Journal of Theoretical Educational Science, 17(1), 1-27, January 2024
Kuramsal Eğitimbilim Dergisi, 17(1), 1-27, Ocak 2024
[Online]: http://dergipark.org.tr/akukeg
DOI number: http://doi.org/10.30831/akukeg.1207350

Cross-National Measurement of Mathematics Intrinsic Motivation: An Investigation of Measurement Invariance with MG-CFA and the Alignment Method Across Fourteen Countries

Matematikte İçsel Motivasyonun Ülkeler Arası Ölçümü: On Dört Ülkede MG-CFA ve Hizalama Yöntemi ile Ölçme Değişmezliğinin İncelenmesi

Mahmut Sami YİĞİTER*

Received: 19 November 2022 **Research Article** **Accepted:** 27 November 2023

ABSTRACT: One of the main objectives of international large-scale assessments is to make comparisons between different countries, education policies, education systems, or subgroups. One of the main criteria for making comparisons between different groups is to ensure measurement invariance. The purpose of this study was to test the measurement invariance of the mathematics intrinsic motivation scale across 14 countries. For this purpose, the "students like learning mathematics" scale, which measures intrinsic motivation for mathematics, was included in the TIMSS 2019 cycle. The study sample consisted of a total of 152992 students, 70192 4th grade and 82800 8th grade students from 14 different countries participating in the TIMSS 2019 cycle. Measurement invariance was tested with Multi-Group Confirmatory Factor Analysis (MG-CFA) and Alignment Method. The mathematics intrinsic motivation scale provides only configural invariance according to MG-CFA at the 4th grade level, whereas the scale provides approximate invariance according to the alignment method. At the 8th grade level, the scale provides configural and metric invariance according to MG-CFA, whereas the scale provides approximate invariance according to the alignment method. The results indicate that the mathematics intrinsic motivation scale provides approximate measurement invariance at both grade levels and that comparisons can be made between the scores of the identified countries.

Keywords: Mathematics, motivation, intrinsic motivation, measurement invariance, cross-national difference, students like learning mathematics, alignment method.

ÖZ: Geniş ölçekli uluslararası değerlendirmelerin temel amaçlarından biri, farklı ülkeler, eğitim politikaları, eğitim sistemleri veya alt gruplar arasında karşılaştırmalar yapmaktır. Farklı gruplar arasında karşılaştırma yapmanın temel ölçütlerinden biri de ölçme değişmezliğinin sağlanmasıdır. Bu çalışmanın amacı, matematik içsel motivasyon ölçeğinin 14 ülke arasında ölçme değişmezliğini test etmektir. Bu amaçla, matematiğe yönelik içsel motivasyonu ölçen "öğrenciler matematik öğrenmeyi sever" ölçeği TIMSS 2019 döngüsüne dahil edilmiştir. Çalışmanın örneklemi TIMSS 2019 döngüsüne katılan 14 farklı ülkeden 70192 4. sınıf ve 82800 8. sınıf öğrencisi olmak üzere toplam 152992 öğrenciden oluşmaktadır. Ölçme değişmezliği, Çok Gruplu Doğrulayıcı Faktör Analizi (MG-CFA) ve Hizalama Yöntemi ile test edilmiştir. Matematik içsel motivasyon ölçeği, 4. sınıf düzeyinde MG-CFA'ya göre sadece yapısal değişmezliği sağlarken, hizalama yöntemine göre yaklaşık değişmezliği sağlamaktadır. 8. sınıf düzeyinde ise ölçek, MG-CFA'ya göre konfigüral ve metrik değişmezliği sağlarken, hizalama yöntemine göre yaklaşık değişmezliği sağlamaktadır. Sonuçlar, matematik içsel motivasyon ölçeğinin her iki sınıf düzeyinde de yaklaşık ölçme değişmezliğini sağladığını ve belirlenen ülkelerin puanları arasında karşılaştırmalar yapılabileceğini göstermektedir.

Anahtar kelimeler: Matematik, motivasyon, içsel motivasyon, ölçme değişmezliği, ülkeler arası farklılık, öğrenciler matematik öğrenmeyi sever, hizalama yöntemi.

* Inst. MSc., Social Sciences University of Ankara, Ankara, Turkey, mahmutsamiyigiter@gmail.com, https://orcid.org/0000-0002-2896-0201

Citation Information
Yiğiter, M. S. (2024). Cross-national measurement of mathematics intrinsic motivation: An investigation of measurement invariance with MG-CFA and alignment method across fourteen countries. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science], 17*(1), 1-27.

ISSN: 1308-1659

The basis of effective mathematics teaching is to support positive attitudes towards learning mathematics and to encourage learning mathematics. According to social cognitive theory, an individual's self-efficacy in a subject affects his/her motivation in that subject (Schunk & DiBenedetto, 2020). Then, academic success emerges with the effect of motivation on performance (Yıldırım, 2011). An individual's motivation to accomplish a task enables him/her to orient himself/herself towards that field and to work persistently in that field. Many studies have shown that student motivation and academic achievement are related (Ahmed et al., 2010; Cleary & Chen, 2009; İlter, 2021; Woolley et al., 2010).

International Large-Scale Assessment (ILSA) programmes have been started to be carried out with the participation of many countries since the end of the 20th century in order to make comparisons between the education systems of countries and to determine student achievement (Cardoso, 2020). Trends in International Mathematics and Science Study (TIMSS) is an assessment study that evaluates the academic achievement of 4th and 8th-grade students in the fields of mathematics and science skills every four years and monitors the achievement differences between countries over time and the results of countries' attempts to increase the level of achievement (Mullis & Martin, 2017). In addition to measuring mathematics and science skills, TIMSS includes many scales and scales that measure cognitive and affective characteristics of students and teachers, such as self-confidence, motivation, school belonging, and peer bullying (Yin & Fishbein, 2019). Students' affective characteristics for mathematics achievement have been measured since 1995. One of the scales included in the TIMSS 2019 student survey is the "Students Like Learning Mathematics" scale, which measures mathematics intrinsic motivation. There are many studies reporting that intrinsic motivation has positive and significant effects on mathematics achievement (Akben-Selcuk, 2017; Guo et al., 2015; Hooper et al., 2020). Therefore, it is important that the scales measuring the characteristics of the mathematics intrinsic motivation scale perform valid and reliable measurements in order to accurately and objectively reveal the relationships between intrinsic motivation and mathematics achievement in cross-country and cross-cultural comparisons.

One of the main goals of ILSAs is to make comparisons between different countries, education systems, subgroups, and individuals (Engel & Rutkowski, 2021). There is increasing methodological discussion about the use of data from ILSAs for making comparisons (Gustafsson, 2018; Rutkowski & Svetina, 2014). One of the basic and critical criteria for making comparisons between different groups is the establishment of measurement invariance (Putnick & Bornstein, 2016). Measurement invariance is a statistical property that analyses whether the scale has equivalent psychometric values between the different groups or sub-groups to which it is applied (Raykov, 2004). A measurement tool should measure the construct in a psychometrically equivalent way in each subgroup. If the psychometric properties of the measurements obtained from subgroups differ, it would not be correct to generalize the results (Başusta & Gelbal, 2015). Therefore, a measurement tool should measure the construct equivalently in each subgroup. With measurement invariance, showing that the factor loadings, inter-dimensional correlations, and error variances of a scale are the same in each group will show that the measurement tool has an equivalent structure in different groups (Jöreskog & Sörbom, 1993). Researchers obtain evidence on whether

the scale measures the same construct in subgroups (Millsap & Olivera-Ogilar, 2012; Uyar & Doğan, 2014). Failure to provide measurement invariance is a validity problem for the measurement tool. Therefore, interpretations regarding the results of group comparisons based on such a measurement tool may also be incorrect (Vandenberg & Lance, 2000). Demonstrating that measurement invariance is established will also provide validity evidence for the measurement tool. He et al. (2019), in their study on cross-cultural comparability with TIMSS and PISA data, state that comparisons made without examining measurement invariance may lead to inaccurate results, hence the importance of testing measurement invariance.

