ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2023; 14(3); 185-199

Investigation of The Measurement Invariance of Affective Characteristics Related to TIMSS 2019 Mathematics Achievement by Gender*

Mehmet ATILGAN**   Kaan Zülfikar DENİZ***

Abstract

This research examines whether the affective characteristics of the TIMSS 2019 Turkey mathematics application provide measurement invariance according to gender. The research sample consists of 4048 8th-grade students participating in the TIMSS in 2019. Research data were downloaded from the international website of TIMSS. The research data collection tools are "Sense of School Belonging", "Students Confident in Mathematics", "Students Like Learning Mathematics", and "Students Value Mathematics" scales. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were performed in the context of validity analyses to examine measurement invariance. In terms of reliability, the Cronbach Alfa internal consistency coefficient was calculated. Accordingly, out of the four scales in the study, only "Students Confident in Mathematics" scale could not be confirmed in confirmatory factor analysis. Therefore, while "Students Confident in Mathematics" scale was not examined for measurement invariance, the other three scales were examined within the scope of measurement invariance. For measurement invariance, research data were tested with Multiple Group Confirmatory Factor Analysis (MG-CFA), one of the Structural Equation Modeling (SEM) techniques. As a result of the analyses, while the strict invariance model was provided in "Students Like Learning Mathematics" scale and "Students Value Mathematics" scale, strong invariance/scale invariance model was provided in "Sense of School Belonging" scale. It was concluded that there was no gender bias in the three scales for which MG-CFA was performed, and the mean scores were comparable according to gender. In this context, it can be said that "Sense of School Belonging", "Students Like Learning Mathematics", and "Students Value Mathematics" scales are valid in determining the differences according to gender.

Keywords: TIMSS, affective variables, measurement invariance, MG-CFA, SEM

Introduction

Raising qualified people is one of the most critical issues for countries. Education systems play a significant role in raising qualified people. States change their education policies over time and make arrangements in their education systems to train qualified people with the desired characteristics.

In Turkey, regulations have been made in the education system over time. These arrangements are made through the findings obtained from the national and international measurement and evaluation practices in which Türkiye has participated. Türkiye has been participating in international educational studies such as TIMSS (Trends in International Mathematics and Science Study), PISA (Programme for International Student Assessment), and PIRLS (Progress in International Reading Literacy Study) for many years. Türkiye participates in these studies to compare the education system of Türkiye with the education systems of others, to reveal the situation of Türkiye on an international scale, to eliminate the deficiencies in the education system based on the findings of these studies, and to make adjustments in education policies.

*This study has been produced from Master's Thesis that was conducted under the supervision of the Prof. Dr. Kaan Zülfikar Deniz and prepared by Mehmet ATILGAN.

**Research Assistant, Uşak University, Faculty of Education, Uşak-Türkiye, mehmet.atilgan@usak.edu.tr, ORCID ID: 0000-0003-1297-4630
***Prof. Dr., Ankara University, Faculty of Education, Ankara-Türkiye, zlfkrdnz@yahoo.com, ORCID ID: 0000-0003-0920-538X

To cite this article:
Atılgan, M. & Deniz, K. Z. (2023). Investigation of the measurement invariance of affective characteristics related to TIMSS 2019 mathematics achievement by gender. *Journal of Measurement and Evaluation in Education and Psychology, 14*(3), 185-199. https://doi.org/10.21031/epod.1221365

Received: 19.12.2022
Accepted: 7.09.2023

One of the most numerous international education studies that Türkiye has participated in is the TIMSS, organized by the IEA (International Association for the Evaluation of Educational Achievement). TIMSS was first implemented in 1995. Türkiye attended TIMSS for the first time in 1999 and, finally, in 2019 (Ministry of National Education-MoNE, 2020).

TIMSS is an educational study aiming to evaluate the knowledge and skills of 4th and 8th-grade students in mathematics and science. Since many variables affect students' success, detailed data about students, teachers, schools, and parents are collected through questionnaires within the scope of TIMSS (MoNE, 2016). Data on determining affective characteristics such as motivation, interest, and attitude are collected through TIMSS student questionnaires (Mullis et al., 2016).

Bloom (2012) states that affective characteristics have a 25% effect on students' academic success. In the literature, there are also studies showing that affective characteristics affect mathematics achievement (Doğan & Barış, 2010; İlhan & Öner-Sünkür, 2012; Kesici, 2018; Kesici & Aşılıoğlu, 2017; Lay et al., 2015; Mohammadpour, 2012; Ölçüoğlu & Çetin, 2016; Sarı & Ekici, 2018; Sarıer, 2020; Yücel & Koç, 2011).

Demographic variables are influential on academic achievement. Studies in the literature aim to determine at what level demographic variables such as gender, socioeconomic level, age, class, and geographical region affect success. Among these variables, studies on the gender variable attract attention. The number of studies comparing the mean scores of gender groups to examine the effect of TIMSS mathematics achievement is relatively high (Aydın, 2015; Hanci, 2015; Kilic & Askin, 2013; Louis & Mistele, 2012; Patterson et al., 2003; Sarıer, 2020; Wang et al., 2012; Webster & Fisher, 2000).

In the 8th-grade Türkiye sample of TIMSS 2019, the mean mathematics scores for male and female groups are 490 and 501, respectively. However, the difference in scores between the averages was not statistically significant (MoNE, 2020). The measurement results obtained regarding the comparison of the groups may vary depending on different characteristics of the individuals. However, the source of the differences may only sometimes be individuals. The reason can sometimes be the measurement tool itself. When comparing the measurements according to the groups, it is assumed that the measurement tool measures the same feature for all groups. In other words, measurement invariance is ensured (Başusta & Gelbal, 2015). However, it is crucial to prove that measurement invariance is ensured to conduct comparison studies with groups more validly and reliably.

Measurement invariance is defined as the same perception and interpretation of the items in the measurement tool in all groups subject to measurement (Byrne & Watkins, 2003). In the scales developed to reveal a latent structure, measurement invariance appears as one of the psychometric properties (Öncü, 2019). The measurement tool should measure the same structure in the groups to ensure measurement invariance. Factor loadings, correlations between factors, and error variances of the scale items should be equal to measure the same structure in groups (Byrn et al., 1989). There is a consensus in the literature that to compare mean scores by groups, measurement invariance should be tested, evidence of strong/scalar invariance model should be obtained, and comparison of mean scores without these conditions may not yield significant results (Başusta & Gelbal, 2015; Cheung & Rensvold, 2000; Gregorich, 2006; Kıbrıslıoğlu, 2015; Öğretmen, 2006; Salzberger et al., 1999; Vandenberg & Lance, 2000; Wicherts, 2007; Wu et al., 2007). The purpose of statistical analyses to test measurement invariance is to determine whether the established structural model is the same in subgroups and which of the parameters included in the structural model are invariant (Mulaik, 2007).

