International Journal of Assessment Tools in Education
2020, Vol. 7, No. 4, 657–673
https://doi.org/10.21449/ijate.714218
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

Measurement Invariance Testing with Alignment Method: Many Groups Comparison

Gozde Sirganci [1], **Gizem Uyumaz** [2,*], **Alperen Yandi** [3]

[1]Yozgat Bozok University, Department of Educational Sciences, Yozgat, Turkey
[2]Giresun University, Department of Educational Sciences, Giresun, Turkey
[3]Bolu Abant İzzet Baysal University, Department of Educational Sciences, Bolu, Turkey

ARTICLE HISTORY

Received: Apr. 03, 2020
Revised: Sept. 30, 2020
Accepted: Oct. 15, 2020

KEYWORDS

Multi-group confirmatory factor analysis,
Validity,
Modification indexes,

Abstract: It is necessary to examine the measurement invariance (MI) among groups in studies where different groups are compared by using a measurement instrument. Most of the studies, measurement invariance is tested with multiple group confirmatory factor analysis. This model applies many model adjustments based on the modification indexes. Therefore, it is not practical due to too many large modification indexes while testing MI over many groups. Besides scalar model is a poor model fit when comparing many groups and so does not hold MI. In this study, the aim is to explain the basic concepts and processes of the alignment method which is offered as a new method for testing MI and illustrate an application on the real data set. In this study, measurement invariance among 56 countries including Turkey is tested with alignment method in order to set an example for researchers. For this purpose, the Instrumental Motivation Scale data, which is one of the psychological measurement instruments used in PISA 2015, was used. As a result of MG-CFA, it was found that configural invariance was ensured. The fit indexes of CFI and TLI were calculated as 0.982 and 0.946 respectively in this stage. After that, metric invariance was tested by considering the difference of fit indices obtained for the two stages. It was found that the metric invariance could not be provided. Alignment results show which countries hold MI and which do not. Besides it provides information which items have the most invariants for groups that hold MI.

1. INTRODUCTION

Validity is an important psychometric property that must be examined in every study that has been conducted with measuring instruments. Bias is one of the most important sources of systematical error that affect the validity (Messick,1995). Test bias, defined as a systematic error is the measurement, captures the idea that there are construct-irrelevant components that result in systematically higher or lower scores on the measurement for the groups under examination (American Educational Research Association [AERA] & National Council on Measurement in Education [NCME], 1999). In bias, the scores of individuals contain systematical error depending on their subgroup (Camilli & Shepard, 1994; Zumbo 1999). While

CONTACT: Gizem UYUMAZ ✉ gizemuyumaz@gmail.com Giresun University, College of Education, Department of Educational Sciences- Giresun, Turkey

ISSN-e: 2148-7456 /© IJATE 2020

determining bias on item level, item response function of the item is analyzed. During the investigations made when determining bias, measurement invariance concept has been encountered. Measurement invariance is the statistical property of the correlation being the same between the observed variable (items) and latent variable (measured trait) among the subgroups (Drasgow & Kanfer, 1985; Widaman & Reise, 1997). In studies where different groups are compared by means of a measurement instrument, measurement invariance among the groups is needed to be investigated and the invariance must be proven. Measurement invariance is a validity issue; and if the measurement invariance cannot be obtained among the subgroups, we cannot make comparisons among groups.

For testing measurement invariance, there are methods based on two different theories. The first one is based on Item Response Theory (IRT), and the other is based on Structural Equation Modelling (SEM). One of the most used methods for examining measurement invariance under SEM is Multiple Group Confirmatory Factor Analysis (MG-CFA). MG-CFA is used in order to determine whether the factor structure of a scale is equal in multiple samples or in multiple subgroups (according to gender, socio-economic level, nation, religion, culture, etc.) (Jöreskog, Sörbom, Toit & Toit, 2001). Four hierarchical models are tested with MG-CFA: configural, metric (weak factorial), scalar (strong factorial), and strict invariance (Byrne, Shavelson & Muthen, 1989; Vandenberg & Lance, 2000). A hierarchical order exists among the models after the first model is confirmed the testing of the second model starts. While testing the models, the number of limited parameters is increased gradually. For configural invariance, no limitation is used in order to equalize any parameters between the groups. Metric invariance assumes that the factor loadings of the across groups are equal. In this way factor variances across groups and structural relations can be comparable. Scalar invariance assumes that both the factor loadings and the measurement intercept (thresholds with categorical items) are invariant among the groups, and only in this way, it becomes possible to compare factor means and variances among the groups. The strict invariance holds the value of the residual variances equal across groups (Muthén & Asparouhov, 2018). Equations related to models are presented in Equation 1-3.