Mathematics Intrinsic Motivation

Intrinsic motivation is defined as the enjoyment received while engaging in an activity and the drive to perform this activity (OECD, 2013). Students' willingness to learn mathematics stems from the fact that they find mathematics interesting and fun (Ryan & Deci, 2009). Intrinsically motivated students take action to "learn" rather than doing something for a purpose. Extrinsically motivated students see the task as a means to achieve a goal and complete this task with concepts such as "grade, reward, competition, performance, external evaluation." For example, a student who studies for an exam to get a good grade is motivated (extrinsically) by the grade he/she gets. Students who study only because they enjoy it are intrinsically motivated. Previous research has revealed that students' intrinsic motivation can be positively influenced by the teacher's approach, curriculum, methods applied in the lesson, learning environment design, and practices (Freiberger et al., 2012; Middleton, 1995; Weidinger et al., 2017). Mueller et al. (2011) state that solving open-ended questions and discussing in the classroom environment can increase intrinsic motivation. Similarly, problem-based learning increases students' intrinsic motivation (Henderson & Landesman, 1995). There are studies indicating that applied learning and active learning activities increase intrinsic motivation (Barak & Asad, 2012; Nugent et al., 2010). In addition, when students accomplish a difficult task, their motivation increases (Middleton, 1995).

Intrinsic motivation is a source of energy and a precondition for behaviour (Malone & Lepper, 2021). Previous studies show that there is a strong relationship between intrinsic motivation and mathematics achievement (Hooper et al., 2020; İlhan & Çetin, 2013; Mullis et al., 2017; Tavani & Losh, 2003; Zembat et al., 2018). Intrinsic Motivation for Mathematics is measured with nine items under the "Students Like Learning Mathematics" scale at both 4th and 8th-grade levels in TIMSS 2019.

Measurement Invariance with Multi-Group Confirmatory Factor Analysis

Measurement invariance investigates whether a latent construct is measured consistently across different groups, categories, or times (Cheung & Rensvold, 2002; Sözer et al., 2021). In other words, if the psychometric values obtained from the same scale differ in different groups, it can be said that measurement invariance cannot be provided. Individuals in different groups who are equivalent to each other in terms of the measured feature (construct) are expected to get the same observed score from a test. If the individuals are the same in terms of the construct measured, but their observed scores are different, it can be said that measurement invariance of the scale (test) cannot be ensured (Schmitt & Kuljanin, 2008). If measurement invariance cannot be proved, it is not correct to interpret the results of intergroup comparisons. The reason

for this is that it cannot be known whether the difference between the groups is due to a real construct difference or the difference between the responses to the scale items (Adibatmaz & Yildiz, 2020; Horn & McArdle, 1992). Therefore, it is important to test for measurement invariance before making inferences about measurements from two or more groups (Yiğiter, 2023).

One of the most commonly used methods in the literature to test measurement invariance is Confirmatory Factor Analysis (CFA) (Schmitt & Kuljanin, 2008; Van De Schoot et al., 2015). Measurement invariance is tested in four hierarchical stages with the Multi-Group CFA (MG-CFA) method. These stages are configural invariance, metric invariance, scalar invariance, and strict invariance (Meredith, 1993).

Configural Invariance

It is the first hierarchical stage of measurement invariance. At this stage, whether the groups have the same factor structure is tested. For this purpose, the equivalence of factors and the pattern of factor loadings are analysed at this stage (Taris et al., 1998). No parameter restriction is made at this stage. If configural invariance is provided, it can be stated that the groups measure the same construct (Wu et al., 2007). If configural invariance is not provided, it is stated that the groups measure different constructs, and the further stages of measurement invariance are not passed. Configural invariance is also referred to as structural invariance in the literature.

Metric Invariance

When it is shown that configural invariance is provided, a metric invariance test can be performed (Milfont & Fischer, 2010). In metric invariance, the equality of factor loadings in different groups is tested. In other words, factor loadings estimated from one group are fixed to the other group, and the fit indices of the model are examined. If metric invariance is provided, comparisons between groups based on factor loadings can be defended (Gregorich, 2006). Metric invariance is also known as weak invariance (Meredith, 1993).

Scalar Invariance

If metric invariance is provided, the scalar invariance stage is proceeded. In this stage, the equivalence of both factor loadings and regression constants between groups is tested. In other words, at this stage, where the equivalence of factor variance and covariances between groups is tested, the equivalence of factor loadings is also examined. If scalar invariance is provided, it means that the means and factor loadings of the observed variables can be compared (Gregorich, 2006). Scalar invariance is also known as strong invariance.

Strict Invariance

At this stage, which is the last step of measurement invariance, in addition to the restrictions in the previous stages, the equality of error variances is also tested (Vandenberg & Lance, 2000). Scales that claim to measure the same construct across groups should provide strict invariance. By ensuring strict invariance, measurement invariance will be fully provided.

Measurement Invariance with Multi-Group Alignment Method

In measurement invariance with MG-CFA - especially when the number of groups is large - it becomes difficult to ensure model fit across stages. MG-CFA assumes strict invariance, which may be an unreachable goal when the number of groups increases. When measurement invariance is rejected in MG-CFA, partial models with free estimation of some item parameters can be tested, but there is no guarantee that these models will also provide measurement invariance (Asparouhov & Muthén, 2014). On the other hand, when measurement invariance is rejected with MG-CFA, the causes of invariance are not properly identified by the analysis. Moreover, the probability of incorrect calculation will increase as a result of MG-CFA's pair-by-pair comparison of groups. Therefore, MG-CFA is not practical when comparing the measurement invariance of a large number of groups (Sırgancı et al., 2020).

A more recent approach, the Alignment method, greatly simplifies the measurement invariance analysis. It allows testing the invariance of parameters according to items and groups. In other words, in the alignment method, it can be determined which group contributes to measurement invariance. The alignment method proposed by Asparouhov and Muthén (2014) provides a result that minimises parameter invariance between groups in an iterative process, similar to rotation in exploratory factor analysis (Glassow et al., 2021).

Aim and Significance of the Study

The purpose of this study is to examine the measurement invariance of the mathematics intrinsic motivation scale in the TIMSS 2019 cycle according to 14 countries. When the invariance studies in the literature were examined, it was seen that measurement invariance studies were carried out according to culture, region, language, and gender (Alatlı, 2020; Bağdu Söyler et al., 2021; Ertürk & Erdinç-Akan, 2018; Uyar & Doğan, 2014;). Studies examining measurement invariance according to the mathematics intrinsic motivation scale are quite limited (He et al., 2019). No study was found that tested the measurement invariance of the mathematics intrinsic motivation scale across cultures with the TIMSS 2019 data. This study is important both because it focuses on the cross-cultural measurement invariance of the mathematics intrinsic motivation scale and because it has not been investigated before.

Related Studies

Glasgow et al. (2019) examined the measurement invariance of mathematics teachers' Job satisfaction, School emphasis on academic success, School condition and resources, Safe and orderly school, and teacher Self-efficacy scales obtained from TIMMS 2015 teacher surveys across 46 countries. The results show that only three constructs provide metric invariance. The results of measurement invariance with the Alignment Optimisation method show that all five constructs provide approximate invariance so that these constructs can be validly compared across educational systems.

Sırgancı et al. (2020), in their study explaining the basic concepts and processes of the alignment method, compared the measurement invariance of 56 countries on the Instrumental Motivation Scale data in the PISA 2015 cycle with both MG-CFA and alignment method. MG-CFA findings show that the scale provides only configural invariance. Then, the measurement invariance findings with the alignment method

provide more detailed information about which countries and which items contribute better to measurement invariance.

Tekin and Cobanoglu-Aktan (2021) examined the measurement invariance of collaborative problem-solving skills in the PISA 2015 cycle between Singapore, Norway, and Turkey with the MG-CFA method. The results of the study show that the construct provides only configural invariance but not metric, scalar, and strict invariance.

Method

This study is descriptive research since it aims to determine whether the "Students Like Learning Mathematics" scale from TIMSS 2019, which is included in the TIMSS 2019 student questionnaire and measures mathematics intrinsic motivation, does not change according to 14 different countries (Büyüköztürk et al., 2017).

Population and Sample

More than 580,000 students from 64 countries around the world participated in TIMSS 2019, which was administered by the International Association for the Evaluation of Educational Achievement (IEA). Approximately 330.000 of these students are in Grade 4, and 250.000 are in Grade 8. The sample of this study consists of a total of 152.992 students, 70.192 of whom are 4th graders and 82.800 of whom are 8th graders, who participated in TIMSS 2019 from 14 countries. In determining the sample, countries were determined by taking into account the differentiation in terms of language, continent, culture, and achievement rankings. In addition, since the research was conducted for both the 4th and 8th-grade levels, the countries that participated in TIMSS 2019 at both grade levels were selected for the current study. The distribution of the sampled countries according to their sample sizes, languages, and continents is given in Table 1.

