EFA, the items are collected under three factors with eigenvalues greater than 1. Accordingly, the first three eigenvalues are respectively 6.10; 1.72 and 1.49. Eigenvalues are 0.63 and less from the fourth factor. Based on these results, it can be stated that the items are grouped under three dimensions. When factor loadings are examined, it can be seen that items that have factor loadings in more than one dimension and whose difference between factor loadings are 0.1 or less are accepted as overlapping (Büyüköztürk, 2009). Accordingly, 3 of the 28 items (ASBM01B, ASBM03A and ASBM03D) were excluded from the data set because they were overlapping. As a result, a 3-dimensional structure that accounted for 36.006% of the total variance was obtained. Accordingly, there are 8 items in the dimension called liking mathematics, and the factor loadings of the items vary between 0.309 and 0.790. There are 10 items in the second dimension, called interest in mathematics. Factor loadings of the items in this dimension range from 0.314 to 0.590. In the third dimension, which is called self-confidence in mathematics, there are 7 items and the factor loadings of the items vary between 0.350 and 0.669. Table 2 contains the statistics of the structure reached as a result of EFA.

In addition to EFA, Velicer's maximum average partial (MAP) analysis was used to decide the number of factors. MAP results are included in Table 3. When Table 3 is examined, it is seen that the smallest average squared correlation takes the lowest value in the fourth step. The number of steps up to the fourth step gives the number of factors and it is seen that the number of dimensions according to the TR^2 value is three. O'Connor (2000) stated that the fourth power of partial correlation is an effective criterion. Accordingly, when the TR^4 value is examined, it is seen that it takes its smallest value in the fourth step. In this regard, the TR^4 value shows that the number of dimensions is three. Finally, when EFA and MAP results are evaluated together,

it can be stated that the results support each other and the affective characteristics for mathematics has a three-factor structure.

Table 2. *Questionnaire items, factor loadings and factors.*

Item code	Item	Liking Mathematics	Interest in Mathematics	Self-Confidence in Mathematics
ASBM01A	I enjoy learning mathematics	.790		
ASBM01C	Mathematics is boring	.488		
ASBM01D	I learn many interesting things in mathematics	.309		
ASBM01E	I like mathematics	.798		
ASBM01F	I like any schoolwork that involves numbers	.538		
ASBM01G	I like to solve mathematics problems	.630		
ASBM01H	I look forward to mathematics lessons	.648		
ASBM01I	Mathematics is one of my favorite subjects	.702		
ASBM02A	I know what my teacher expects me to do		.314	
ASBM02B	My teacher is easy to understand		.424	
ASBM02C	I am interested in what my teacher says		.498	
ASBM02D	My teacher gives me interesting things to do		.327	
ASBM02E	My teacher has clear answers to my questions		.590	
ASBM02F	My teacher is good at explaining mathematics		.510	
ASBM02G	My teacher lets me show what I have learned		.476	
ASBM02H	My teacher does a variety of things to help us learn		.499	
ASBM02I	My teacher tells me how to do better when I make a mistake		.576	
ASBM02J	My teacher listens to what I have to say		.587	
ASBM03B	Mathematics is harder for me than for many of my classmates			.669
ASBM03C	I am just not good at mathematics			.692
ASBM03E	Mathematics makes me nervous			.577
ASBM03F	I am good at working out difficult mathematics problems			.350
ASBM03G	My teacher tells me I am good at mathematics			.377
ASBM03H	Mathematics is harder for me than any other subject			.661
ASBM03I	Mathematics makes me confused			.631

Table 3. *Eigen Values Regarding Partial Correlations Obtained from the MAP Test.*

|	TR^2	TR^4		TR^2	TR^4		TR^2	TR^4
0	0.06555	0.00991	9	0.03196	0.00768	18	0.17084	0.07851
1	0.01779	0.00076	10	0.03951	0.01096	19	0.18635	0.07830
2	0.01421	0.00056	11	0.05099	0.01526	20	0.20042	0.08979
3	**0.00842***	**0.00025****	12	0.06017	0.01774	21	0.26284	0.14586
4	0.01010	0.00040	13	0.06642	0.01780	22	0.38916	0.26106
5	0.01249	0.00108	14	0.07968	0.02544	23	0.54993	0.42485
6	0.01594	0.00296	15	0.09420	0.03611	24	100.000	100.000
7	0.01969	0.00448	16	0.11363	0.04385			
8	0.02518	0.00684	17	0.13486	0.05890			