Configural:
$$y_{ig} = v_g + \lambda_g f_{ig} + \varepsilon_{ig} \quad (1)$$
$$E(f_g) = \alpha_g = 0,\ V(f_g) = \psi_g = 1$$

Metric:
$$y_{ig} = v_g + \lambda_g f_{ig} + \varepsilon_{ig} \quad (2)$$
$$E(f_g) = \alpha_g = 0,\ V(f_g) = \psi_g$$

Scalar:
$$y_{ig} = v_g + \lambda_g f_{ig} + \varepsilon_{ig} \quad (3)$$
$$E(f_g) = \alpha_g,\ V(f_g) = \psi_g$$

g: number of groups, i: number of independent observations in group g, f_g : latent variable, λ_g:factor loading, v_g: measurement intercept, α_g:factor mean ve ψ_g:factor variance

To provide model fit at any phase (model is usually rejected at strict invariance phase) may depend on large number of modifications. In circumstances where so many large modification indexes are presented MGFA fails. Because the presence of modification indexes enabling so many large valued changes shows that to achieve an acceptable model a long model modification line is needed. In this situation, the sources that ruin invariance cannot be defined suitably. Therefore, it is not guaranteed to achieve a suitable model at the end of the modifications (Asparouhov & Muthén, 2014). MGCFA is based on dual comparisons across groups. Since comparisons are made for each item when the number of groups is larger, the

number of dual comparisons will increase exponentially; and this will increase the possibility of miscalculation of the measurement invariance, and will make the method unfavorable (Muthén & Asparouhov, 2018; Rutkowski & Svetina, 2014). On the other hand, for methods based on IRT, it is unlikely to talk about metric invariance. Because, while only the regression slope is constant, the regression intercept is not constant; it is hard to say that an item will be perceived the same by individuals. Especially in the analysis of many groups, scalar invariance rarely fits the data set (Muthén & Asparouhov, 2018). Because of the reasons mentioned here, while testing scalar invariance, comparison of factor means among groups is nearly impossible, either for SEM or IRT based conventional methods.

MG-CFA is impractical in comparing too many groups. In contrast, the alignment method automates and greatly simplifies the measurement invariance analysis. In addition, the alignment method can be used for determining invariance of parameters singly and which item provides the invariance mostly in measurement instrument. This situation is important for determining the best-fit CFA model that provides partial measurement invariance when estimating factor mean and variance of the groups. Also, the alignment method determines which group contributes to the measurement invariance by a single analysis. For this reason, in studies where measurement invariance is examined, when confronted with a large number of modifications, and especially when the number of groups is large; a new method is needed in order to investigate whether there is invariance across the groups or not. In this study it is aimed to explain basic terminologies and processes of the Alignment method suggested by Asparouhov and Muthén (2014) and explain an application example.

The Alignment method does not presume exact measurement invariance. It can estimate factor mean and variance parameters in each group while discovering the most suitable measurement invariance pattern. The strong aspect of the alignment method is that it is based on configural model and can predict the most suitable models for a large number of groups. In the configural model, since measurement intercepts and loadings are free across groups, factor means and variances cannot be defined. However, the model sets the metric of factor by fixing the factor mean to zero and the factor variance to 1. In the configural model, since the factor mean and variance are not defined latent characteristic (factor) cannot be compared across groups; that means it scales differently for each group. It is not possible to compare factorial scores of individuals situated in different groups and intergroup factor mean. The Alignment method can predict factor mean and variance for each group without assuming measurement invariance and by discovering the most suitable measurement invariance pattern. With this aspect, the method gives information about the level of measurement invariance along with intergroup factor mean and variance by calculating approximate measurement invariance. Thereby, which measurement parameters are approximately constant, and which are not specified (Asparouhov & Muthén, 2014; Kim, Cao, Wang & Nguyen, 2017). In other words, the alignment method can estimate factor loadings (λ_g), measurement intercepts (υ_g), factor means (α_g) and variances (ψ_g) by predicting the number of variable item parameters and the model that can hold impaired measurement variance at the minimum level (Muthén & Asparouhov, 2018).