While the methods in Structural Equation Modeling (SEM) and Item Response Theory (IRT) approaches are primarily preferred in determining measurement invariance, methods based on the Latent Class Analysis (LCA) approach have also been used in recent years (Yandı et al., 2017). 80% of measurement invariance studies are conducted with approaches based on SEM (Vandenberg & Lance,2000). The MG-CFA (Multiple Group Confirmatory Factor Analysis) method is most frequently used in SEM-based approaches. Measurement invariance can be tested by examining the equality of mean covariance structures with the MG-CFA method (Yandı et al., 2017).

In TIMSS and PISA literature, measurement invariance was examined according to countries in some studies (Ercikan & Koh, 2005; Karakoc-Alatli et al., 2016; Öncü, 2019; Rutkowski & Rutkowski, 2013;

Tavlıca, 2019; Wu et al., 2007; Ma & Qin, 2021; Meng et al., 2019; Polat, 2019; Scherer et al., 2016). The subjects of these studies are mathematics and science achievement, socioeconomic level, affective variables related to mathematics and science, and using information and communication technologies. Some studies also examined measurement invariance according to gender (Ertürk & Erdinç-Akan, 2018b; Polat, 2019), geographical regions (Ölçüoğlu & Çetin, 2016; Polat, 2019), and household resources (Cakici-Eser, 2021). The subjects of these studies are affective variables related to mathematics and science, home environment, and school environment.

There are quite a lot of studies (Aydın, 2015; Kilic & Askin, 2013; Louis & Mistele, 2012; Patterson et al., 2003; Sarıer, 2020; Wang et al., 2012; Webster & Fisher, 2000) comparing gender groups on TIMSS mathematics achievement without testing measurement invariance. The literature states that measurement invariance must first be ensured. If it is not ensured, the comparisons may not yield meaningful results (Başusta & Gelbal, 2015; Cheung & Rensvold, 2000; Gregorich, 2006; Kıbrıslıoğlu, 2015; Salzberger et al., 1999; Vandenberg & Lance, 2000; Wicherts, 2007; Wu et al., 2007). For this reason, it is vital to test the measurement invariance before examining the effect of gender on mathematics achievement. Thanks to measurement invariance analysis, the way of interpreting the items of the subgroups can be determined, and it can be tested whether there is a bias of the subgroups in the scale's items (Byrne, 1998; Gregorich, 2006; Kıbrıslıoğlu, 2015; Millsap & Olivera-Aguilar, 2012). Failure to provide measurement invariance indicates that some items in the scale are biased. Some studies conducted according to gender have provided measurement invariance (Bofah & Hannula, 2015; Demir, 2020; Demir, 2017; Gungor & Atalay-Kabasakal, 2020; Jung, 2019; Polat, 2019; Uyar, 2021). However, measurement invariance cannot be achieved in some (Ertürk & Erdinç-Akan, 2018b; Gülleroğlu, 2017; Uzun & Öğretmen, 2010). Since some studies point to gender bias, this study aims to test the measurement invariance of affective characteristics related to TIMSS 2019 mathematics achievement according to gender groups. Measurement invariance studies, which indicate biases according to gender, show that the degree of accuracy of decisions taken about individuals may be inadequate (Öğretmen, 2006).

In examining the relationships between TIMSS student questionnaires and mathematics achievement, “Sense of School Belonging” (Akyüz & Pala, 2010; Akyüz & Satıcı, 2013; Işlak, 2020; Koç, 2019; Sarı et al., 2017; Sarıer, 2020), “Students Confident in Mathematics” (Akyüz-Aru, 2020; Akyüz & Pala, 2010; Atar, 2011; Aydın, 2015; Demir et al., 2010; Ertürk & Erdinç-Akan, 2018a; Işlak, 2020; Khine et al., 2015; Koç, 2019; Oral & McGivney, 2013; Usta & Demirtaşlı, 2018), “Students Like Learning Mathematics” (Erşan, 2016; Ertürk & Erdinç-Akan, 2018a; Khine et al., 2015; Koç, 2019; Oral & McGivney, 2013), and “Students Value Mathematics” (Doğan & Barış, 2010; Khine et al., 2015) scales have been used by some researchers. These affective variables are frequently used in the TIMSS literature.

In the literature, although many studies examine the effect of “Sense of School Belonging” scale on mathematics achievement (Akyüz & Pala, 2010; Akyüz ve Satıcı, 2013; Işlak, 2020; Koç, 2019; Sarı et al., 2017; Sarıer, 2020), no measurement invariance research has been found. While measurement invariance was confirmed in some studies testing the measurement invariance of “Students Confident in Mathematics” scale (Bofah & Hannula, 2015; Cakici-Eser, 2021; Polat, 2019; Uyar, 2021), measurement invariance could not be achieved in some (Ertürk & Erdinç-Akan, 2018b). In the studies that test the measurement invariance of “Students Like Learning Mathematics” (Bofah & Hannula,2015; Cakici-Eser, 2021; Ertürk & Erdinç-Akan, 2018b; Polat, 2019; Shukla & Konold, 2014) and “Students Value Mathematics” (Bofah & Hannula,2015; Polat, 2019; Uyar, 2021) scales, measurement invariance was achieved, and no research was found in which measurement invariance could not be achieved.

Accordingly, this study aimed to examine the measurement invariance of “Sense of School Belonging”, “Students Confident in Mathematics”, “Students Like Learning Mathematics”, and “Students Value Mathematics” scales of TIMSS 2019 Turkey 8th-grade in the context of gender.

Method

This research, which aims to examine the measurement invariance of the affective characteristics of the students in the Turkish sample who participated in the TIMSS 2019 mathematics according to gender, is descriptive. Studies that aim to reveal a situation without intervening are a type of descriptive research (Fraenkel & Wallen, 2006; Karasar, 2011).

Participants

TIMSS 2019 was held with the participation of 4077 eighth-grade students from 181 schools in Turkey. However, it was determined that 29 of these students left all the scales in the student questionnaire blank. For this reason, 29 students were excluded from the analysis, and the participant group consisted of 4048 students. The descriptive statistics of the participant group are shown in Table 1.

Table 1

Descriptive Statistics of Participant Group

Gender	*f*	%	Age average
Female	2009	49.63	13.89
Male	2039	50.37	13.92

Data Collection Tools

The data obtained from "Sense of School Belonging", "Students Confident in Mathematics", "Students Like Learning Mathematics", and "Students Value Mathematics" scales in the TIMSS 2019 mathematics student questionnaire were used in this research. Data were downloaded from the TIMSS research international website (https://timss2019.org/international-database/).