* The smallest average squared correlation
**The smallest average squared correlation's 4th power

Confirmatory factor analysis (CFA) was performed with Lisrel 8.24 to verify the model by EFA. EFA was run with ML algoritm and results showed there is no need for modifications. In the established model, it was found that $x^2 = 2605.21$, $df = 272$, $x^2 / df = 9.57$. In the model established with CFA, the x^2/df ratio is expected to be ≤ 3.00. However, x^2 statistics is sensitive to sample size, and as the sample size increases, this ratio exceeds 3 (Kline, 2011). Therefore, the model x^2/df value obtained was not interpreted as a model-data misfit and other fit statistics were examined. Accordingly, it was found that RMSEA=0.056, SRMR=0.052, CFI =0.96, and NNFI =0.96. Since all of these values indicated good fit, it was concluded that the model was validated (Kline, 2011). Correlations between factors are r_{12}=0.54; r_{23}=-0.53 and r_{13}=-0.34The path diagram for the model is presented in Figure 1.

After the mathematical affective characteristics model was verified, measurement invariance tests were carried out. Accordingly, the data set was analysed by MGCFA in configural, metric, scalar, and strict invariance stages separately for each home resources variable. The values of fit statistics, x^2, df, RMSEA, SRMR, NNFI and CFI were examined in each invariance stage. In addition, the ΔCFI values revealing the change in the CFI in the transition from the unconstrained model to the constrained model were examined in order to decide if invariance was achieved. In the literature, measurement invariance is examined according to the chi-square difference test and the difference in CFI. In various studies, the lack of significance of chi-square has been shown as evidence for measurement invariance (Hirschfeld & von Brachel, 2014; Brannick, 1995; Kelloway, 1995). However, as the chi-square is sensitive to the sample size, it tends to be significant in large samples. This situation is also valid for this study. Similarly, Cheung and Rensvold (2002) stated that the ΔX^2 test is sensitive to the sample size, the complexity of the model and is less effective in making practical decisions. Cheung and Rensvold (2002) examined 20 different fit indices in their study and stated that the strongest statistics to be examined in the test of intergroup invariance are ΔCFI, ΔGamma line, and ΔMcDonald's NCI. For these reasons, in making the decision about measurement invariance, it is taken as a reference whether the |ΔCFI| is <0.01 or not as stated by Wu, Li and Zumbo (2007).