One advantage of the alignment method is that it has the same model fit with the configural model. The method minimizes the distortions of measurement invariance by predicting group-specific factor mean (α_g) and variance (ψ_g). Although these parameters cannot be defined without applying strong invariance, this is possible by using a series of constraints that optimize the simplicity function in the alignment method. The simplicity function "F" is optimized with few parameters that are not substantially invariant, and many parameters that are not nearly invariable, rather than many parameters that are not moderately invariant. This alignment method includes a simplification function similar to the rotation (Jennrich, 2006) used in Exploratory Factor Analysis (EFA) (Muthén & Asparouhov, 2014).

The measurement invariance is tested in two steps by the alignment method. In the first step, the configural model in which factor loading and measurement intercepts are free across groups and the factor means are fixed to 0 and the variances to 1 in all groups are estimated. This configural model, which is called the base model "M_0", is the best fit model among the multi-group factor analysis models since it does not contain parameter constraints across groups. In the second step, alignment optimization is done. At this stage, factor means and variances set free, and factor means and variances are calculated by a simplicity function that minimizes the distortions of measurement invariance. This simplicity function consists of the loss function (f) for each group pair, where each measurement intercepts and factor loadings values are components (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2018).

The model, which is not defined by the simplicity function, is defined by adding factor means and variances to the configural model. The simplicity function is presented in Equation 4.

$$F = \sum_p \sum_{g_1<g_2} w_{g_1,g_2} f(\lambda_{pg_1} - \lambda_{pg_2}) + \sum_p \sum_{g_1<g_2} w_{g_1,g_2} f(v_{pg_1} - v_{pg_2}) \quad (4)$$

$$w_{g_1,g_2} = \sqrt{N_{g_1} N_{g_2}} \quad (5)$$

w: factor weight, N: sample size of the group

The proposed final aligned model has the same fit as the M_0 model. Although the aligned model tries to minimize the amount of invariance, it does not compromise the model fit. The relationship between the M_0 model and the last aligned model is in line with the relationship between the non-rotated model in exploratory factor analysis which has the best fit between a fixed number of factors and all EFAs and the rotated model which has the same fit with the non-rotated model without compromising the fit of the model (Muthén & Asparouhov, 2014).

There are two different alignment optimizations in the alignment method: FIXED and FREE. In FIXED alignment optimization, the factor means of the first group is restricted to 0 and its variance to 1. In FREE optimization, there is no restriction on the factor mean and variance of the first group, and these parameters are considered as additional parameters that should be estimated (Kim, Cao, Wang & Nguyen, 2017).

2. METHOD

This study is a descriptive research which aims to assess the measurement invariance via alignment method of the Instrumental Motivation Scale (INSTSCIE) which is in science learning applied in PISA (The Programme for International Student Assessment) 2015 and to introduce of "alignment method" in this assessment. In the literature, there are some studies about exact MI which is done with PISA 2015 data. To debated this new MI method's results with the studies in the literature, PISA 2015 data is used in this study.

2.1. Study Group

PISA 2015 has been implemented in 72 countries, of which 35 are OECD members. In this study, which aims to explain the basic concepts and processes of the alignment method for measurement invariance and to introduce an example using the alignment method for researchers, the data of 406,961 participants from 57 countries which answered the INSTSCIE were used. Information about the countries in the study and the number of participants in the countries are shown in Table 1.