"Sense of School Belonging" scale consists of 5 items and a single factor, with a 4-point Likert-type rating. The items are scored as "1= disagree a lot", "2= disagree a little", "3= agree a little", and "4= agree a lot". There is no reverse-coded item. Higher scores on the scale indicate a higher sense of belonging to the school. Regarding the validity of the TIMSS Turkey sample data set, item factor loadings for this scale ranged from 0.58 to 0.77, and the total explained variance rate was 51%. Regarding reliability, the Cronbach Alpha coefficient was 0.76 (Yin & Fishbein, 2020).

"Students Confident in Mathematics", "Students Like Learning Mathematics", and "Students Value Mathematics" scales consist of 9 items and a single factor, with a 4-point Likert-type rating. Items 2, 3, 5, 8, and 9 were reverse-coded for "Students Confident in Mathematics" scale, and items 2 and 3 were reverse-coded for "Students Like Learning Mathematics" scale. There is no reverse-coded item for "Students Value Mathematics" scale. Higher scores on these scales indicate higher self-confidence, liking, and value in mathematics. In the TIMSS Turkey sample data set, item factor loadings ranged from 0.62 to 0.80 for "Students Confident in Mathematics", 0.61 to 0.89 for "Students Like Learning Mathematics", and 0.58 to 0.81 for "Students Value Mathematics". The total explained variance was 54% for "Students Confident in Mathematics", 62% for "Students Like Learning Mathematics", and 51% for "Students Value Mathematics". The Cronbach Alpha coefficient was 0.89 for "Students Confident in Mathematics", 0.89 for "Students Like Learning Mathematics", and 0.88 for "Students Value Mathematics" (Yin & Fishbein, 2020).

Data Analysis

Data analysis was carried out in three stages. In the first stage, the processes of examining the missing data, extreme values, and normality were followed, sequentially. In the second stage, EFA (Exploratory Factor Analysis) and CFA (Confirmatory Factor Analysis) were performed to create affective trait models associated with mathematics achievement. Cronbach Alpha internal consistency coefficients for each affective trait model created for reliability were calculated. In the last stage, MG-CFA (Multiple

Group Confirmatory Factor Analysis) was performed to determine the measurement invariance according to gender groups in the validated models.

SPSS IBM 20.0 and R Studio were used to analyze the data. For MG-CFA, semTools (Jorgensen et al., 2021) and lavaan (Rosseel, 2012) packages were used.

Results

Before testing the measurement invariance for each scale, missing data, extreme values, and normality were examined in terms of the suitability of the data for analysis. As a result of missing data analysis, 47 participants for "Students Value Mathematics", 90 for "Students Like Learning Mathematics", 52 for "Students Confident in Mathematics", and 16 for "Sense of School Belonging" scales were excluded from the analysis. The extreme value analysis converted each scale's items into Z scores. No extreme values were found except -4 and +4 (Çokluk et al., 2016; Harrington, 2009; Pituch & Stevens, 2016; Tabachnick & Fidell, 2013). In the normality examination, the skewness and kurtosis values were calculated separately for all the items in each scale. It has been determined that all related scales have skewness and kurtosis values except -1 and +1 (Çokluk et al., 2016; Hair et al., 2014; Harrington, 2009; Raykov & Marcoulides, 2006). For this reason, the relevant scales did not show a normal distribution.

In the second stage, the validity and reliability of the scores collected from scales were discussed. According to the EFA, the item factor loadings for "Students Value Mathematics" ranged from 0.58 to 0.81, for "Students Like Learning Mathematics" 0.60 to 0.89, for "Students Confident in Mathematics" 0.62 to 0.79, and for "Sense of School Belonging" 0.57 to 0.77. The total explained variance was 50.39% for "Students Value Mathematics", 61.19% for "Students Like Learning Mathematics", 52.45% for "Students Confident in Mathematics", and 50.49% for "Sense of School Belonging". The DWLS (Diagonally Weighted Least Squares) method is used when the number of categories in the scoring of the items in the Likert-type graded scales at the ranking level is less than five, and the multivariate normality requirement cannot be met in the data set (Kline, 2015; Mindrila, 2010; Schumacker & Beyerlein, 2000). Therefore, the DWLS method was preferred as the estimation method in CFA. $\chi2$, CFI, TLI, RMSEA and SRMR goodness-of-fit indices are used in this study to evaluate CFA results. The criterion values are presented in Table 2 (Çokluk et al., 2016; Harrington, 2009; Kline, 2015; Tabachnick & Fidell, 2013).

Table 2

Criterion Values in Goodness of Fit Indices

Fit Index	Good Fit			Acceptable Fit		
χ^2		*p*>	.05		*p*>	.05
CFI	0.95 ≤	CFI	≤ 1.00	0.90 ≤	CFI	≤ 0.95
TLI	0.95 ≤	TLI	≤ 1.00	0.90 ≤	TLI	≤ 0.95
RMSEA	0 ≤	RMSEA	≤ 0.05	0.05 ≤	RMSEA	≤ 0.08
SRMR	0 ≤	SRMR	≤ 0.05	0.05 ≤	SRMR	≤ 0.08

CFA was conducted for "Students Value Mathematics" scale with two modifications. Covariance was established between M3 and M4 items and M1 and M2 items with the recommendation of the R program. CFA was conducted for "Students Like Learning Mathematics" scale with a modification. Covariance was established between M2 and M3 items with the recommendation of the R program. Without modifications, CFA was conducted for "Students Confident in Mathematics" and "Sense of School Belonging" scales. After all these procedures, the CFA results for the four scales are presented in Table 3.

Table 3
CFA Results of Affective Scales

Scales	χ^2	CFI	TLI	RMSEA	SRMR
Students Value Mathematics	135.667 ($p< .05$)	.993	.990	.033 *(.028, .039)	.036
Students Like Learning Mathematics	116.191 ($p< .05$)	.998	.997	.030 *(.024, .035)	.027
Students Confident in Mathematics	1510.635 ($p< .05$)	.954	.938	.117 *(.112, .122)	.094
Sense of School Belonging	9.307 ($p> .05$)	.999	.997	.015 *(.000, .029)	.016

*Lower and upper confidence interval for RMSEA

Table 3 shows that CFI, TLI, RMSEA and SRMR values for "Students Value Mathematics" and "Students Like Learning Mathematics" scales indicate good fit, and the $\chi2$ value is not within acceptable fit ranges. When the literature is examined, it is stated that the sample size affects the $\chi2$ (Kline, 2015). When all goodness-of-fit indices are evaluated together, it can be said that "Students Value Mathematics" and "Students Like Learning Mathematics" scales are confirmed. CFI value for "Students Confident in Mathematics" scale indicates a good fit. TLI value is acceptable, and the $\chi2$, RMSEA, and SRMR values are not at acceptable ranges. When all goodness-of-fit indices are evaluated together, it can be said that "Students Confident in Mathematics" scale cannot be confirmed. For "Sense of School Belonging" scale, $\chi2$, CFI, TLI, RMSEA and SRMR values indicate a good fit. When all goodness-of-fit indices are evaluated together, it can be said that "Sense of School Belonging" scale is confirmed.