Figure 1. *Path diagram for mathematical affective characteristics model.*
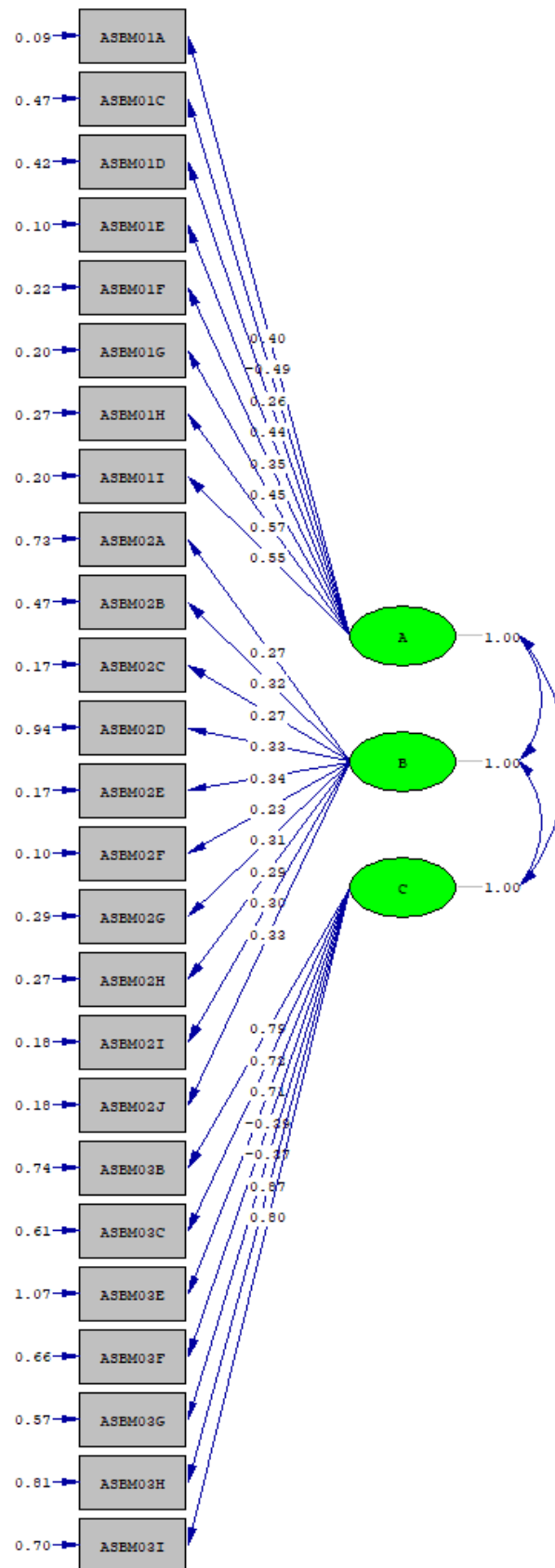
Chi-Square=2827.10, df=272, P-value=0.00000, RMSEA=0.056

3. FINDINGS

The MGCFA method was used for the measurement invariance test with Lisrel 8.54 in the study. In the analysis EM algoritm and covariance matrix were used. However, the validation of the model was tested first in each of the subgroups where invariance would be examined. Accordingly, the mathematical affective characteristics model was validated separately for two groups (according to the responses "yes, I have" and "no, I haven't") of the internet connection variable. The same procedure was carried out for the heating system, cooling system, and dishwasher variables. The fit statistics for the model verified in the groups created based on each variable are given in Table 4.

Table 4. *CFA fit statistics of the groups based on home resources variables.*

Variable	Group	x^2	df	x^2/df	RMSEA	SRMR	NNFI	CFI
Internet	Yes	3322.49	272	12.215	0.056	0.052	0.96	0.96
connection	No	2498.51	272	12.862	0.057	0.055	0.94	0.95
Heating system	Yes	2779.62	272	10.208	0.055	0.052	0.96	0.96
	No	2777.77	272	10.212	0.057	0.055	0.95	0.95
Cooling system	Yes	2313.34	272	8.504	0.057	0.055	0.95	0.96
	No	3443.07	272	12.658	0.055	0.053	0.95	0.96
Dishwasher	Yes	3968.37	272	14.589	0.056	0.052	0.96	0.96
	No	1804.79	272	6.635	0.057	0.058	0.94	0.95

According to Table 4, the x^2/df value was found to be >3 in the model established for each subgroup of variables. Since the x^2 statistics is sensitive to the sample size, x^2, df, x^2/df were reported in the following phases of the study, but other statistics were taken as a basis to decide if the model was validated. In Table 4, from fit statistics, it was found that RMSEA was <0.06, SRMR <0.08, and NNFI> 0.90, and this corresponds to a good fit; also that CFI $\geq$ 0.95 corresponds to a perfect fit (Hu & Bentler, 1999; Klein, 2011). These results show that the model is validated in the subgroups of each home resources variable. After the model was verified separately in each subgroup, measurement invariance analyses were initiated.

3.1. Measurement Invariance According to Internet Connection Variable

Whether or not there is an internet connection at home is one of the common questions asked regarding home resources in all countries. In Turkey, 58.4% of students (n=3576) have an internet connection at home and the remaining 41.6% (n=2549) do not. The results of measurement invariance across groups concerning students with and without internet connection are presented in Table 5.