Table 1

Countries in the Sample and Their Characteristics

Country Name	ISO Country Code	Language	Continent	4th Grade	8th Grade
Australia	36	English	Oceania	5664	8898
Chile	152	Spanish	South America	4039	4061
Hungary	348	Hungarian	Europe	4433	4537
Italy	380	Italian	Europe	3666	3600
Japan	392	Japanese	Asia	4162	4443
Morocco	504	Moroccan Arabic and others	Africa	7645	8440
Portugal	620	Portuguese	Europe	4256	3348
Russia	643	Russian	Europe-Asia	3993	3890
Saudi Arabia	682	Arabic	Asia	5334	5634
Singapore	702	English, Malay, and others	Asia	4362	4165
South Africa	710	Afrikaans	Africa	11729	20717

Sweden	752	Swedish	Europe	3816	3907
Turkey	792	Turkish	Asia	3998	3978
Pakistan	926	Urdu	Asia	3095	3182
Total				70192	82800

As seen in Table 1, 14 countries in the sample are located on five different continents. In addition, the official languages of each of these countries are different.

Data Source

The data were obtained from the database at "https://timss2019.org/international-database/". The TIMSS administration includes mathematics and science achievement tests as well as student, teacher, school, and home questionnaires. In addition, there are also items that examine the affective characteristics of students in mathematics and science. This study was limited to the "Students Like Learning Mathematics" scale, which measures intrinsic motivation for mathematics. The items and codes in this scale are given in Table 2.

Table 2

Items and Codes in the Scale

Item Code		Description
4. Grade	8. Grade	
ASBM02A	BSBM16A	I enjoy learning mathematics
ASBM02B	BSBM16B	I wish I did not have to study mathematics [R]
ASBM02C	BSBM16C	Mathematics is boring [R]
ASBM02D	BSBM16D	I learn many interesting things in mathematics
ASBM02E	BSBM16E	I like mathematics
ASBM02F	BSBM16F	I like any schoolwork that involves numbers
ASBM02G	BSBM16G	I like to solve mathematics problems
ASBM02H	BSBM16H	I look forward to mathematics lessons
ASBM02I	BSBM16I	Mathematics is one of my favourite subjects

There are nine items in this scale, as seen in Table 2. Items 2 and 3 are reverse-coded questions. All items were scored on a 4-point Likert rating scale with the options "agree a lot, agree a little, disagree a little, disagree a lot."

Data Analysis

All analyses in this study were performed with the open-source R program. "dplyr" (Wickham et al., 2019) for data manipulation, "lavaan" (Rosseel, 2012) for CFA and MG-CFA analyses, "sirt" (Robitzsch, 2019) for alignment method, "naniar" (Tierney et al., 2021) for missing data analysis, "mvdalab" (Afanador et al., 2016) for missing data imputation, "Performance Analytics" (Peterson et al., 2018) for normality analysis. Grade 4 and Grade 8 data obtained from the TIMSS 2019 database were

analysed separately. Before starting the data analysis, missing data, outliers, normality, and multicollinearity were examined.

4th Grade Level

It is seen that there are a total of 73336 students participating in the exam at the 4th grade level from 14 countries. Firstly, missing data, which are questions not answered by the students, were analysed. It was observed that the rate of missing data on a variable basis varied between 2.5% (1878/73336) and 4.8% (3546/73336). On the basis of all data, the rate of missing data reaches 11.3% (8323/73336). Little MCAR test was performed to examine the randomness of the missing data. The Little MCAR test results show that the missing data is random and does not contain any pattern (LittleMCAR=5091; df=2604; p>0.05). Since the missing data is more than 10%, missing data assignment was preferred instead of the listwise deletion method in order not to lose the analysis power. Before assigning missing data, 1065 participants who left the entire scale used in the study blank were deleted. Then, the missing data assignment was performed with the EM algorithm. The extreme value analysis was performed in two stages. Before the extreme value analysis, participants who answered carelessly were identified. Participants who give the same responses to questions with reverse coding and questions with normal coding in the scale exhibit careless responding behaviour (Woods, 2006). It is stated that these participants have disruptive effects on the factor structure (Kam, 2019). In the 4th grade data, 1560 participants with careless responding behaviour were identified and excluded from the sample (Kam & Meyer, 2015). In the second stage, total and standardised Z scores were calculated according to the scales. According to the z score, participants who were outside the [-3,+3] range were determined as outliers (Kaliyaperumal et al., 2015). According to the Z scores, 519 observations outside this range were identified as outliers and deleted from the data. Skewness and kurtosis values of the variables were analysed to determine whether the data were normally distributed. Since the kurtosis and skewness coefficients were in the range of [-1.5,+1.5], it was decided that the data were normally distributed (Tabachnick & Fidell, 2013). In order to examine the multicollinearity, the VIF (variance inflation factor) value was calculated for all items. The highest VIF value was found to be 3.55 in the item coded BSBM16E. It was decided that there was no multicollinearity problem since a multicollinearity problem would occur if the VIF value was greater than 5 (Kline, 2011).

8th Grade Level

It is seen that there are 84345 students from 14 countries who participated in the exam at the 8th-grade level. Firstly, missing data, which are questions not answered by the students, were analysed. It was observed that the missing data rates on a variable basis varied between 1.5% (1275/84345) and 3.7% (3142/84345). On the basis of all data, the missing data rate reaches 8.21% (6927/84345). Little MCAR test was performed to examine the randomness of missing data. The Little MCAR test results show that the missing data is random and does not contain any pattern (LittleMCAR=3.179; sd=1701; p>0.05). Before assigning missing data, 860 participants who left the entire scale used in the study blank were deleted. Missing data assignment was made with the EM algorithm. The extreme value analysis was performed in two

stages. Before the extreme value analysis, participants who carelessly responded were identified. In this data, 685 participants who responded carelessly were identified and removed from the sample (Kam & Meyer, 2015). Since there was no value outside the range of [-3,+3] according to Z scores, it was decided that there was no outlier (Kaliyaperumal et al. , 2015). The fact that the kurtosis and skewness coefficients are in the range of [-1.5,+1.5] indicates that the data are normally distributed (Tabachnick & Fidell, 2013). The highest VIF value calculated to examine the multicollinearity problem was found to be 4.37 in the item coded BSBM16E. It was decided that there was no multicollinearity problem since a multicollinearity problem would occur if the VIF value was greater than 5 (Kline, 2011).

Multi-Group CFA analyses were performed using the lavaan package in R (Rosseel, 2012). Another reason why this package was preferred is that it allows the use of sample weights when estimating with the MG-CFA model. In large-scale assessments, all participants in the population (all students at the relevant grade level) cannot be included in the sample due to time and financial limitations. In order to overcome this limitation and to ensure the generalisability of the sample to the population, sampling weights are used (Arıkan et al., 2020). Student weights in the TIMSS 2019 data were added to the model as sampling weights. In this study, country code as a categorical variable and items of the intrinsic motivation scale as ordinal variables were used. It is recommended to use WLS (weighted least squares), WLSMV (robust weighted least squares), or ULS (unweighted least squares) methods that are robust to violations of assumptions as estimation methods (Brown, 2006; Koğar & Yılmaz Koğar, 2015). Therefore, the WLSMV estimation method, which is reported to give good results in the MG-CFA model, was used as the estimation method (Forero et al., 2009).

Measurement invariance was analysed by testing four hierarchical stages with MG-CFA. These four stages are configural invariance, metric invariance, scalar invariance, and strict invariance (Vandenberg & Lance, 2000). To examine the model-data fit between the stages, χ^2, χ^2/df, $RMSEA$, $SRMR$, TLI, CFI and ΔCFI values were reported. Acceptable levels of these values are presented in Table 3 (Hu & Bentler, 1999).