Regarding reliability, the Cronbach Alpha coefficient was calculated for three scales except for "Students Confident in Mathematics" scale because the scale was not confirmed by CFA. Cronbach Alpha coefficients were calculated as 0.87, 0.92, and 0.75 for "Students Value Mathematics", "Students Like Learning Mathematics", and "Sense of School Belonging" scales, respectively. A Cronbach Alpha coefficient of 0.70 and above indicates that the level of reliability is good, and a value between 0.60 and 0.70 indicates that the level of reliability is acceptable (Hair et al., 2014). In this context, the reliability of the three scales is reasonable.

Measurement Invariance

In the data analysis, measurement invariance according to gender was tested for three affective scales, which CFA confirmed at the last stage. MG-CFA method was used to test the measurement invariance. Measurement invariance by MG-CFA method, structural invariance, weak/metric invariance, strong/scalar invariance, and strict invariance models are examined by looking for evidence. Measurement invariance models have a 4-stage hierarchical structure (Byrne et al., 1989; Stark et al., 2006; Vandenberg & Lance, 2000; Wu et al., 2007). The model with the minor parameter constraints is the structural invariance model, and the model with the most parameter constraint is the strict invariance model. Due to the hierarchical structure of the invariance models, if there is no evidence that measurement invariance is provided for the model with fewer parameter constraints, there will be no evidence that measurement invariance is provided for the models with more parameter constraints. The goodness of fit values at that stage is considered when looking for evidence for invariance models. Then, the $\Delta\chi^2$ value between it and the previous model, which has fewer parameter limitations, is considered. Suppose the $\Delta\chi^2$ value is not statistically significant ($p> .05$), and the goodness-of-fit values of the model with more parameter limitations are within acceptable values. In that case, evidence of measurement invariance is obtained for the model with more parameter limitations. However, since the $\Delta\chi^2$ value is affected by the sample size, the p-value in the $\Delta\chi^2$ test tends to be significant. For this reason, it is stated by some researchers that ΔCFI (Comparative Fit Index Differences), ΔRMSEA (Root Mean Square Error of Approximation Differences), and ΔSRMR (Standardized Root Mean Square Residual Differences) values can be considered instead of $\Delta\chi^2$ (Chen, 2007; Cheung & Rensvold, 2002; French & Finch, 2006; Meade et al., 2008). When comparing the models, if the ΔCFI value is between -0.01 and +0.01, evidence is obtained that the model with more parameter constraint provides measurement invariance (Cheung & Rensvold, 2002). Chen (2007) states that besides the 0.01 change in ΔCFI value, changes of 0.015 for the ΔRMSEA value and 0.030 for the ΔSRMR value are acceptable in the weak/metric invariance stage, while changes of 0.015 for the ΔRMSEA and ΔSRMR values are acceptable in the scalar/strong invariance and strict invariance stages. Considering all these reasons, in this study, while comparing the invariance models, ΔCFI, ΔRMSEA, and ΔSRMR values were also considered, in addition to the $\Delta\chi^2$ value. This study considered that at least two of the difference tests were within the desired criteria while deciding that models with measurement invariance were provided. MG-CFA results by gender for "Sense of School Belonging" scale are presented in Table 4.

Table 4

MG-CFA Results by Gender for "Sense of School Belonging" Scale

	Structural Invariance	Metric Invariance	Scalar Invariance	Strict Invariance
χ^2	$p > .05$	$p > .05$	$p > .05$	$p < .05$
CFI	1.000	1.000	1.000	0.991
TLI	0.999	1.000	1.000	0.992
RMSEA	0.009	0.002	0.000	0.024
	*(0.000, 0.026)	*(0.000, 0.022)	*(0.000, 0.018)	*(0.015, 0.033)
SRMR	0.015	0.017	0.018	0.039
ΔCFI	-	0.000	0.000	-0.009
ΔRMSEA	-	-0.006	-0.002	0.024
ΔSRMR	-	0.002	0.001	0.021
$\Delta\chi^2$	-	$p > .05$	$p > .05$	$p < .05$

*Lower and upper confidence interval for RMSEA

According to Table 4, all the goodness-of-fit values calculated for the structural invariance model indicate a good fit. For this reason, "Sense of School Belonging" scale provides the structural invariance model. All the goodness-of-fit values calculated for the weak/metric invariance model indicate a good fit. The $\Delta\chi^2$, ΔCFI, ΔRMSEA, and ΔSRMR values between the structural and the weak/metric invariance models are all within the benchmark values. For this reason, "Sense of School Belonging" scale provides the weak/metric invariance model. All the goodness-of-fit values calculated for the strong/scalar invariance model indicate a good fit. The $\Delta\chi^2$, ΔCFI, ΔRMSEA, and ΔSRMR values between the weak/metric and the strong/scalar invariance models are all within the benchmark values. For this reason, "Sense of School Belonging" scale provides a strong/scalar invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the strict invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. Only the ΔCFI value is among the benchmark values for the difference tests between the strong/scale and strict invariance models. $\Delta\chi^2$, ΔRMSEA, and ΔSRMR values are outside the criterion values. In this study, "Sense of School Belonging" scale does not provide the strict invariance model since at least two of the difference tests were determined as a prerequisite for obtaining evidence of measurement invariance. MG-CFA results by gender for "Students Like Learning Mathematics" scale are presented in Table 5.

Table 5

MG-CFA Results by Gender for "Students Like Learning Mathematics" Scale

	Structural Invariance	Metric Invariance	Scalar Invariance	Strict Invariance
χ^2	$p < .05$	$p < .05$	$p < .05$	$p < .05$
CFI	0.998	0.998	0.997	0.997
TLI	0.998	0.997	0.997	0.997
RMSEA	0.027	0.29	0.030	0.030
	*(0.021, 0.033)	*(0.023, 0.034)	*(0.025, 0.036)	*(0.025, 0.035)
SRMR	0.026	0.029	0.031	0.033
ΔCFI	-	-0.001	-0.001	-0.009
ΔRMSEA	-	0.002	0.002	0.000
ΔSRMR	-	0.003	0.003	0.001
$\Delta\chi^2$	-	$p < .05$	$p < .05$	$p < .05$

*Lower and upper confidence interval for RMSEA

According to Table 5, the χ^2 value, one of the goodness-of-fit values calculated for the structural invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. For this reason, "Students Like Learning Mathematics" scale provides the structural invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the weak/metric invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the structural and weak/metric invariance models are among the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criteria values, "Students Like Learning Mathematics" scale provides the weak/metric invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the strong/scalar invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the weak/metric and the strong/scalar invariance models are among the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criteria values, "Students Like Learning Mathematics" scale provides a strong/scalar invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the strict invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the strong/scale and strict invariance models are within the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criteria values, "Students Like Learning Mathematics" scale provides the strict invariance model. MG-CFA results by gender for "Students Value Mathematics" scale are presented in Table 6.