Table 5. *Measurement invariance according to internet connection.*

Invariance type	x^2	df	x^2/df	RMSEA	SRMR	NNFI	CFI	ΔCFI
Configural	5821.00	544	10.700	0.056	0.055	0.95	0.96	-
Metric	5903.51	566	10.430	0.056	0.054	0.96	0.96	0.00
Scalar	6030.11	575	10.487	0.056	0.060	0.95	0.96	0.00
Strict	7082.32	597	11.863	0.060	0.061	0.95	0.95	-0.01

According to Table 5, it is seen that RMSEA is <0.08, SRMR <0.08, NNFI $\geq$ 0.95, and CFI $\geq$ 0.95. ΔCFI was calculated as 0.00 when changing from configural to metric, from metric to scalar and it became -0.01 when switching from scalar to strict. Based on model fit indexes and ΔCFI, the mathematical affective characteristics model ensures all stages of measurement invariance across groups of internet connection variable. According to this result, the factor

structure, item factor loadings, item constants, and error variances of mathematical affective characteristics do not differ depending on whether there is an internet connection at home or not. According to this result, mathematical affective characteristics can be compared significantly concerning the internet connection variable and it can be concluded that the possible differences are due to internet connection.

3.2. Measurement Invariance According to The Heating System Variable

One of the country-specific indicators of wealth concerning home resources is heating systems in TIMSS 2015 4th grade Turkey administration. Accordingly, 49.1% (n=3010) of the participating students have a heating system in their houses, whereas 50.9% (n=3115) do not. The findings regarding the invariance of the mathematical affective characteristics model across the sub-groups of the heating system are given in Table 6.

Table 6. *Measurement invariance according to heating system.*

Invariance type	x^2	df	x^2/df	RMSEA	SRMR	NNFI	CFI	ΔCFI
Configural	5829.38	544	10.716	0.056	0.052	0.95	0.96	-
Metric	5897.45	566	10.420	0.055	0.054	0.95	0.96	0.00
Scalar	6009.45	572	10.506	0.056	0.058	0.95	0.96	0.00
Strict	6857.99	597	11.487	0.059	0.061	0.95	0.95	-0.01

According to the results in Table 6, error indices for RMSEA were found to be <0.08 and for SRMR <0.08; and fit indices for NNFI and CFI were obtained as ≥ 0.95 for all invariance types. ΔCFI was calculated as 0.00 when changing from configural to metric, from metric to scalar, it takes -0.01 value when switching from scalar to strict invariance. These values obtained are within the accepted range indicating that invariance is achieved. The established model ensures all stages of measurement invariance in subgroups of the heating system. Accordingly, the factor structures, factor loadings, regression constants, and error variances obtained in both groups are equal. The differences of mathematical affective properties according to the heating system can be examined and the differences can be explained on the basis of home resource addressed.

3.3. Measurement Invariance According to Cooling System

In TIMSS 2015 Turkey administration, a cooling system is one of the home resources asked as a country-specific indicator of wealth. Students who have air conditioner-like devices as a cooling system account for 37.3% (n=2286) of all participants, and those who do not have a cooling system such as an air conditioner account for 62.7% (n=3839). Findings regarding the mathematical affective characteristics model invariance across groups based on the cooling system are given in Table 7.

Table 7. *Measurement invariance according to cooling system.*

Invariance type	x^2	df	x^2/df	RMSEA	SRMR	NNFI	CFI	ΔCFI
Configural	5756.41	544	10.582	0.056	0.055	0.95	0.96	-
Metric	5811.24	566	10.267	0.055	0.055	0.96	0.96	0.00
Scalar	5857.85	572	10.241	0.055	0.065	0.96	0.96	0.00
Strict	6123.63	597	10.257	0.055	0.068	0.96	0.96	0.00

According to Table 7, RMSEA and SRMR were found to be <0.08, NNFI and CFI were found to be ≥ 0.95. Since CFI was 0.96 in all invariance models, all values of ΔCFI were equal to 0.00. When the model fit statistics are evaluated together, it is seen that the mathematical affective characteristics model is invariant based on the groups of the cooling system.

Accordingly, it was shown that the factor structure, factor loadings, regression constants, and error variances of the mathematical affective characteristics model were equal in the two groups. Therefore, the mathematical affective characteristics model can be significantly compared and interpreted according to the cooling system variable.