Table 3

Acceptable Ranges of Goodness of Fit Indices

Fit Indices	Acceptable Fit	Good Fit
χ^2/df	$3 < \chi^2/df < 5$	$0 < \chi^2/df < 3$
CFI	0.95 < CFI < 0.97	0.97 < CFI < 1
TLI	0.95 < TLI < 0.97	0.97 < TLI < 1
RMSEA	0.05 < RMSEA < 0.08	0.00 < RMSEA < 0.05
SRMR	0.05 < SRMR < 0.08	0.00 < SRMR < 0.05

Note. (Hu & Bentler, 1999)

In the MG-CFA method, once it was determined that the fit at a particular stage was satisfactory, the analysis proceeded to the next stage. There are studies suggesting

that the significance can be tested according to the difference of chi-square values ($\Delta\chi^2$) in determining whether inter-stage invariance is achieved (Schmitt & Kuljanin, 2008). However, the chi-square difference test rejects the null hypothesis with too much power as the sample size increases. Therefore, Cheung and Rensvold (2002) suggested examining the change in CFI value (ΔCFI) as an alternative to $\Delta\chi^2$. In this study, a difference of ΔCFI less than or equal to 0.01 was used as a criterion to determine whether inter-stage invariance was achieved (Cheung & Rensvold, 2002).

In measurement invariance with the Alignment Method, the fit of a configural model without restriction between groups was first assessed. Then, it was optimised with a component loss function to minimise the invariance between the means of each factor and the variances of the groups under the configural model (Asparouhov & Muthén, 2014). The tolerance criteria proposed by Robitzsch (2020) were used (factor loadings (λ=.40) and intercepts (ν=.20)). The alignment strength for the parameters was determined as .25 (Fischer & Karl, 2019). The equivalence of the parameters was interpreted with the R^2 value. R^2 values close to 1 indicate that there is more invariance (Asparouhov & Muthén, 2014). In determining the measurement invariance, the cut-off criterion of 25% of the invariant parameter ratio of the intercept and slope parameters was used (Asparouhov & Muthén, 2014).

Results

In this section, the findings obtained from the mathematics intrinsic motivation scale for 4th and 8th grade levels according to TIMSS 2019 data are presented. Analyses were conducted separately for both grade levels. Firstly, CFA was conducted. Finally, measurement invariance was tested with MG-CFA.

CFA Results

CFA analysis was performed to check the unidimensional factor structure. The results of the CFA analyses are presented under separate subheadings according to the grade level.

CFA Results (4th Grade Level)

The fit indices obtained from the CFA analysis are presented in Table 4.

Table 4

Fit Statistics of the CFA Model (4th Grade)

χ^2	df	χ^2/df	RMSEA	*SRMR*	*TLI*	*CFI*
14905.4	27	552.0	0.049	0.048	0.985	0.989

The χ^2/df value is expected to be less than 5 to ensure model-data fit. However, since the χ^2/df ratio is especially affected by the sample size, it is recommended to use other indexes in model-data fit. The results indicate an RMSEA of 0.049, an SRMR of 0.048, a CFI index of 0.989, and a TLI index of 0.985. According to these values, it can be concluded that the model-data fit is at an acceptable level (Schreiber et al., 2006).

The standardised factor loadings, AVE (average variance extracted) values, and reliability coefficients obtained from the CFA model are presented in Table 5.

Table 5

Standardised Factor Loadings, AVE Values, and Reliability Coefficients

Items	Factor Loadings	AVE	Cronbach Alfa	McDonald's Omega
ASBM02A	0.689			
ASBM02B	0.483			
ASBM02C	0.577			
ASBM02D	0.504			
ASBM02E	0.802	%52.55	0.902	0.905
ASBM02F	0.679			
ASBM02G	0.726			
ASBM02H	0.827			
ASBM02I	0.876			

According to the CFA results, factor loadings are expected to be 0.30 and above (Harrington, 2009). These results show that all of the factor loadings are at a sufficient level. The AVE (average variance extracted) values in the Table 5 are calculated from the factor loadings and give information about what percentage of the variance of the items the factor explains. Generally, this value is desired to be 50% and above (Hair et al., 2014). It can be said that the scale has a good AVE value. On the other hand, internal consistency coefficients provide evidence of whether a homogeneous structure is measured in addition to providing evidence of reliability (Aybek, 2022). Cronbach's Alpha and McDonald's Omega coefficients of the scales are presented in Table 5 on internal consistency. According to Büyüköztürk (2011), a reliability coefficient of 0.70 and above is considered sufficient for scale reliability. Kline (2011) categorises 0.90 and above as excellent, 0.80 and above as good, and 0.70 and above as acceptable. As seen in the Table 5, both Cronbach's Alpha and McDonald's Omega coefficients are more than 0.70 and are considered to be at a good level.

The results of the separately conducted CFAs for each country are presented in Appendix 1. It is observed that the fit indices for Morocco, South Africa, and Saudi Arabia are notably low. In addition, it is noticeable that the factor loadings of the reverse-coded items in these countries are quite low. It can be stated that the CFA results of other countries are at a good level in terms of fit indices, factor loadings, AVE, and reliability values.

CFA Results (8th Grade Level)

The fit indices obtained from the CFA analysis are presented in Table 6.

Table 6

Fit Statistics of the CFA Model (8th Grade)

χ^2	df	χ^2/df	RMSEA	$SRMR$	TLI	CFI
22238.6	27	823.6	0.048	0.039	0.993	0.994

Table 6 shows that RMSEA is 0.048, SRMR is 0.039, CFI index is 0.994, and TLI index is 0.993. According to these values, it can be said that the model-data fit is at a good level (Schreiber et al., 2006).

The standardised factor loadings, AVE (average variance extracted) values, and reliability coefficients obtained from the CFA model are presented in Table 7.

Table 7

Standardised Factor Loadings, AVE Values, and Reliability Coefficients

Items	Factor Loadings	AVE	Cronbach Alfa	McDonald's Omega
BSBM16A	0.796			
BSBM16B	0.597			
BSBM16C	0.677			
BSBM16D	0.644			
BSBM16E	0.929	%60.60	0.928	0.931
BSBM16F	0.780			
BSBM16G	0.840			
BSBM16H	0.820			
BSBM16I	0.958			

When Table 7 is analysed, it is seen that all of the factor loadings are at a sufficient level. It can be said that the AVE value is at a good level. In addition, both Cronbach's Alpha and McDonald's Omega coefficients are more than 0.70 and are accepted to be at a good level.

CFA results separately by country are presented in Appendix 2. It can be stated that the CFA fit indices of all countries are at a good or acceptable level.

Measurement Invariance Results

Measurement invariance was analysed in two stages. Firstly, under the known MG-CFA model, configurational invariance, metric invariance, scalar invariance, and strict invariance were tested in hierarchical order. In the second stage, measurement invariance was analysed by the alignment method. In the MG-CFA model, the ΔCFI

value was taken into account in determining whether the invariance was achieved between two hierarchical stages. When this value is ΔCFI<0.01 between the stages, it is interpreted that measurement invariance is provided at the relevant stage (Cheung & Rensvold, 2002). In this section, the measurement invariance of the mathematics intrinsic motivation scale according to the 4th and 8th-grade data was carried out and reported separately according to the grade level.

Measurement Invariance Results (4th Grade)

The measurement invariance findings of the mathematics intrinsic motivation scale with the MG-CFA method according to 14 countries are presented in Table 8.

Table 8

Measurement Invariance by Countries (4th Grade)

Stage	χ^2	df	χ^2/df	$RMSEA$	$SRMR$	TLI	CFI	ΔCFI
Configural	5506.1	378	14.5	0.053	0.043	0.984	0.987	-
Metric	12872.1	482	26.7	0.072	0.068	0.971	0.972	0.015

When the configural invariance findings are analysed, it is evident that the scale's structure is compatible with configural invariance in all country groups ($RMSEA$ < .08, $SRMR$ < .08, TLI > .95, CFI > .95). After configural invariance, metric invariance was analysed. However, as the difference in CFI value between the configural invariance and metric invariance stages is 0.015 (ΔCFI>0.01), it is apparent that metric invariance cannot be established across countries.

According to the results of the analyses conducted with the MG-CFA method, when the ΔCFI (ΔCFI < 0.01) values and goodness of fit statistics ($RMSEA$ < .08, $SRMR$ < .08, TLI > .95, CFI > .95) of the 4th-grade data were taken into consideration, it was concluded that 14 different countries provided only configural invariance. According to this result, it can be stated that the item-factor structure is equivalent between groups, but factor loadings, variances, covariances, and error variances are not equivalent between groups.

4th-grade level Mathematics Intrinsic Motivation scale measurement invariance with alignment method findings are presented in Table 9 and Table 10.