Table 1. *Frequencies and percentages related to the study group*

	Code	Country	f	%		Code	Country	f	%		Code	Country	f	%
1	Tur	Turkey	5608	.38	**0**	Grc	Georgia	5239	1.29	**9**	Prt	Portugal	6982	1.72
2	Aus	Australia	13377	.29	**1**	Hkg	Hong Kong	5158	1.27	**0**	Qat	Qatar	10682	2.62
3	Aut	Austria	6708	.65	**2**	Hun	Hungary	5341	1.31	**1**	Rus	Russia	5477	1.35
4	Bel	Belgium	8754	.15	**3**	Isl	Iceland	3150	0.77	**2**	Sgp	Singapore	5971	1.47
5	Bra	Brazil	18276	.49	**4**	Irl	Ireland	5473	1.34	**3**	Svk	Slovakia	5759	1.42
6	Bgr	Bulgaria	5143	.26	**5**	Ita	Italy	10815	2.66	**4**	Svn	Slovenia	5913	1.45
7	Can	Canada	18706	.60	**6**	Jpn	Japan	6404	1.57	**5**	Esp	Spain	6474	1.59
8	Chl	Chile	6731	.65	**7**	Kor	South Korea	5443	1.34	**6**	Swe	Sweden	5071	1.25
9	Tap	Taipei	7576	.86	**8**	Lva	Latvia	4678	1.15	**7**	Che	Switzerland	5545	1.36
10	Col	Colombia	11019	.71	**9**	Ltu	Lithuania	6047	1.49	**8**	Tha	Thailand	7856	1.93
11	Cri	Costa Rica	5686	.40	**0**	Lux	Luxembourg	4925	1.21	**9**	Are	United Arab Emirates	12940	3.18
12	Hrv	Croatia	5447	.34	**1**	Mac	Macau (China)	4414	1.08	**0**	Tun	Tunisia	4532	1.11
13	Cze	Czech Republic	6397	.57	**2**	Mex	Mexican	7209	1.77	**1**	Gbr	United kingdom	13082	3.21
14	Dnk	Denmark	6440	.58	**3**	Mne	Serbia	4945	1.22	**2**	Usa	USA	5414	1.33
15	Dom	Dominican Republic	3992	.98	**4**	Nld	Netherlands	5078	1.25	**3**	Ury	Uruguay	5412	1.33
16	Est	Estonia	5312	.31	**5**	Nzl	New Zealand	4239	1.04	**4**	Qch	B-S-J-G (China)	9564	2.35
17	Fin	Finland	5621	.38	**6**	Nor	Norway	5093	1.25	**5**	Qes	Spain B	31003	7.62
18	Fra	France	5312	.31	**7**	Per	Peru	6535	1.61	**6**	Quc	Massachusette	1534	0.38
19	Deu	Germany	5353	.32	**8**	Pol	Poland	4336	1.07	**7**	Que	North Carolina	1770	0.43
												Toplam	406961	100

The country with the highest number of participants is the Spain regions with 31.003 participants (7.62%). The statewith the lowest number of participants is Massachusette with 1,534 participants (0.38%). Turkey has participated the PISA 2015 with 5,608 students (1.38%).

2.2. Data Collection Tool

In this study, data of ST113 (INSTSCIE-Instrumental Motivation to Learn Science), one of the psychological measurement tools used in the PISA 2015, was used. According to OECD (2016):

> *"PISA 2015 focused on science learning in school by including several questions about the learning environment in the science classroom. They asked how often specific activities happened in the school science course."*

With ST113, it is aimed to measure the perspective of in-school scientific issues of the students.

The Instrumental Motivation to Learn Science is one of the subscale of the measurement tool related to the Disciplinary Climate in Science Classes. Items are presented in Table 2.

Table 2. *Items of Instrumental Motivation to Learn Science*

Code	Item
ST113Q01TA	Making an effort in my <school science> subject(s) is worth it because this will help me in the work I want to do lat
ST113Q02TA	What I learn in my <school science> subject(s) is important for me because I need this for what I want to do later on
ST113Q03TA	Studying my <school science> subject(s) is worthwhile for me because what I learn will improve my career prospects.
ST113Q04TA	Many things I learn in my <school science> subject(s) will help me to get a job.

2.3. Data Analysis

In the data analysis process, missing values and extreme values were examined and whether or not the assumptions of MG-CFA which were normality, linearity, homoscedasticity and multicollinearity were provided for every single country. The amount of missing values in all data set was below 5% thus it was deleted.

In this study, to set an example for practitioners, the approximate measurement invariance across the countries was determined by the alignment method. The analyzes of the Alignment method for approximate measurement invariance and MG-CFA that tested the exact measurement invariance were performed in the Mplus 7.1 program. The Mplus script of the alignment analysis is attached in the appendix.

In the analysis of measurement invariance with MG-CFA, model fit was examined by taking into consideration the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR). Sokolov (2019) stated that RMSEA and TLI values give erroneous results in determining the measurement invariance with MG-CFA especially when testing metric invariance and CFI and SRMR values should be taken into consideration in such studies. Cut-off criteria for the fit indexes recommended by Sokolov (2019) for use in many group comparisons are presented in Table 3.