Table 6

MG-CFA Results by Gender for "Students Value Mathematics" Scale

	Structural Invariance	Metric Invariance	Scalar Invariance	Strict Invariance
χ^2	$p< .05$	$p< .05$	$p< .05$	$p< .05$
CFI	0.995	0.989	0.986	0.983
TLI	0.992	0.986	0.985	0.984
RMSEA	0.030	0.040	0.041	0.042
	*(0.024, 0.036)	*(0.035, 0.045)	*(0.036, 0.046)	*(0.038, 0.047)
SRMR	0.033	0.044	0.046	0.056
ΔCFI	-	-0.006	-0.003	-0.003
ΔRMSEA	-	0.010	0.002	0.001
ΔSRMR	-	0.011	0.002	0.010
$\Delta\chi^2$	-	$p< .05$	$p< .05$	$p< .05$

*Lower and upper confidence interval for RMSEA

According to Table 6, the χ^2 value, one of the goodness-of-fit values calculated for the structural invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. For this reason, "Students Value Mathematics" scale provides the structural invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the weak/metric invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the structural and the weak/metric invariance models are among the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criteria values, "Students Value Mathematics" scale provides the weak/metric invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the strong/scalar invariance model, indicates an unacceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the weak/metric and the strong/scalar invariance models are among the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criterion values, it can be said that "Students Value Mathematics" scale provides a strong/scalar invariance model. The χ^2 value, one of the goodness-of-fit values calculated for the strict invariance model, indicates an unacceptable fit. SRMR value indicates an acceptable fit. Other goodness of fit values indicate a good fit. The ΔCFI, ΔRMSEA, and ΔSRMR values between the strong/scale and strict invariance models are within the benchmark values. Only the $\Delta\chi^2$ value was found to be statistically significant. Since three of the four difference tests are among the criteria values, “Students Value Mathematics” scale provides the strict invariance model.

Discussion and Conclusion

This study aimed to examine the measurement invariance of affective scales in the TIMSS 2019 Turkey 8th-grade mathematics student questionnaire in the context of gender. For this purpose, the validity of the relevant affective structures for the Turkish sample was tested by performing CFA separately for the four scales. Then, to determine the measurement invariance, MG-CFA was performed according to gender in the scales confirmed by CFA.

“Students Confident in Mathematics” scale could not be verified by CFA. However, many studies in the literature examining the effect of self-confidence on success in mathematics (Akyüz-Aru, 2020; Akyüz & Pala, 2010; Atar, 2011; Aydın, 2015; Demir et al., 2010; Ertürk & Erdinç-Akan, 2018a; Işlak, 2020; Khine et al., 2015; Koç, 2019; Oral & McGivney, 2013; Usta & Demirtaşlı, 2018). It can be said that “Students Confident in Mathematics” scale is not valid for the mathematics data of the 8th-grade sample of TIMSS Turkey. For this reason, the validity of the results of studies in which this scale will be used in the data of the TIMSS 2019 Turkey 8th-grade mathematics sample in the future will also be low. The measurement invariance of “Students Confident in Mathematics” scale was not examined within the scope of this study since it is not statistically significant to perform MG-CFA analyses of a structure that CFA cannot verify. As a matter of fact, in some of the studies testing the measurement invariance of self-confidence in mathematics, measurement invariance is ensured (Bofah & Hannula, 2015; Cakici-Eser, 2021; Polat, 2019; Uyar, 2021), while measurement invariance cannot be achieved in some (Ertürk & Erdinç-Akan, 2018b). The finding of this study shows parallelism with studies that cannot provide measurement invariance.

“Sense of School Belonging” scale was validated by CFA. For this reason, the sense of belonging to the school is valid for the mathematics data of the 8th-grade sample of TIMSS Turkey. Although there are many studies (Akyüz & Pala, 2010; Akyüz & Satıcı, 2013; Işlak, 2020; Koç, 2019; Sarı et al., 2017; Sarıer, 2020) examining the effect of belonging to school on mathematics achievement in the literature, no study of measurement invariance of this affective variable was found. It is crucial to test the measurement invariance of belonging to the school, whose effect on mathematics achievement is the subject of research. As a result of the MG-CFA for this scale, evidence could be obtained that the scale provided strong/scalar invariance but no evidence that it provided strict invariance. Since this scale provides strong/scalar invariance, the factor score is zero in gender subgroups, while the regression constants are equal. The mean scores on the factor and observed variables are comparable. The differences between the mean scores of the subgroups arise from the latent variable (Başusta & Gelbal, 2015). As a result, it can be said that this scale provides measurement invariance according to gender, there is no bias in the items according to gender, and the mean scores are comparable according to gender.

“Students Like Learning Mathematics” scale was validated by CFA. For this reason, liking mathematics is valid for the mathematics data of the 8th-grade sample of TIMSS Turkey. In the literature, there are many studies (Erşan, 2016; Ertürk & Erdinç-Akan, 2018a; Khine et al., 2015; Koç, 2019; Oral & McGivney, 2013) examining the effect of liking mathematics on success in mathematics. It is crucial to test the measurement invariance of the affective variable of liking mathematics, whose effect on mathematics achievement is the subject of research. As a result of the MG-CFA conducted for this scale, evidence was obtained that the scale provides strict invariance. Since strict invariance is provided in this scale, it was concluded that the error variances for the measured items were equal in gender groups (Widaman & Reise, 1997). In the literature, in the studies in which the measurement invariance of the affective variable of liking mathematics was tested (Bofah & Hannula,2015; Cakici-Eser, 2021; Ertürk & Erdinç-Akan, 2018b; Polat, 2019; Shukla & Konold, 2014), evidence was obtained regarding the measurement invariance. There was no study in which measurement invariance could not be achieved in the measurement invariance studies conducted with the affective variable of liking mathematics. The finding of this study is in parallel with the studies in the literature. As a result, it can be said that this