Table 9

Each item's Alignment Results of 14 Countries

Stage	Items	Mean	SD	Min	Max	R^2	%
Factor Loadings	ASBM02A	0.64	0.05	0.52	0.72	0.991	%0.0
	ASBM02B	0.50	0.12	0.30	0.67		
	ASBM02C	0.59	0.11	0.38	0.73		
	ASBM02D	0.47	0.09	0.36	0.63		
	ASBM02E	0.78	0.03	0.74	0.82		

	ASBM02F	0.61	0.05	0.50	0.72		
	ASBM02G	0.69	0.05	0.60	0.75		
	ASBM02H	0.79	0.03	0.74	0.84		
	ASBM02I	0.85	0.03	0.79	0.88		
Intercept	ASBM02A	1.66	0.07	1.56	1.79	0.995	%8.7
	ASBM02B	1.85	0.20	1.46	2.30		
	ASBM02C	1.84	0.14	1.54	2.08		
	ASBM02D	1.58	0.10	1.43	1.73		
	ASBM02E	1.72	0.03	1.67	1.80		
	ASBM02F	1.85	0.15	1.64	2.15		
	ASBM02G	1.89	0.09	1.80	2.12		
	ASBM02H	2.00	0.08	1.82	2.14		
	ASBM02I	2.02	0.07	1.89	2.17		

The alignment method was invariant for the factor loadings (R^2=.991) and the intercepts of the items (R^2=.995) of the Mathematics Intrinsic Motivation Scale, as seen in Table 9. While there is no non-invariant parameter in the factor parameters (0.0%), 11 parameters are non-invariant according to the item intercepts (8.7%). Since the percentage of non-invariant parameters is lower than 25%, it can be stated that the scale provides approximate measurement invariance (Asparouhov & Muthén, 2014).

Table 10 presents the invariance findings of intercept and factor loading parameters on an item and country basis. Countries in brackets and bold are labelled as non-invariant since they show more variability than the tolerance parameter.

Table 10

Each Item's Alignment Results of 14 Countries

Items	Intercepts	Factor Loadings
ASBM02A	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02B	36, (**152**), 348, 380, (**392**), (**504**), **(620), (643)**, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02C	36, 152, 348, 380, 392, (**504**), **(620),** 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02D	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02E	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926

ASBM02F	**(36)**, 152, 348, 380, 392, 504, 620, 643, **(682)**, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02G	**(36)**, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02H	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, (**710**), 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
ASBM02I	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926

Note. 36: Australia, 152: Chile, 348: Hungary, 380: Italy, 392: Japan, 504: Morocco, 620: Portugal, 643: Russia, 682: Saudi Arabia, 702: Singapore, 710: South Africa, 752: Sweden, 792: Turkey, 926: Pakistan.

When Table 10 is examined, it is seen that all factor loadings of Mathematics Intrinsic Motivation are invariant in 14 countries at the 4th-grade level. According to the intercept coefficients, it can be stated that five parameters in item ASBM02B, two parameters in item ASBM02C, two parameters in item ASBM02F, one parameter in item ASBM02G, and one parameter in item ASBM02H are non-invariant. Findings on an item basis show that 4 out of 9 items are invariant across all countries, while in 5 items, some countries are non-invariant in the intercept coefficients. On the other hand, it is seen that the negatively rooted items ASBM02B and ASBM02C are the items that violate invariance the most. Findings by country suggest that factor loading parameters are invariant across all countries. According to the intercept parameter, Australia is non-invariant in two items: Morocco in two, Portugal in two, Japan in one, Russia in one, Saudi Arabia in one, Chile in one, and South Africa in one item. It can be stated that both factor loadings and intercept parameters of Hungary, Italy, Singapore, Sweden, Turkey, and Pakistan are equivalent. As Asparouhov and Muthén (2014) suggested, since the non-invariant parameter ratio of factor loading and intercept parameters is below 25%, it can be stated that the Mathematics Intrinsic Motivation scale provides approximate invariance and all groups can be compared.

Measurement Invariance Results (8th Grade)

The measurement invariance findings of the mathematics intrinsic motivation scale with the MG-CFA method according to 14 countries are presented in Table 11.

Table 11

Measurement Invariance by Countries (8th Grade)

Stage	χ^2	df	χ^2/df	*RMSEA*	*SRMR*	*TLI*	*CFI*	ΔCFI
Configural	5480.9	378	14.5	0.049	0.036	0.993	0.994	-
Metric	12919.7	482	26.8	0.067	0.054	0.986	0.987	0.007
Scalar	25908.5	586	44.2	0.086	0.076	0.977	0.974	0.013

When analyzing the configural invariance findings, the scale structure is found to be compatible across all country groups ($RMSEA < .08$, $SRMR < .08$, $TLI > .95$, CFI

> .95). After configural invariance, metric invariance was analysed. Since the difference in CFI value between the configural invariance and metric invariance stages (ΔCFI) was 0.008, it was decided that metric invariance was also provided according to the countries. Additionally, other fit indices are also observed to be at a satisfactory level at the metric invariance stage ($RMSEA$=0.066, $SRMR$=0.064, TLI= 0.987). After metric invariance, scalar invariance was analysed. Since the ΔCFI value between metric invariance and scalar invariance is 0.013 (ΔCFI>0.01), it is decided that scalar invariance is not provided. It should also be noted that the RMSEA value at the scalar invariance stage showed incompatibility ($RMSEA$ > .08).

According to the results of the analyses conducted with the MG-CFA method, considering the ΔCFI (ΔCFI < 0.01) values and goodness of fit statistics ($RMSEA$ < .08, $SRMR$ < .08, TLI > .95, CFI > .95) of the 8th-grade data, it was concluded that 14 different countries provided configural and metric invariance. According to this result, it can be stated that item-factor structure and factor loadings are equivalent between groups, but variances, covariances, and error variances are not equivalent between groups.

8th-grade level Mathematics Intrinsic Motivation scale measurement invariance with alignment method findings are presented in Table 12 and Table 13.

Table 12

Each Item's Alignment Results of 14 Countries

Stage	Items	Mean	SD	Min	Max	R^2	%
Factor Loadings	BSBM16A	0.78	0.05	0.69	0.85	0.995	%0.0
	BSBM16B	0.65	0.09	0.51	0.80		
	BSBM16C	0.68	0.06	0.59	0.80		
	BSBM16D	0.61	0.06	0.51	0.71		
	BSBM16E	0.91	0.01	0.88	0.91		
	BSBM16F	0.72	0.03	0.67	0.79		
	BSBM16G	0.82	0.03	0.76	0.87		
	BSBM16H	0.77	0.06	0.69	0.91		
	BSBM16I	0.93	0.06	0.80	1.04		
Intercept	BSBM16A	1.96	0.07	1.86	2.09	0.997	%8.7
	BSBM16B	2.12	0.19	1.93	2.67		
	BSBM16C	2.23	0.15	1.95	2.43		
	BSBM16D	1.97	0.08	1.86	2.17		
	BSBM16E	2.07	0.06	2.01	2.18		
	BSBM16F	2.33	0.14	2.07	2.53		
	BSBM16G	2.26	0.07	2.20	2.41		
	BSBM16H	2.48	0.14	2.15	2.66		
	BSBM16I	2.46	0.08	2.33	2.58		

The alignment method was invariant for the factor loadings (R^2=.995) and the intercepts of the items (R^2=.997) of the Mathematics Intrinsic Motivation Scale, as seen in Table 12. While there is no non-invariant parameter in the factor parameters (0.0%), 11 parameters are non-invariant according to the item intercepts (8.7%). Since the percentage of non-invariant parameters is lower than 25%, it can be stated that the scale provides approximate measurement invariance (Asparouhov & Muthén, 2014).

Table 13 presents the invariance findings of intercept and factor loading parameters on an item and country basis. Countries in brackets and bold are labelled as non-invariant since they show more variability than the tolerance parameter.

Table 13

Each Item's Alignment Results of 14 Countries

Item	Intercepts	Factor Loadings
BSBM16A	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16B	36, 152, 348, 380, 392, **(504)**, 620, 643, 682, 702, 710, 752, **(792)**, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16C	36, 152, 348, 380, 392, 504, **(620)**, 643, **(682)**, 702, 710, **(752)**, **(792)**, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16D	36, 152, 348, 380, 392, 504, **(620)**, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16E	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16F	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, **(710)**, 752, 792, **(926)**	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16G	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16H	36, 152, 348, 380, 392, **(504)**, 620, 643, 682, 702, 710, 752, **(792)**, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926
BSBM16I	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926	36, 152, 348, 380, 392, 504, 620, 643, 682, 702, 710, 752, 792, 926

Note. 36: Australia, 152: Chile, 348: Hungary, 380: Italy, 392: Japan, 504: Morocco, 620: Portugal, 643: Russia, 682: Saudi Arabia, 702: Singapore, 710: South Africa, 752: Sweden, 792: Turkey, 926: Pakistan.