Table 3. *Cut-off criteria of Goodness of Fit Indexes for Multiple Group Measurement Invariance Comparisons*

	Cut-off Values			Relative Cut-off Values	
Fit Index	Configural inv.	Metric inv.	Scalar inv.	Configural inv.	Metric inv.
CFI	>0.985	>0.980	>0.970	>-0.010	>-0.010
TLI	-	-		-	>-0.005
RMSEA	-	-		-	<0.005
SRMR	<0.020	<0.040	<0.045	<0.010	<0.010

With the alignment procedure in the Mplus 7.1, the measurement invariance is tested with an algorithm based on the calculation of the largest number of groups where the difference between the parameters is not significant by making binary parameter comparisons. The table comparing factor means and variances for all groups shows on the top in the Mplus output file. The countries/groups which measurement invariance is not provided significantly are shown in bold in brackets. Another output is the table in which the factor means of all countries are ordered from high to low and the significant differences across them are shown through the z test. In addition, the contribution of each item's interceps and factor loading to the optimized simplicity function is calculated by the measure of R^2. The R^2 is a useful descriptive statistic that gives

the degree of noninvariance that can be absorbed by group-varying factor means and variances (Muthén & Asparouhov, 2014; 2018). In the configural model, it shows how much of the parameter variation across groups for each measurement parameter can be explained by factor means and variation in factor variances. R^2 value close to "1" implies a high degree of invariance and to "0" a low degree of invariance.

3. FINDINGS

3.1. Results of Multi-Group Confirmatory Factor Analysis

The analyzes carried out with MG-CFA was stopped at the stage where measurement invariance were not provided. Therefore, the results given in Table 4 belong to the results of configural and metric invariance.

Table 4. *Configural and metric invariance results*

	Models	χ^2	df	p	CFI	SRMR
	Configural	21055.716	114	0.000	0.982	0.018
ST113	Metric	26378.988	282	0.000	0.978	0.046
	Configural vs metric	5323.272	168	0.000	-0.004	0.028

According to Table 4, the CFI is 0.003 points below the cut-off value proposed by Sokolov (2019). On the other hand, it is seen that SRMR is lower than the accepted cut-off value of 0.02. Considering these two values together, it can be said that the configural model is provided for 57 countries. After determining that configural model invariance was provided, analyzes were carried out for the metric invariance stage. The CFI and the SRMR, calculated at the metric invariance stage were outside the recommended cut-off value in Table 3 (CFI> 0.98, SRMS <0.04). When the relative cut-off values are examined, it is seen that the cut-off value of ΔSRMR is greater than 0.01 and the cut-off value of ΔCFI is less than -0.01. Considering the cut-off values of fit indexes and their relative cut-off values, it was concluded that the metric invariance was not provided. In addition, the chi-square test (Fan & Sivo, 2009), which is used as a complement to alternative fit indexes in testing measurement invariance, was also examined and it was found that the difference between chi-square fit index values was significant. Therefore, it was supported that the metric invariance was not provided with the chi-square test.

In the examinations, it has been determined that the measurement invariance is impaired in the weak invariance stage. However, for the 57 countries, it is not possible to determine across which countries the measurement invariance is impaired by MG-CFA. To reach this information, a double comparison of 57 groups (57x56/2=1.596) should be done, which will bring the work and time load to a high level. At this point, the alignment method was used to determine for which countries the measurement invariance was provided/impaired at the item level. With this method, the measurement invariance of the items on the scale has been revealed.

3.2. Results of the Alignment Method

In this section, it is shown how the alignment method solves the problem of comparing the factor means found by traditional multi-group factor analysis under scalar invariance. Maximum Likelihood estimation was used for measurement invariance analysis. In the alignment analysis, since the factor mean of the 29th country (Lithuania) is closest to zero, FIXED alignment optimization was used and this country was taken as the reference country whose factor mean was restricted to zero. Table 5 shows each item's intercept and factor loading values for 57 countries.