3.4. Measurement Invariance According to The Dishwasher Variable

Dishwasher was considered a country-specific indicator of wealth in TIMSS 2015 4th grade Turkey administration. Accordingly, 71.4% (n=4376) of the participating students had a dishwasher at home, whereas 28.6% (n=1749) did not. The measurement invariance results of the mathematical affective characteristics model based on the dishwasher variable are presented in Table 8.

Table 8. *Measurement invariance according to the dishwasher variable,*

Invariance type	x^2	df	x^2/df	RMSEA	SRMR	NNFI	CFI	ΔCFI
Configural	5773.16	544	10.612	0.056	0.052	0.95	0.96	-
Metric	5860.76	566	10.355	0.055	0.052	0.96	0.96	0.00
Scalar	5955.19	572	10.411	0.055	0.054	0.96	0.96	0.00
Strict	7297.59	597	12.224	0.061	0.055	0.95	0.95	-0.01

According to Table 8, RMSEA and SRMR values were <0.08; fit statistics NNFI and CFI were ≥ 0.95. ΔCFI was calculated as 0.00 when changing from configural to metric, from metric to scalar, and as -0.01when changing from scalar to strict. When the statistics in Table 8 are evaluated together, the mathematical affective characteristics model ensures measurement invariance across the dishwasher-based groups. The factor structure, factor loadings, regression constants, and error variances of the model are identical across groups. Mathematical affective characteristics can be meaningfully compared and interpreted based on the dishwasher.

4. DISCUSSION and CONCLUSION

This study investigated whether the mathematical affective characteristics model proposed based on the TIMSS 2015 4th grade Turkey administration showed measurement invariance according to home resources or not. As a result of the study, it was shown that the variables of internet connection, heating system, cooling system, dishwasher, which are considered within the scope of home resources, provide configural, metric, scalar, and strict measurement invariance across the subgroups, respectively. Accordingly, the means, variances, covariances, and item residual variances of the model are identical across the subgroups of each established home resource. The results indicate that the mean of observed scores obtained from the mathematical affective characteristics scale can be compared according to home resources. The results of further research to be obtained by making comparisons are meaningful and possible differences can be attributed to the relevant home resource.

Although there is no similar study in the literature examining measurement invariance based on home resources, there are various measurement invariance studies based on large-scale exams. One of these is the study by Hansson and Gustafsson (2013), which examined whether or not the socio-economic status is invariant according to the ethnic structure using TIMSS 2003 data. In their study, Hansson and Gustafsson (2013) tested the invariance of the latent SES variable between Swedish and non-Swedish groups and found that configural invariance was achieved, but scalar invariance was not. Ertürk and Erdinç-Akan (2018) and Polat (2019) focused on the mathematical affective characteristics questionnaire of the TIMSS 2015 administration in their studies. Accordingly, Ertürk and Erdinç-Akan (2018) examined the gender-based measurement invariance of variables related to mathematics achievement based on the 4th grade administration. According to the results of the study, they found that the liking

mathematics scale provides strict invariance, and interest in mathematics and mathematical self-confidence scales provide configural invariance. Polat (2019) investigated the invariance of both mathematical as well as the cultural affective characteristics questionnaire according to cultures (Turkey, Singapore, and Saudi Arabia) and regions (NUTS-Level 1), and gender based on the TIMMS 2015 8th grade administration. The study showed that the established mathematical and science affective characteristics models provide scalar invariance across cultures and regions and strict invariance across genders.

Some of the invariance studies carried out based on PISA are the studies by Kıbrıslıoğlu (2015), Güngör and Kabasakal (2020), and Uyar and Uyanık (2019). Kıbrıslıoğlu (2015) investigated the invariance of the PISA 2012 mathematical learning model across cultures (Turkey, China-Shanghai, and Indonesia) and genders. The study showed that the model provides only configural invariance across cultures. The study examined gender-based measurement invariance based on all of the data set obtained from three cultures, and as a result, the study showed that the mathematical learning model provides strict invariance across genders. Güngör and Kabasakal (2020) investigated the measurement invariance of instrumental motivation and science self-efficacy scales in science teaching according to gender and regions based on PISA 2015 Turkey administration. Güngör and Kabasakal (2020) reported that only configural invariance was achieved based on gender, and metric invariance was achieved across regions. Uyar and Uyanık (2019) established a learning model for science by using the questionnaire in the PISA 2015 administration and investigated the invariance of the established model according to gender in Turkey sample and the invariance of the established model in Turkey-Singapore samples according to cultures. As a result, Uyar and Uyanık (2019) found that across genders metric invariance and across cultures configural invariance was achieved.