When Table 13 is examined, it is seen that all factor loadings of Mathematics Intrinsic Motivation are invariant in 14 countries at the 8th-grade level. According to the intercept coefficients, two parameters in item BSBM16B, four parameters in item BSBM16C, one parameter in item BSBM16D, two parameters in item BSBM16F, and one parameter in item BSBM16H are non-invariant. Findings on an item basis show that 4 out of 9 items are invariant across all countries, whereas in 5 items, there are

differences in the intercept coefficients of some countries. On the other hand, it can be said that the negatively rooted items BSBM16B and BSBM16C are the items that violate the invariance the most. The country-based findings suggest that the factor loading parameters are invariant across all countries. According to the intercept parameter, Morocco is non-invariant in two items: Portugal in two, Saudi Arabia in one, South Africa in one, Sweden in one, Turkey in three, and Pakistan in one item. Both factor loadings and intercept parameters are equivalent for Australia, Chile, Hungary, Italy, Japan, Russia, and Singapore. As Asparouhov and Muthén (2014) suggested, since the non-invariant parameter ratio of factor loading and intercept parameters is below 25%, it can be stated that the Mathematics Intrinsic Motivation scale provides approximate invariance at the 8th-grade level, and all groups can be compared.

Discussion and Conclusion

The fact that the factor structures of the groups are different from each other in the measurement made with the same scale on two or more groups may indicate that the measurement tool is perceived differently in each group or that it measures a different psychological characteristic. If this is the case, it would not be correct to compare the measurements obtained from these groups. The reason for the different factor structures of the groups may be application conditions, cultural reasons, and linguistic translation problems (Başusta, 2010). According to the results of measurement invariance, it can be evaluated whether the items in the measurement tool are perceived in the same way by the participants according to groups such as different genders, languages, and cultures.

In this study, it was examined whether the mathematics intrinsic motivation scale in the TIMSS 2019 application showed measurement invariance according to 14 different countries. The analyses were performed with MG-CFA and MG Alignment methods. Since the stages in measurement invariance with MG-CFA are nested models, $\Delta\chi^2$ and ΔCFI values are generally used in the literature to compare the models. In this study, ΔCFI value was used to decide the measurement invariance between the stages due to the large sample size. Other goodness of fit statistics is also reported.

According to the measurement invariance results of the study with MG-CFA, the TIMSS 2019 4th grade mathematics intrinsic motivation scale provides only configural invariance across 14 different countries. It was observed that metric, scalar, and strict invariance were not achieved. Similarly, the 8th-grade mathematics intrinsic motivation scale provides configural and metric invariance across 14 different countries. However, it was observed that it did not provide scalar and strict invariance. Following these results, measurement invariance was analysed with the Alignment method. The results of the alignment method applied to factor loading and intercept parameters showed that the mathematics intrinsic motivation scale at the 4th-grade level provided approximate measurement invariance and that the scores of the countries could be compared. Similarly, at the 8th-grade level, the results showed that the mathematics intrinsic motivation scale also provided approximate measurement invariance and that the scores of the countries could be compared.

In the literature, there are studies examining the measurement invariance of different scales used in large-scale assessments across countries. He et al. (2019) examined the measurement invariance of 29 countries according to Instrumental Motivation, Enjoyment of Science, and Sense of School Belonging scales with PISA

and TIMSS data. The results show that all three scales provide only metric invariance. Raižienė et al. (2021) examined the measurement invariance of the motivational constructs (competitiveness, work mastery, and fairness of failure) model with the PISA 2018 application with EU countries and concluded that the model provides metric invariance. Tekin and Çobanoğlu-Aktan (2021) conducted a measurement invariance study comparing Turkey, Norway, and Singapore on the collaborative problem-solving scale with PISA 2015 data. The results of the study reported that these three countries met only the configural invariance stage. Ersözlü et al. (2022) reported that metric invariance and partial scalar invariance of the mathematics anxiety scale were met in their study examining measurement invariance between Australian and Russian groups. Karakoç-Alatlı et al. (2016) examined the measurement invariance of the TIMSS 2011 mathematics test between four different countries (Turkey, England, Japan, and the USA) and concluded that the test provided only configural invariance. According to measurement invariance with MG-CFA, the results of the current study are similar to the results of these studies. However, it should be noted that these studies did not examine measurement invariance with the Alignment method but only with MG-CFA. On the other hand, MG-CFA is not practical in comparing a large number of groups (i.e., countries) since it makes many pairwise comparisons between countries and may produce incorrect results (Sırgancı et al., 2020). In addition, when measurement invariance cannot be achieved with MG-CFA, it does not provide any information about the groups that violate invariance. Therefore, measurement invariance should be examined with the Alignment method, which overcomes these disadvantages of MG-CFA (Asparouhov & Muthén, 2014).

Glasgow et al. (2019) examined the measurement invariance of mathematics teachers' Job satisfaction, School emphasis on academic success, School condition and resources, Safe and orderly school, and teacher Self-efficacy scales obtained from TIMMS 2015 teacher surveys across 46 countries. The results showed that only three constructs achieved metric invariance, while the other constructs remained in configural invariance. In the present study, the results of measurement invariance using the alignment method show that all five constructs achieve approximate invariance, indicating that these constructs can be validly compared across educational systems. Sirganci et al. (2020), in their study explaining the basic concepts and processes of the alignment method, examined the measurement invariance of 56 countries on the Instrumental Motivation Scale data in the PISA 2015 cycle, MG-CFA results show that the scale provides only configural invariance. The measurement invariance findings with the alignment method provide more detailed information about which countries and which items contribute better to measurement invariance. Kaya et al. (2023) examined the measurement invariance of the Self-Efficacy scale data obtained from PISA 2018 on 79 countries. The researchers report that only configural invariance was met in the cross-country comparison. Jami and Kemmelmeier (2021), in their study examining the measurement invariance of the Subjective well-being scale in 36 countries, MG-CFA findings showed that metric invariance was achieved, while the alignment method was found to provide approximate measurement invariance of the scale. The results of the current study are similar to the results of these studies in the literature (Glasgow et al., 2019; Sırgancı et al., 2020; Jami & Kemmelmeier, 2020). Ertürk and Oyar (2021) examined the measurement invariance between different

countries with different methods using the "Mathematics Liking Scale" obtained from PISA 2015 data. Researchers stated that when determining the method to be used in measurement invariance studies, attention should be focused on meeting the assumptions and the structure of the data (Ertürk & Oyar, 2021). Therefore, in line with both the findings of this study and Sırgancı et al. (2020), the Alignment Method should be preferred in examining the measurement invariance of scales/tests with a large number of groups.

When the measurement invariance results on an item basis were analysed by alignment method, it was concluded that four items at the 4th-grade level and four items at the 8th-grade level were invariant according to both factor loadings and intercept parameters. On the other hand, the second and third items in the scale have negatively worded items, and it can be seen that the intercept parameters of these two items are the most non-invariant parameters according to the intercept parameter among the countries. This result shows that items negatively worded have a decreasing effect on measurement invariance. Therefore, it is important to translate the items with negative roots in a more comprehensible way and in accordance with the student's level.

When the results of measurement invariance by alignment method on a country basis were analysed, it was concluded that the parameters of 14 countries were equivalent in six countries at the 4th-grade level and seven countries at the 8th-grade level according to both factor loadings and intercept parameters. Morocco, Portugal, Saudi Arabia, and South Africa were found to have non-invariant parameters in the intercept parameters at both grade levels.

In conclusion, as Asparouhov and Muthén (2014) suggested, since the factor loading and the non-invariant parameter ratio of the intercept parameters are below 25%, the Mathematics Intrinsic Motivation scale provides approximate measurement invariance at both 4th and 8th-grade levels and the scores of the countries are comparable.

Limitations and Recommendations

In this study, the measurement invariance of the mathematics intrinsic motivation scale was compared across countries at both the 4th and 8th grades. Therefore, countries that completed this scale at both 4th and 8th-grade levels were selected, and other countries were not included in the scope of the study. This is a limitation of the study. Future studies can be carried out to cover all countries according to one grade level of this scale.

Scores obtained from large-scale exams such as PISA and TIMSS can be compared with the scores obtained by applying the same scale in different grades with measurement invariance, and the differentiation of the factor structure of different grade levels can be examined.