Table 5. *Each item's Alignment results of 57 countries*

Item	Intercepts	Factor Loadings
ST113.1	1 2 3 **(4)** **(5)** 6 7 8 **(9)** 10 **(11)** 12 13 **(14)** 15 16 **(17)** **(18)** 19 20 **(21)** **(22)** **(23)** 24 **(25)** **(26)** **(27)** 28 **(29)** 30 31 **(32)** **(33)** 34 35 **(36)** 37 38 **(39)** 40 41 42 43 **(44)** **(45)** **(46)** 47 **(48)** 49 50 **(51)** **(52)** **(53)** 54 **(55)** 56 57	1 2 3 4 **(5)** 6 **(7)** 8 9 10 **(11)** 12 13 **(14)** 15 16 17 **(18)** 19 **(20)** **(21)** **(22)** 23 24 **(25)** **(26)** 27 28 **(29)** 30 **(31)** 32 **(33)** **(34)** **(35)** 36 **(37)** 38 **(39)** **(40)** **(41)(42)** 43 **(44)** **(45)** **(46)** 47 **(48)** **(49)** **(50)** **(51)** **(52)** **(53)** **(54)** **(55)** **(56)** 57
ST113.2	**(1)** **(2)** 3 4 5 6 **(7)** **(8)** **(9)** 10 **(11)** **(12)** **(13)** **(14)** **(15)** **(16)** **(17)** 18 19 20 21 22 23 **(24)** **(25)** **(26)** **(27)** **(28)** **(29)** 30 31 **(32)** **(33)** 34 **(35)** 36 **(37)** **(38)** 39 40 **(41)** 42 43 **(44)** 45 **(46)** 47 **(48)** 49 **(50)** **(51)** 52 53 54 55 56 57	**(1)** **(2)** 3 4 5 6 7 8 **(9)** 10 11 12 **(13)** **(14)** 15 **(16)** **(17)** 18 19 **(20)** **(21)** **(22)** **(23)** 24 **(25)** **(26)** **(27)** **(28)** **(29)** 30 31 32 **(33)** 34 35 **(36)** 37 38 **(39)** **(40)** **(41)** 42 43 **(44)** 45 46 47 **(48)** 49 50 **(51)** 52 53 **(54)** 55 56 57
ST113.3	**(1)** 2 3 4 **(5)** 6 7 8 9 10 **(11)** **(12)** **(13)** 14 15 16 **(17)** 18 19 20 21 **(22)** **(23)** 24 25 **(26)** 27 **(28)** **(29)** 30 **(31)** 32 **(33)** 34 35 36 37 **(38)** **(39)** **(40)** **(41)** 42 43 **(44)** 45 **(46)** 47 **(48)** **(49)** 50 51 **(52)** **(53)** **(54)** 55 56 **(57)**	1 2 3 **(4)** **(5)** **(6)** **(7)** **(8)** **(9)** 10 **(11)** 12 13 14 **(15)** 16 17 **(18)** 19 20 21 **(22)** 23 **(24)** **(25)** **(26)** 27 **(28)** **(29)** 30 **(31)** **(32)** **(33)** **(34)** 35 36 37 **(38)** **(39)** **(40)** **(41)** 42 **(43)** **(44)** **(45)** **(46)** 47 **(48)** 49 **(50)** **(51)** 52 **(53)** **(54)** **(55)** 56 57
ST113.4	**(1)** **(2)** 3 **(4)** 5 6 **(7)** 8 **(9)** **(10)** **(11)** 12 13 **(14)** 15 **(16)** **(17)** **(18)** 19 **(20)** **(21)** 22 **(23)** **(24)** 25 **(26)** **(27)** 28 **(29)** 30 31 **(32)** 33 34 **(35)** 36 **(37)** 38 **(39)** **(40)** 41 42 43 **(44)** **(45)** 46 **(47)** **(48)** **(49)** 50 **(51)** **(52)** 53 54 **(55)** 56 **(57)**	1 2 3 **(4)** 5 6 **(7)** 8 **(9)** **(10)** 11 **(12)** 13 14 15 **(16)** **(17)** **(18)** 19 20 21 22 23 24 **(25)** **(26)** 27 28 **(29)** 30 31 **(32)** **(33)** 34 35 **(36)** 37 38 39 **(40)** **(41)** 42 43 **(44)** **(45)** **(46)** **(47)** **(48)** **(49)** 50 51 **(52)** 53 54 **(55)** 56 57