When the above-mentioned studies are evaluated together, it is seen that strict invariance based on gender is ensured under certain conditions in large-scale exams, and there are no studies ensuring strict invariance based on cultures. However, there are no studies carried out based on home resources in large-scale exams or SES in general that can be compared with the findings of the present study. In this regard, researchers are recommended that they investigate measurement invariance based on variables such as parental education level, parental income, number of siblings, along with other home resources not included in this study, and to address variables that ensure strict invariance in comparisons across groups.

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

Derya CAKICI ESER ⓘ https://orcid.org/0000-0002-4152-6821

5. REFERENCES

Acar Güvendir, M. (2017). Determination of the relationship between the students mathematical literacy and home and school educational resources in program for international student assessment - (PISA 2012). *Mersin University Journal of the Faculty of Education*, *13*(1). https://doi.org/10.17860/mersinefd.305762

Alivernini, F. (2011). Measurement invariance of a reading literacy scale in the Italian context: A psychometric analysis. *Procedia Social and Behavioral Sciences*, *15*, 436-441. https://doi.org/10.1016/j.sbspro.2011.03.117

Azina, I. N., & Halimah, A. (2012). Student factors and mathematics achievement: Evidence from TIMSS 2007. *Eurasia Journal of Mathematics, Science and Technology Education*, *8*(4), 249-255. https://doi.org/10.12973/eurasia.2012.843a

Başusta, N. B., & Gelbal, S. (2015). Examination of measurement invariance at groups' comparisons: A study on PISA student questionnaire, *Hacettepe University Journal of Education*, *30*(4), 80-90.

Bouhlila, D. S. (2014). The impact of socioeconomic status on students' achievement in the Middle East and North Africa: An essay using the TIMSS 2007 database. *International Perspectives on Education and Society*, *24*, 199-226 https://doi.org/10.1108/S1479-367920140000024017

Bofah, E. A. T., & Hannula, M. S. (2017). Home resources as a measure of socio-economic status in Ghana. *Large-scale Assessments in Education,* 5(1), 1-15. https://doi.org/10.1186/s40536-017-0039-5

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, 16(3), 201-213. https://doi.org/10.1002/job.4030160303

Büyüköztürk, Ş. (2009) *Sosyal Bilimler İçin Veri Analizi El Kitabı*, Pegem Akademi

Caponera, E., & Losito, B. (2016). Context factors and student achievement in the IEA studies: Evidence from TIMSS. *Large-scale Assessments in Education*, *4*(1),12. https://doi.org/10.1186/s40536-016-0030-6

Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods*, *15*(2), 167-198. https://doi.org/10.1177/1094428111421987

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness of fit indices for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French version of TIMSS. *International Journal of Testing*, *5*, 23-35. https://doi.org/10.1207/s15327574ijt0501_3

Ertürk, Z., & Erdinç-Akan, O. (2018). The investigation of the variables effecting TIMSS 2015 mathematics achievement with SEM, *Journal of Theoretical Educational Science*, *2*, 14-34.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*, 78-94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Gustafsson, J. E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, *57*, 16-30. https://doi.org/10.1016/j.stueduc.2016.09.004

Hansson, Å., & Gustafsson, J.-E. (2013). Measurement invariance of socioeconomic status across migrational background. *Scandinavian Journal of Educational Research*, *57*(2), 148–166. https://doi.org/10.1080/00313831.2011.625570

Hirschfeld, G., & Von Brachel, R. (2014). Improving multiple-group confirmatory factor analysis in R–A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, and Evaluation*, *19*(1), 7. https://doi.org/10.7275/qazy-2946

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, M., P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

International Association for the Evaluation of Educational Achievement, [IEA], (2019). Chapter 3 TIMSS 2019 context questionnaire framework, https://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks-Chapter-3.pdf

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. IL: Scientific Software International, Inc.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior, 16*(3), 215-224. https://doi.org/10.1002/job.4030160304

Kıbrıslıoğlu, N. (2015). *The investigation of measurement invariance PISA 2012 mathematics learning model according to culture and gender: Turkey - China (Shangai) - Indonesia*, [Graduate Thesis, Hacettepe University].