In this study, it was concluded that questions with negative roots had a disruptive effect on measurement invariance. The effect of negatively rooted questions on measurement invariance can be addressed in future studies.

The findings of this study - similar to the literature - show that in cross-country comparisons, MG-CFA findings do not indicate the presence of measurement invariance, while the findings of the alignment method indicate approximate

measurement invariance. Therefore, it is recommended that measurement invariance be tested with the alignment method in future multi-group measurement invariance studies.

Conflicts of Interest

There is no conflict of interest in the research.

Author Bio:

Mahmut Sami Yiğiter completed his master's degree at Ankara Hacı Bayram Veli University. He continues his doctoral studies in Hacettepe University in the field of Measurement and Evaluation in Education. He is currently working as a lecturer at Social Sciences University of Ankara. His research focuses on measurement and evaluation in education, measurement invariance, cross-cultural comparisons, meta-analysis and methodological methods.

References

Adıbatmaz, F. B. K., & Yildiz, H. (2020). The effects of distractors to differential item functioning in peabody picture vocabulary test. *Journal of Theoretical Educational Science*, *13*(3), 530-547.

Afanador, N. L., Tran, T., Blanchet, L., & Baumgartner, R. (2016). mvdalab-package 3.

Ahmed, W., Minnaert, A., Van der Werf, G., & Kuyper, H. (2010). Perceived social support and early adolescents' achievement: The mediational roles of motivational beliefs and emotions. *Journal of Youth and Adolescence*, *39*(1), 36–46. doi:10.1007/s10964-008-9367-7

Akben-Selcuk, E. (2017). Personality, motivation, and math achievement among Turkish students: Evidence from PISA data. *Perceptual and Motor Skills*, *124*(2), 514–530. https://doi.org/10.1177/0031512516686505

Arikan, S., Özer, F., Şeker, V., & Ertaş, G. (2020). The importance of sample weights and plausible values in large-scale assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(1), 43-60. doi: https://doi.org/10.21031/epod.602765

Barak, M., & Asad, K. (2012). Teaching image-processing concepts in junior high schools: Boys' and girls' achievements and attitudes towards technology. *Research in Science and Technological Education, 30*(1), 81–105. https://doi.org/10.1080/02635143. 2012. 656084

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

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). *Scientific research methods*. Pegem.

Cardoso, M. E. (2020). Policy evidence by design: International large-scale assessments and grade repetition. *Comparative Education Review*, *64*(4), 598-618. https://doi.org/10.1086/710777

Cheung G. W., Rensvold R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1-27. https://doi.org/10.1177/014920639902500101

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

Cleary, T. J., & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology*, *47*(5), 291–314. https://doi.org/10.1016/j.jsp.2009.04.002

Engel, L. C., & Rutkowski, D. (2021). Costs of big data. In C. Wyatt-Smith, B. Lingard, & E. Heck (Eds.), *Digital disruption in teaching and testing* (pp. 124–135). Routledge.

Ertürk, Z., & Erdinç-Akan, O. (2018). TIMSS 2015 matematik başarısı ile ilgili bazı değişkenlerin cinsiyete göre ölçme değişmezliğinin incelenmesi. *Journal of Theoretical Educational Science*, *UBEK-2018*, 204-226.

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, *10,* 1507. https://doi.org/10.3389/fpsyg.2019.01507

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Freiberger, V., Steinmayr, R., & Spinath, B. (2012). Competence beliefs and perceived ability evaluations: How do they contribute to intrinsic motivation and achievement?. *Learning and Individual Differences*, *22*(4), 518-522. https://doi.org/10.1016/j.lindif.2012.02.004

Glassow, L. N., Rolfe, V., & Hansen, K. Y. (2021). Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: an alignment optimization approach. *Educational Assessment Evaluation and Accountability*, *33*(1), 105–137. https://doi.org/10.1007/s11092-020-09348-2

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 Suppl 3), S78.

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*(8), 1163–1176. https://doi.org/10.1037/a0039440

Gustafsson, J.-E. (2018). International large scale assessments: Current status and ways forward. *Scandinavian Journal of Educational Research*, *62*(3), 328–332. https://doi.org/10.1080/00313831.2018.1443573

He, J., Barrera-Pedemonte, F., & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education Principles Policy and Practice*, *26*(4), 369–385. https://doi.org/10.1080/0969594x.2018.1469467

Henderson, R. W. & Landesman, E. M. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *Journal of Educational Research*, *88*(5), 290–300.

Hooper, M., Mullis, I. V., Martin, M. O., & Fishbein, B. (2020). TIMSS 2019 context questionnaire framework. *TIMSS*, 59-78.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3–4), 117–144. https://doi.org/10.1080/03610739208253916

Hu, L.-T., & Bentler, P. M. (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

Ilter, İ. (2021). The relationship between academic amotivation and academic achievement: A study on middle school students. *Kuramsal Eğitimbilim Dergisi*, *14*(3), 389–410. https://doi.org/10.30831/akukeg.847145

İlhan, M., & Çetin, B. (2013). Matematik odaklı epistemolojik inanç ölçeği (MOEİÖ): Geçerlik ve güvenirlik çalışması. *Kurumsal Eğitimbilim Dergisi*, *6*(3), 362-368.

Jami, W. A., & Kemmelmeier, M. (2020). Assessing well-being across space and time: Measurement equivalence of the WHO-5 in 36 European countries and over 8 years. *Journal of Well-Being Assessment*, *4*(3), 419–445. https://doi.org/10.1007/s41543-021-00042-8

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific software international.

Kaliyaperumal, S. K., Kuppusamy, M., & Gounder, A. S. (2015). Outlier detection and missing value in time series ozone data. *International Journal of Scientific Research in Knowledge*, *3*(9), 220–226. https://doi.org/10.12983/ijsrk-2015-p0220-0226

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*, 1258. https://doi.org/10.3389/fpsyg.2019.01258

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kaya, S., Eryilmaz, N., & Yuksel, D. (2023). A cross-cultural comparison of self-efficacy as a resilience measure: Evidence from PISA 2018. *Youth & Society*. https://doi.org/10.1177/0044118x231186833

Kline, R. B. (2011). Convergence of structural equation modeling and multilevel modeling. In W. Vogt & M. Williams (Eds.), The SAGE handbook of innovation in social research methods (pp. 562–589). SAGE Publications. https://doi.org/10.4135/9781446268261.

Koğar, H., & Yilmaz Koğar, E. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Journal of Measurement and Evaluation in Education and Psychology*, *6*(2), 351-364. https://doi.org/10.21031/epod.94857

Malone, T. W., & Lepper, M. R. (2021). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction* (pp. 223-254). Routledge.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Middleton, J. (1995). A study of intrinsic motivation in the mathematics classroom: A personal constructs approach. *Journal for Research in Mathematics Education, 26*(3), 254–279. https://doi.org/10.2307/749130

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-130.

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.

Mueller, M., Yankelewitz, D., & Maher, C. (2011). Sense making as motivation in doing mathematics: Results from two studies. *The Mathematics Educator*, *20*(2), 33–43.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS and PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Muthén ©n, B., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00978

Nugent, G., Barker, B., Grandgenett, N., & Adamchuk, V. (2010). Impact of robotics and geospatial technology interventions on youth stem learning and attitudes. *Journal of Research on Technology in Education, 42*(4), 391–408.

OECD. (2013). *Students' drive and motivation.* In PISA 2012 results: Ready to learn (Volume III): Students' engagement, drive, and self-beliefs. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/ 9789264201170-7-en.

Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., ... & Wuertz, D. (2018). Package 'performanceanalytics'. *R Team Cooperation*, *3*, 13-14.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review: DR*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Raižienė, S., Ringienė, L., Laukaityte, I., & Jakaitienė, A. (2021). Measurement invariance of pisa 2018 motivational constructs across EU countries. *EDULEARN21 Proceedings* (pp. 7081-7081). IATED.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, *35*(2), 299–331. https://doi.org/10.1016/s0005-7894(04)80041-8

Robitzsch, A. (2020). sirt: Supplementary item response theory models. R package version 3.4-64. https://CRAN.R-project.org/package=sirt

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2012). lavaan: An*R*Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Ryan, R.M., & Deci E.L., (2009) Promoting self-determined school engagement: motivation, learning, and well-being. In D. B. Miele & K. R. Wentzel (Eds.), *Handbook on motivation at school* (pp 171–196). Routledge.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003

Shores, M. L., & Shannon, D. M. (2007). The effects of self-regulation, motivation, anxiety, and attributions on mathematics achievement for fifth and sixth grade students. *School Science and Mathematics*, *107*(6), 225–236. Retrieved from http://ssmj.tamu.edu

Sırgancı, G., Uyumaz, G., & Yandi, A. (2020). Measurement invariance testing with alignment method: Many groups comparison. *International Journal of Assessment Tools in Education*, *7*(4), 657–673. https://doi.org/10.21449/ijate.714218

Sözer, E., Eren, B., & Kahraman, N. (2021). Investigating measurement invariance for longitudinal assessments: An application using repeated data over four weeks. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *41*(2), 729–763. https://doi.org/10.17152/gefad.873885

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

Taris, T. W., Bok, I. A., & Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *The Journal of Psychology*, *132*(3), 301–316. https://doi.org/10.1080/00223989809599169

Tavani C.M., & Losh S.C. (2003) Motivation, self-confidence, and expectations as predictors of the academic performances among our high school students. *Child Study J, 33*(3), 141–151.

Tierney, N., Cook, D., McBain, M., & Fay, C. (2021). naniar: Data structures, summaries, and visualisations for missing data (R package version 0.6.1)[Computer software].

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 Education Sciences*, *2014*(3), 30-43.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, & M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, *6*, 1064. https://doi.org/10.3389/fpsyg.2015.01064

Vandenberg, 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*(1), 4–70. https://doi.org/10.1177/109442810031002

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *British Journal of Educational Psychology*, *87*(2), 187-204.

Wickham, H., François, R., Henry, L., Müller, K., & Wickham, M. H. (2019). Package 'dplyr'. *A Grammar of Data Manipulation. R package version*, *8*.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186-191.

Woolley, M. E., Strutchens, M. E., Gilbert, M. C., & Martin, W. (2010). Mathematics success of black middle school students: Direct and indirect effects of teacher expectations and reform practices. *Negro Educational Review*, *61*(1), 41–59.

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*. University of Massachusetts Amherst. https://doi.org/10.7275/MHQA-CD89

Yiğiter, M. S. (2023). Matematik duyuşsal özellik faktörlerinin cinsiyete göre ölçme değişmezliğinin incelenmesi: TIMSS 2019 Türkiye örneği. *Anadolu Üniversitesi Eğitim Fakültesi Dergis*i, *7*(4), 859–882. https://doi.org/10.34056/aujef.1198134

Yildirim, S. (2011). Self-efficacy, intrinsic motivation, anxiety and mathematics achievement: Findings from Turkey, Japan and Finland. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *5*(1), 277-291.

Yin, L., & Fishbein, B. (2019). Creating and interpreting the TIMSS 2019 context questionnaire scales. *Methods and procedures: TIMSS*, 16-1.

Zembat, R., Akşin-Yavuz, E., Tunçeli, H. İ., Yılmaz, H. (2018). Öğretmenlik mesleğine yönelik tutum ile akademik motivasyon ve başarı arasındaki ilişkinin incelenmesi. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science], 11*(4), 789-808.

Appendix 1

CFA Result (4th Grade)

Country	i1	i2	i3	i4	i5	i6	i7	i8	i9	cfi	tli	rmsea	srmr	Cronbach	Omega	AVE
Australia	0.826	0.750	0.824	0.565	0.930	0.732	0.838	0.933	1.007	0.998	0.998	0.030	0.025	0.941	0.944	0.657
Chile	0.680	0.547	0.661	0.473	0.851	0.627	0.773	0.869	0.912	0.985	0.980	0.063	0.053	0.899	0.903	0.519
Hungary	0.849	0.647	0.675	0.520	0.936	0.721	0.720	0.938	0.970	0.995	0.994	0.042	0.036	0.919	0.923	0.581
Italy	0.802	0.760	0.757	0.442	0.864	0.697	0.681	0.895	0.932	0.998	0.997	0.026	0.025	0.926	0.930	0.605
Japan	0.744	0.417	0.531	0.602	0.899	0.711	0.803	0.816	0.865	0.996	0.995	0.044	0.044	0.934	0.939	0.642
Morocco	0.331	0.262	0.297	0.403	0.471	0.408	0.470	0.516	0.520	0.921	0.894	0.062	0.067	0.778	0.768	0.278
Portugal	0.667	0.583	0.684	0.458	0.811	0.609	0.735	0.787	0.881	0.993	0.991	0.041	0.043	0.919	0.922	0.577
Russia	0.686	0.492	0.606	0.446	0.741	0.642	0.663	0.798	0.870	0.997	0.996	0.033	0.032	0.905	0.909	0.536
Saudi Arabia	0.547	0.359	0.487	0.555	0.706	0.576	0.667	0.746	0.751	0.954	0.938	0.072	0.066	0.863	0.862	0.420
Singapore	0.740	0.723	0.724	0.574	0.858	0.671	0.810	0.809	0.945	0.998	0.997	0.033	0.030	0.933	0.936	0.621
South Africa	0.517	0.285	0.340	0.508	0.656	0.603	0.593	0.622	0.667	0.934	0.911	0.080	0.062	0.796	0.794	0.312
Sweden	0.722	0.699	0.881	0.655	0.938	0.756	0.775	0.898	0.974	0.998	0.997	0.033	0.029	0.944	0.946	0.664
Turkey	0.540	0.526	0.633	0.325	0.657	0.449	0.634	0.735	0.780	0.988	0.984	0.038	0.045	0.885	0.889	0.484
Pakistan	0.768	0.680	0.768	0.529	0.879	0.719	0.800	0.909	0.977	0.997	0.996	0.034	0.030	0.935	0.938	0.633

Note. i1: item1 factor loading, AVE : average variance extracted

Appendix 2

CFA Result (8th Grade)

Country	i1	i2	i3	i4	i5	i6	i7	i8	i9	cfi	tli	rmsea	srmr	Cronbach	Omega	AVE
Australia	0.863	0.722	0.692	0.636	0.926	0.724	0.843	0.795	0.944	0.997	0.996	0.041	0.031	0.938	0.940	0.641
Chile	0.748	0.610	0.696	0.558	0.900	0.701	0.829	0.753	0.904	0.994	0.992	0.054	0.045	0.930	0.932	0.609
Hungary	0.875	0.635	0.671	0.647	0.968	0.805	0.802	0.734	0.852	0.996	0.995	0.045	0.035	0.932	0.934	0.614
Italy	0.929	0.808	0.815	0.587	1.039	0.824	0.869	0.790	1.056	0.999	0.999	0.032	0.023	0.948	0.951	0.688
Japan	0.795	0.515	0.582	0.669	0.902	0.666	0.812	0.697	0.866	0.995	0.994	0.048	0.043	0.942	0.945	0.664
Morocco	0.733	0.486	0.598	0.668	0.862	0.682	0.722	0.857	0.903	0.990	0.986	0.056	0.048	0.908	0.911	0.540
Portugal	0.869	0.774	0.776	0.694	0.982	0.724	0.892	0.836	1.004	0.998	0.997	0.035	0.027	0.947	0.948	0.674
Russia	0.694	0.518	0.568	0.545	0.792	0.656	0.733	0.675	0.844	0.996	0.994	0.043	0.038	0.922	0.924	0.582
Saudi Arabia	0.790	0.557	0.711	0.750	0.994	0.832	0.903	0.936	1.035	0.991	0.989	0.068	0.053	0.928	0.931	0.607
Singapore	0.767	0.754	0.663	0.554	0.862	0.655	0.818	0.724	0.984	0.997	0.996	0.041	0.034	0.935	0.938	0.632
South Africa	0.632	0.546	0.563	0.514	0.812	0.651	0.758	0.681	0.865	0.988	0.984	0.054	0.045	0.890	0.893	0.488
Sweden	0.809	0.719	0.801	0.669	0.925	0.798	0.828	0.744	0.910	0.996	0.995	0.046	0.035	0.943	0.945	0.657
Turkey	0.700	0.721	0.809	0.519	0.897	0.752	0.840	0.872	1.016	0.993	0.991	0.053	0.041	0.918	0.921	0.571
Pakistan	0.777	0.703	0.687	0.611	0.868	0.706	0.824	0.768	0.887	0.997	0.996	0.043	0.035	0.933	0.934	0.614

Note. i1: item1 factor loading, AVE : average variance extracted