scale provides measurement invariance according to gender, there is no bias in the items according to gender, and the mean scores are comparable according to gender.
“Students Value Mathematics” scale was validated by CFA. For this reason, valuing mathematics is valid for the mathematics data of the 8th-grade sample of TIMSS Turkey. In the literature, studies (Doğan & Barış, 2010; Khine et al., 2015) examine the effect of valuing mathematics on mathematics achievement. It is crucial to test the measurement invariance of the affective variable of valuing mathematics, whose effect on mathematics achievement is the subject of research. As a result of the MG-CFA conducted for this scale, evidence was obtained that the scale provides strict invariance. Since strict invariance is provided in this scale, it was concluded that the error variances for the measured items were equal in gender groups (Widaman & Reise, 1997). In the literature, in studies where the measurement invariance of the affective variable of valuing mathematics was tested (Bofah & Hannula, 2015; Polat, 2019; Uyar, 2021), evidence was obtained that the measurement invariance was provided. In the studies of measurement invariance conducted with the affective variable of valuing mathematics, no study was found in which measurement invariance could not be achieved. The finding of this study is in parallel with the studies in the literature. As a result, it can be said that this scale provides measurement invariance according to gender, there is no bias in the items according to gender, and the mean scores are comparable according to gender.
The study's large sample size affects the χ^2 goodness-of-fit index used in CFA and $\Delta\chi^2$ values used to compare the differences between models in MG-CFA. In future studies, the effect of sample size can be reduced by choosing a smaller sample than the entire TIMSS sample. In this study, the structure of “Students Confident in Mathematics” scale could not be confirmed by CFA. In future research, it is recommended that researchers approach this scale with caution. In this study, only measurement invariance was examined. Measurement invariance can reveal whether there are biases in terms of items in subgroups. However, it does not reveal which items have biases. In future research, it can be revealed from which items the biases originate by examining the Differential Item Functioning (DIF) based on IRT for “Students Confident in Mathematics” scale.

Declarations

Author Contribution: Mehmet Atılgan: Conceptualization, methodology, analysis, writing & editing, visualization. Kaan Zülfikar Deniz: Conceptualization, methodology, writing-review & editing, supervision.

Conflict of Interest: No potential conflict of interest was reported by the authors.

Ethical Approval: The study was ethically approved by the Uşak University Social and Humanities Scientific Research and Publication Ethics Committee (decision number: 2022-66, dated 14/04/2022).

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript

Consent to Participate: All authors have given their consent to participate in submitting this manuscript to this journal.

Consent to Publish: Written consent was sought from each author to publish the manuscript.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

References

Akyüz-Aru, S. (2020). *4. sınıf öğrencilerinin fen ve matematik başarısına etki eden değişkenlerin incelenmesi “TIMSS 2015 durum analizi”* [Investigation of variables affecting science and mathematics success of grade 4 students "TIMSS 2015 status analysis"] [Unpublished doctoral dissertation]. Gazi Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Akyüz, G., & Pala, N. M. (2010). PISA 2003 sonuçlarına göre öğrenci ve sınıf özelliklerinin matematik okuryazarlığına ve problem çözme becerilerine etkisi [The effect of student and class characteristics on mathematics literacy and problem solving in PISA 2003]. *İlköğretim Online, 9*(2), 668-678.

Akyüz, G., & Satıcı, K. (2013). PISA 2003 verilerine göre matematik okuryazarlığının çeşitli değişkenler açısından incelenmesi: Türkiye ve Hong Kong-Çin modelleri [Investigation of the factors affecting mathematics literacy using PISA 2003 results: Turkey and Hong Kong-China]. *Kastamonu Üniversitesi Kastamonu Eğitim Dergisi, 21*(2), 503 - 522.

Atar, B. (2011). Tanımlayıcı ve Açıklayıcı Madde Tepki Modellerinin TIMSS 2007 Türkiye Matematik Verisine Uyarlanması [An application of descriptive and explanatory ıtem response models to TIMSS 2007 Turkey mathematics data]. *Eğitim ve Bilim, 36*(159), 255 - 269.

Aydın, M. (2015). *Öğrenci ve okul kaynaklı faktörlerin TIMSS matematik başarısına etkisi* [The effects of student-level and school-level factors on middle school students' mathematics achievement] [Unpublished doctoral dissertation]. Necmettin Erbakan Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Başusta, N. B., & Gelbal, S. (2015). Gruplar arası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği [Examination of Measurement Invariance at Groups' Comparisons: A Study on PISA Student Questionnaire]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 30*(4), 80-90.

Bloom, B. S. (2012). *İnsan nitelikleri ve okulda öğrenme* [Human characteristics and school learning] (D. A. Özçelik, Trans.). Pegem Akademi.

Bofah, E.At., & 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*(4), 1-36. https://doi.org/10.1186/s40536-015-0014-y

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, application and programming*. Lawrence Erlbaum.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155–175. https://doi.org/10.1177/0022022102250225

Cakici-Eser, D. (2021). Investigation of measurement ınvariance according to home resources: TIMSS 2015 mathematical affective characteristics questionnaire. *International Journal of Assessment Tools in Education, 8*(3), 633-648. https://doi.org/10.21449/ijate.817168

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

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling, 9*(2), 233-255.

Cheung, G., W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-cultural Psychology, 31*(2), 188–213. https://doi.org/10.1177/0022022100031002003

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları* (5. Baskı) [Multivariate statistics SPSS and LISREL applications for social sciences (5th ed.)]. Pegem Akademi.

Demir, E. (2017). Testing measurement invariance of the students' affective characteristics model across gender sub-groups. *Educational Sciences:Theory and Practice, 17*(1), 47–62. https://doi.org/10.12738/estp.2017.1.0223

Demir, İ., Kılıç, S., & Ünal, H. (2010). Effects of students' and schools' characteristics on mathematics achievement: Findings from PISA 2006. *Procedia - Social and Behavioral Sciences, 2*(2), 3099-3103. https://doi.org/10.1016/j.sbspro.2010.03.472

Demir, M. C. (2020). *TIMSS 2015 fen duyuşsal özelliklerinin cinsiyet ve bölgelere göre incelenmesi* [An examination of TIMMS 2015 science affective factors with regard to gender and regions] [Unpublished master's dissertation]. Hacettepe Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Doğan, N., & Barış, F. (2010). Tutum, değer ve özyeterlik değişkenlerinin TIMSS-1999 ve TIMSS-2007 Sınavlarında öğrencilerin matematik başarılarını yordama düzeyleri [Prediction levels of attitude, value and self-efficacy variables for students' mathematics achievement in TIMSS-1999 and TIMSS-2007 Exams]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(1), 44-50.

Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing, 5*(1), 23-35. https://doi.org/10.1207/s15327574ijt0501_3

Erşan, Ö. (2016). *TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarılarını etkileyen faktörlerin çok düzeyli yapısal eşitlik modeliyle incelenmesi* [Investigation of the factors affecting mathematics achievement of TIMSS 2011 eighth grade students with multilevel structural equation modeling] [Unpublished master's dissertation]. Hacettepe Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Ertürk, Z., & Erdinç-Akan, O. (2018a). TIMSS 2015 matematik başarısını etkileyen değişkenlerin yapısal eşitlik modeli ile incelenmesi [The Investigation of the Variables Effecting TIMSS 2015 Mathematics

Achievement with SEM]. *Ulusal Eğitim Akademisi Dergisi (UEAD), 2*(2), 14-34. https://doi.org/10.32960/uead.407078

Ertürk, Z., & Erdinç-Akan, O. (2018b). 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]. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science]*, UBEK-2018, 204-226. https://doi.org/10.30831/akukeg.412604

Fraenkel, J. R., & Wallen, N.E. (2006). *How to design and evaluate research in education*. McGraw-Hill.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement ınvariance. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(3), 378–402. https://doi.org/10.1207/s15328007sem1303_3

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*(11), 78-94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Gungor, M., & Atalay-Kabasakal, K. (2020). Investigation of measurement ınvariance 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

Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan Türk öğrencilerin duyuşsal özeliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesi [An investigation of measurement invariance by gender for the Turkish students' affective characteristics who took the PISA 2012 math test]. *Gazi Eğitim Fakültesi Dergisi, 37*(1), 151-175.