Kline, R. B., (2011). *Principles and Practices of Structural Equation Modelling*. The Guilford Press.

Marsh, H. W., Hau, K. T., Artelt, C., Boument, J., & Peschar, J. (2006). OECD's brief selfreport measure of educational psychology's most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6* (4), 311-360. https://doi.org/10.1207/s15327574ijt0604_1

McGaw, B., & Jöreskog, K. G. (1971). Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. *British Journal of Mathematical and Statistical Psychology, 24*(2), 154-168. https://doi.org/10.1111/j.2044-8317.1971.tb00463.x

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*, 525-543. https://doi.org/10.1007/BF02294825

Millsap, R. E., & Olivera-Aguilar, M. (2012). Investigating measurement invariance using confirmatory factor analysis. In R. H. Hoyle, (Ed.) *Handbook of structural equation modeling,* (pp. 380-392), Guilford.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance,* Routledge.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, &Computers, 32*, 396-402. https://doi.org/10.3758/BF03200807

Ölçüoğlu, R., & Çetin, S. (2016). The investigation of the variables that affecting eight grade students' TIMSS 2011 math achievement according to regions, *Journal of Measurement and Evaluation in Education and Psychology 7*(1), 202-220. https://doi.org/10.21031/epod.34424

Polat, M. (2019). *The investigation of measurement invariance of TIMSS-2015 mathematics and science affective characteristics models according to culture, gender and statistical region,* [Graduate Thesis, Hacettepe University].

Schmith, N., & Kuljanin, G. (2008). Measurement invariance: review of practice and implication. *Human Resources Management Review, 18*, 210-222. https://doi.org/10.1016/j.hrmr.2008.03.003

Segeritz, M., & Pant, H. A. (2013). Do they feel the same way about math? Testing measurement invariance of the PISA students' approaches to learning instrument across immigrant groups within Germany. *Educational and Psychological Measurement, 73*(4), 601-630 https://doi.org/10.1177/0013164413481802

Shen, C. (2005). How American middle schools differ from schools of five Asian countries: Based on cross-national data from TIMSS 1999. *Educational Research and Evaluation*, *11*(2), 179-199. https://doi.org/10.1080/13803610500110810

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417-453. https://doi.org/10.3102/00346543075003417

Tabachnick, B. G., & Fidell, L.S. (2007). *Using Multivariate Statistics*, (5. ed.) Pearson Education.

Tavşancıl, E. (2005) *Tutumların Ölçülmesi ve SPSS ile Veri Analizi [Measuring Attitudes and Data Analysis with SPSS]*, Nobel Yayınları

Teo, T. (2010). Gender differences in the intention to use technology: A measurement invariance analysis. *British Journal of Educational Technology*, *41*(6), 120-124. https://doi.org/10.1111/j.1467-8535.2009.01023.x

Uyar. Ş., & Doğan, N. (2014). An investigation of measurement invariance of learning strategies model across different groups in PISA Turkey sample, *International Journal of Turkish Educational Studies, 2*, 30-43.

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

Walzebug, A. (2014). Is there a language-based social disadvantage in solving mathematical items?. *Learning, Culture and Social Interaction*, *3*(2), 159-169. https://doi.org/10.1016/j.lcsi.2014.03.002

White, K. R. (1982). The relation between socioeconomic status and academic achievement, *Psychological Bulletin*, *91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, *12*(3). https://doi.org/10.7275/mhqa-cd89

Yıldırım, S. (2019). Predicting mathematics achievement: The role of socioeconomic status, parental involvement, and self-confidence, *Education and Science*, *44*(198), 99-113. https://doi.org/10.15390/EB.2019.7868