Table 5 shows the findings regarding the invariance of the intercepts and factor loading values of the ST113 coded questionnaire for 57 countries. The results of the alignment analysis are interpreted as the fact that the intercepts and factor loading values differ significantly across the groups (countries) in parentheses. Thus, factor loadings and factor intercepts can be compared across countries which are within the parentheses. For example, the factor loading of the first item does not differ significantly for the countries coded as 1, 2, 3, 4, 6, 8, 9, 10, 12, 13, 15, 16, 17, 19, 23, 24, 27, 28, 30, 32, 36, 38, 43, 47, 57. Metric invariance has been provided for the first item across these countries. Therefore, factor variances and structural relationships can be compared across groups for this item. Also this item does not signiificantly differ from both the intercepts and loading among the countries coded as 1, 2, 3, 6, 8, 10, 12, 13, 15, 16, 19, 24, 28, 30, 38, 43, 47, 57. Accordingly, it can be said that scalar invariance that assuming both intercepts and factor loading parametres are equivalent across groups has been provided across these countries. Therefore, it is possible to compare factor means and intercepts across these countries.

Table 6 shows the coefficient of fit function and R^2 values of each item, which shows how much items contribute to the optimized simplicity function.

Table 6. *Alignment Fit Statistics*

Items	Intercepts		Factor Loadings	
	Fit Function Contribution	R^2	Fit Function Contribution	R^2
ST113.1	-536.658	0.964	-551.396	0.987
ST113.2	-535.435	0.970	-526.201	0.935
ST113.3	-530.307	0.972	-534.924	0.901
ST113.4	-557.961	0.960	-532.381	0.876

When Table 6 is examined, it can be said that all the items of the ST113 questionnaire contributed similarly to the simplicity function. This finding shows that the degree of noninvariance is similar. The R^2 results presented in Table 6 are interpreted in the configural model as able to explain the variation in the intercepts and factor loading values predicted for all groups with the variation in factor means and variances among all groups. The R^2 values of the item coded ST113.4 indicate that the item contributed the least to the simplicity function. In other words, this item has most degree of noninvariant across the groups. Table 7 shows the factor means estimated for all groups by the alignment method and groups that have factor means significantly different on the 0.05 level.

Table 7. *Comparison of Factor Means between Countries*

Ranking	Group	Factor Means	Groups with Significantly Smaller Factor Mean
1	19	.753	47 13 22 30 26 4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
2	3	.729	13 22 30 26 4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
3	34	.715	13 22 30 26 4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
4	47	.692	13 22 30 26 4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
5	13	.581	22 30 26 4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
6	22	.535	4 18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
7	30	.520	18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
8	26	.502	18 27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
9	4	.494	27 43 44 14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
10	18	.455	14 28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
11	27	.441	28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
12	43	.431	28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
13	44	.418	28 25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
14	14	.412	25 36 38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
15	28	.376	38 12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
16	25	.363	12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
17	36	.355	12 2 17 6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
18	38	.330	6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
19	12	.314	16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
20	2	.313	6 16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
21	17	.303	16 31 23 55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
22	6	.277	55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
23	16	.259	55 21 41 46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
24	31	.242	46 9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
25	23	.239	9 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
26	55	.230	9 45 56 20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
27	21	.213	20 53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50

Ranking	Group	Factor Means	Groups with Significantly Smaller Factor Mean
28	41	.206	53 39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
29	46	.200	39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
30	9	.192	39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
31	45	.191	39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
32	56	.169	8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
33	20	.168	39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
34	53	.158	39 52 8 57 10 33 24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
35	39	.117	24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
36	52	.116	24 1 35 51 11 29 7 5 37 42 32 54 40 48 49 15 50
37	8	.097	51 11 29 7 5 37 42 32 54 40 48 49 15 50
38	57	.094	11 29 7 5 37 42 32 54 40 48 49 15 50
39	10	.090	51 11 29 7 5 37 42 32 54 40 48 49 15 50
40	33	.079	11 29 7 5 37 42 32 54 40 48 49 15 50
41	24	.071	11 29 7 5 37 42 32 54 40 48 49 15 50
42	**1**	**.068**	**11 29 7 5 37 42 32 54 40 48 49 15 50**
43	35	.064	29 7 5 37 42 32 54 40 48 49 15 50
44	51	.052	29 7 5 37 42 32 54 40 48 49 15 50
45	11	.022	7 5 37 42 32 54 40 48 49 15 50
46	29	.000	7 5 37 42 32 54 40 48 49 15 50
47	7	-.032	37 42 32 54 40 48 49 15 50
48	5	-.050	37 42 32 54 40 48 49 15 50
49	37	-.086	48 49 15 50
50	42	-.091	48 49 15 50
51	32	-.096	49 15 50
52	54	-.100	49 15 50
53	40	-.107	15 50
54	48	-.119	15 50
55	49	-.127	50
56	15	-.161	
57	50	-.197	

For convenience of the presentation, the groups are ordered from high to low according to factor means and the groups that have factor means that differ on the 0.05 significance level are determined. For example, as seen in Table 7, the factor means of the 19th country estimated by the alignment method is 0.753 and this value of the 19th country is significantly higher than the countries whose codes written in the last column.