Hair Jr, J. F., Black, C. W., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education.

Hanci, A. (2015). *8. sınıf öğrencilerinin öğrenme stilleri ve TIMSS matematik başarılarının farklı değişkenler açısından incelenmesi: Bayburt ili örneği* [Investigation of 8th grade students' learning styles and TIMSS matematics achivements from the aspect of different variable: Bayburt sample] [Unpublished master's dissertation]. Bayburt Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Harrington, D. (2009). *Confirmatory factor analysis*. Oxforda University Press, Inc.

Işlak, O. (2020). *TIMSS 2015 uygulamasına katılan öğrencilerin matematik başarılarının öğrenci, aile ve okul değişkenlerine göre yordanma* [Prediction of mathematics achievement of students attending TIMSS 2015 according to student, family and school variables] [Unpublished doctoral dissertation]. Burdur Mehmet Akif Ersoy Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

İlhan, M., & Öner-Sünkür, M. (2012). Matematik kaygısı ile olumlu ve olumsuz mükemmeliyetçiliğin matematik başarısını yordama gücü [The predictive power of mathematics anxiety and positive and negative perfectionism on math achievement]. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 8*(1), 178-188.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). semTools: Useful tools for structural equation modeling. *R package version 0.5-5*. Retrieved from https://CRAN.R-project.org/package=semTools

Jung, J. Y. (2019). *A Comparison of CFA and ESEM approaches using TIMSS science attitudes items: Evidence from factor structure and measurement invariance* [Unpublished master's dissertation]. Purdue University.

Karakoc-Alatli, B., Ayan, C., Polat-Demir, B., & Uzun, G. (2016). Examination of the TIMSS 2011 fourth grade mathematics test in terms of cross-cultural measurement invariance. *Eurasian Journal of Educational Research, 66*, 389-406. https://doi.org/10.14689/ejer.2016.66.22

Karasar, N (2011). *Bilimsel Araştırma Yöntemi* [Scientific Research Method]. Nobel Yayıncılık.

Kesici, A. (2018). Lise öğrencilerinin matematik motivasyonunun matematik başarısına etkisinin incelenmesi. *OMÜ Eğitim Fakültesi Dergisi, 37*(2), 177-194. https://doi.org/10.7822/omuefd.438550

Kesici, A., & Aşılıoğlu, B. (2017). Ortaokul öğrencilerinin matematiğe yönelik duyuşsal özellikleri ile temel eğitimden ortaöğretime geçiş (TEOG) sınavları öncesi yaşadıkları stresin matematik başarısına etkisi [The Effect of Secondary Students' Affective Features Towards Mathematics and The Stress They Experience Before The TEOG Exam (The Exam For Accessing to Various Types of High Schools) on Their Mathematical Success]. *Kırşehir Eğitim Fakültesi Dergisi, 18*(3), 394-414.

Khine, M. S., Al-Mutawah, M., & Afari, E. (2015). Determinants of affective factors in mathematics achievement: Structural equation modeling approach. *Journal of Studies in Education, 5*(2), 199–211.

Kıbrıslıoğlu, N. (2015). *PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye – Çin (Şangay) – Endonezya örneği* [The investigation of measurement invariance PISA 2012 mathematics learning model according to culture and gender: Turkey - China (Shangai) - Indonesia] [Unpublished master's dissertation]. Hacettepe Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Kilic, S. & Askin, Ö. E. (2013). Parental influence on students' mathematics achievement: the comparative study of Turkey and best performer countries in TIMSS 2011. *Procedia - Social and Behavioral Sciences, 106*, 2000-2007. https://doi.org/10.1016/j.sbspro.2013.12.228

Kline, R. B. (2015). *Principles and practices of structural equation modeling* (4th ed.). The Guilford Press.

Koç, O. (2019). *4. ve 8. sınıf öğrencilerinin TIMSS 2015 matematik başarısını yordayan değişkenlerin belirlenmesi* [Determination of predictive variables of 4th and 8th grade students' on TIMSS 2015 mathematics achievement] [Unpublished master's dissertation]. Akdeniz Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Lay, Y. F., Ng, K. T., & Chong, P. S. (2015). Analyzing affective factors related to eighth grade learners' science and mathematics achievement in TIMSS 2007. *Asia-Pacific Education Researcher, 24*(1), 103–110. https://doi.org/10.1007/s40299-013-0163-0

Louis, R.A., & Mistele, J.M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*, 1163–1190. https://doi.org/10.1007/s10763-011-9325-9

Ma, Y., & Qin, X. (2021). Measurement invariance of information, communication and technology (ICT) engagement and its relationship with student academic literacy: Evidence from PISA 2018. *Studies in Educational Evaluation, 68*, 1–15. https://doi.org/10.1016/j.stueduc.2021.100982

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Ministry of National Education (MoNE) (2020). *TIMSS 2019 Türkiye ön raporu* [TIMSS 2019 Turkey preliminary report]. Ankara: Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.

Ministry of National Education (MoNE). (2016). *TIMSS 2015 ulusal matematik ve fen ön raporu: 4. ve 8. sınıflar* [TIMSS 2015 national math and science preliminary report: 4th and 8th grades]. Ankara.

Meng, L., Qiu, C., & Boyd-Wilson, B. (2019). Measurement invariance of the ICT engagement construct and its association with students' performance in China and Germany: Evidence from PISA 2015 data. *British Journal of Educational Technology, 50*(6), 3233–3251. https://doi.org/10.1111/bjet.12729

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). The Guilford Press.

Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society, 1*(1), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010

Mohammadpour, E. (2012). Factors accounting for mathematics achievement of singaporean eighth-graders. *The Asia-Pacific Education Researcher, 21*(3), 507-518.

Mulaik, S. A. (2007). There is a place for approximate fit in structural equation modelling. *Personality and Individual Differences, 42*(5), 883-891. https://doi.org/10.1016/j.paid.2006.10.024

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Boston, US.

Oral, I., & McGivney, E. (2013). *Türkiye'de matematik ve fen bilimleri alanlarında öğrenci performansı ve başarının belirleyicileri: TIMSS 2011 analizi* [Student performance and determinants of success in mathematics and science in Turkey: TIMSS 2011 analysis]. İstanbul: Eğitim Reformu Girişimi Raporu.

Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin incelenmesi: Türkiye-Amerika Birleşik Devletleri örneği* [The investigation of psychometric properties of the test of progress in international reading literacy (PIRLS) 2001: The model of Turkey-United States of America] [Unpublished doctoral dissertation]. Hacettepe Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Ölçüoğlu, R., & Çetin, S. (2016). TIMSS 2011 sekizinci sınıf öğrencilerinin matematik başarısını etkileyen değişkenlerinin bölgelere göre incelenmesi [The investigation of the variables that affecting eight grade students' TIMSS 2011 math achievement according to regions]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 7*(1), 202-220. https://doi.org/10.21031/epod.34424

Öncü, Ö. (2019). *TIMSS 2015 sekizinci sınıf matematik başarı testinin oecd ülkelerine göre ölçme değişmezliğinin incelenmesi* [An investication into the measurement invariance according to OECD countries of TIMSS 2015 eight grade math achievement test] [Unpublished master's dissertation]. Akdeniz Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Patterson, M., Perry, E., Decker, C., Eckert, R., Klaus, S., Wendling, L., & Papanastasiou, E. (2003). Factors associated with high school mathematics performance in the United States. *Studies in Educational Evaluation, 29*(2), 91-108. https://doi.org/10.1016/S0191-491X(03)00017-8

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (6th Ed.). Routledge.

Polat, M. (2019). *TIMSS-2015 matematik ve fen duyuşsal özellik modellerinin kültürlere, cinsiyete ve bölgelere göre ölçme değişmezliğinin incelenmesi* [The investigation of measurement invariance of TIMSS-2015 mathematics and science affective characteristics models according to culture, gender and statistical region] [Unpublished master's dissertation]. Hacettepe Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education, 8*(3), 259–278. https://doi.org/10.2304/rcie.2013.8.3.259

Salzberger, T., Sinkovics, R., R., & Schlegelmilch, B. B. (1999). Data equivalence in cross-cultural research: a comparison of classical test theory and latent trait theory based approaches. *Australasian Marketing Journal, 7*(2), 23-38. https://doi.org/10.1016/S1441-3582(99)70213-2

Sarı, M. H., & Ekici, G. (2018). İlkokul 4. sınıf öğrencilerinin matematik başarıları ile aritmetik performanslarını etkileyen duyuşsal değişkenlerin belirlenmesi [Determination of affective variables affecting mathematical achievement and arithmetic performance of primary school 4th grade students]. *OPUS International Journal of Society Researches, 8*(15), 1562-1594. https://doi.org/10.26466/opus.451025

Sarı, M. H., Arıkan, S., & Yıldızlı, H. (2017). 8. sınıf matematik akademik başarısını yordayan faktörler-TIMSS 2015 [Factors predicting mathematics achievement of 8th graders in TIMSS 2015]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 8*(3), 246-265.

Sarıer, Y. (2020). TIMSS uygulamalarında Türkiye'nin performansı ve akademik başarıyı yordayan değişkenler [Turkey's performance in TIMSS applications and variables predicting academic achievement]. *Temel Eğitim Dergisi, 2*(2), 6-27.

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7*(110), 1-16. https://doi.org/10.3389/fpsyg.2016.00110

Schumacker, R. E., & Beyerlein, S. T. (2000). Confirmatory factor analysis with different correlation types and estimation methods. *Structural Equation Modeling, 7*(4), 629–636. https://doi.org/10.1207/S15328007SEM0704_6

Shukla, K., & Konold, T. (2014). *Fondness of math and science as measured by the TIMSS student questionnaire: Invariance across U.S. ethnic groups*. Paper presented at the 2014 Annual Meeting of the American Educational Research Association, Philadelphia, USA

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with comfirmatory factor analysis and ıtem response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292–1306. https://doi.org/10.1037/0021-9010.91.6.1292

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Tavlıca, A. (2019). *TIMSS 2015 dördüncü sınıf matematik testinin ölçme değişmezliğinin ülkelere göre incelemesi* [An investigation of measurement invariance for TIMSS 2015 fourth grade mathematics test according to countries] [Unpublished master's dissertation]. Akdeniz Üniversitesi. https://tez.yok.gov.tr/UlusalTezMerkezi/

Usta, H. G., & Demirtaşlı, R. N. (2018). PISA 2012 matematik okuryazarlığı üzerine uluslararası bir karşılaştırma: Türkiye ve Finlandiya [An international comparison according to PISA 2012 mathematical literacy Turkey and Finland]. *Turkish Studies (Elektronik), 13*(11), 1389 - 1420. https://doi.org/10.7827/TurkishStudies.13377

Uyar, S. (2021). Factor structure and measurement invariance of the TIMSS 2015 mathematics attitude questionnaire: Exploratory structural equation modelling approach. *International Journal of Assessment Tools in Education, 8*(4), 855–871. https://doi.org/10.21449/ijate.796862

Uzun, B., & Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin TIMSS-R Türkiye örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi [Assessing the measurement ınvariance of factors that are related to students' science achievement across gender in TIMSS-R Turkey sample]. *Eğitim ve Bilim, 35*(155), 26–35.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the MI literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–69. https://doi.org/10.1177/109442810031002

Wang, Z., Osterlind, S., & Bergin, D. A. (2012). Building mathematics achievement models in four countries using TIMSS 2003. *International Journal of Science and Mathematics Education, 10*(5), 1-28. https://doi.org/10.1007/s10763-011-9328-6

Webster, B. J., & Fisher, D. L. (2000). Accounting for variation in science and mathematics achievement: A multilevel analysis of Australian data third international mathematics and science study (TIMSS). *School Effectiveness and School Improvement, 11*(3), 339–360. https://doi.org/10.1076/0924-3453(200009)11:3;1-G;FT339

Wicherts, J. M. (2007). *Group differences in intelligence test performance* [Unpublished doctoral dissertation]. University of Amsterdam.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association.

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 and Evaluation, 12*, 1-26. https://doi.org/10.7275/mhqa-cd89

Yandı, A., Köse, İ. A., & Uysal, Ö. (2017). Farklı yöntemlerle ölçme değişmezliğinin incelenmesi: PISA 2012 örneği [Examining measurement invariance with different methods: Example of PISA 2012]. *Mersin Üniversitesi Eğitim Bilimleri Dergisi, 13*(1), 243-253. https://doi.org/10.17860/mersinefd.305952

Yin, L., & Fishbein, B. (2020). Creating and interpreting the TIMSS 2019 context questionnaire scales. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 16.1-16.331). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/methods/chapter-16.html

Yücel, Z., & Koç, M. (2011). İlköğretim öğrencilerinin matematik dersine karşı tutumlarının başarı düzeylerini yordama gücü ile cinsiyet arasındaki ilişki [The relationship between the prediction level of elementary school students' math achievement by their math attitudes and gender]. *İlköğretim Online, 10*(1), 133-143.