In exact MI framework, scalar invariance assumes that both the factor loadings and the measurement intercept are invariant among the groups if and only it is possible to compare factor means and variances among the groups. Especially, when there are many groups, scalar invariance rarely fits to the data set (Muthén and Asparouhov, 2018). In the Alignment method, the most appropriate measurement invariance pattern is discovered in which the factor means and factor variances of the groups are comparable. In this method, the maximum number of groups in which the factor means and factor variances across the groups do not differ statistically are estimated. In this way, it is revealed that among which groups that all items of the scale are comparable. Thus, MI are determined in not only item-based level but also scale-based level. Specifically for Turkey, is factor mean is 0.068, ranks 42nd out of 57 countries when ranked from high to low. When all statements taken together, while the factor means of the countries which coded 8 (Chile), 57 (North Carolina), 10 (Colombia), 33 (Serbia) and 24 (Ireland) coded countries' factor means are larger than Turkey, it is smaller for 35 (New Zealand) and 51 (United kingdom). However, the factor mean differences are not statistically significant between Turkey and these countries. This table shows that approximate measurement invariance is provided between Turkey and Chile, North Carolina, Colombia, Serbia, Ireland, New Zealand and the United kingdom. By means of the Alignment method, it is determined with a one-step analysis which countries factor means are comparable of each country included in the analysis.

Figure 1, Figure 2 and Figure 3 show that scatter diagrams between factor means obtained in

the alignment method of countries and factor means obtained in the scalar invariance stage of MG-CFA. The scattering values for all countries are shown in Figure 1. In Figure 2 and Figure 3, the scattering values for the countries are presented on a larger scale by cutting-off the graph in Figure 1 at the level of (0.2, 0.1).

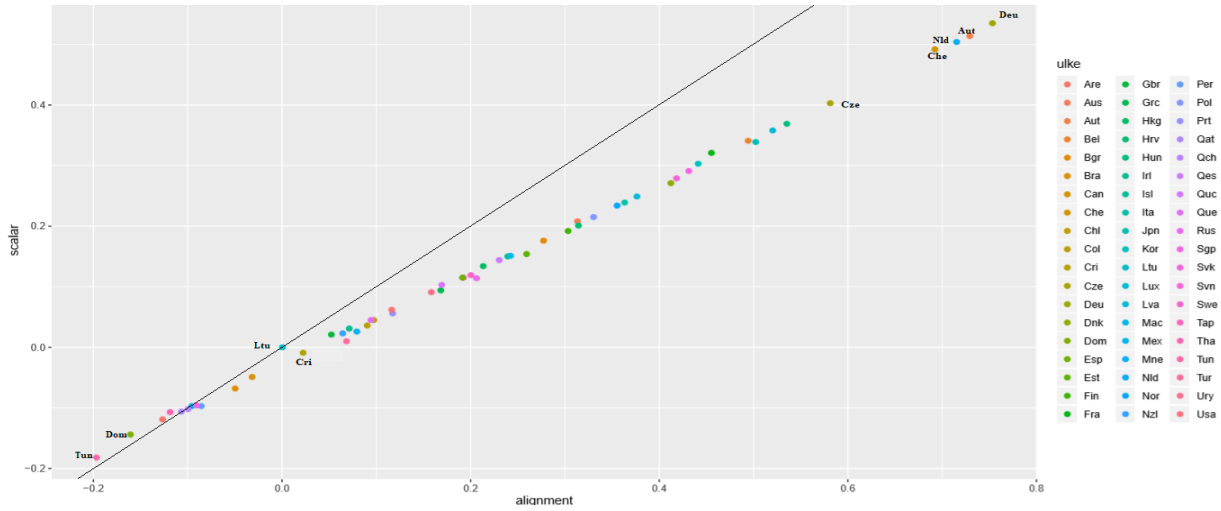

Figure 1. *Factor means obtained from the alignment method versus scalar model (57 countries)*

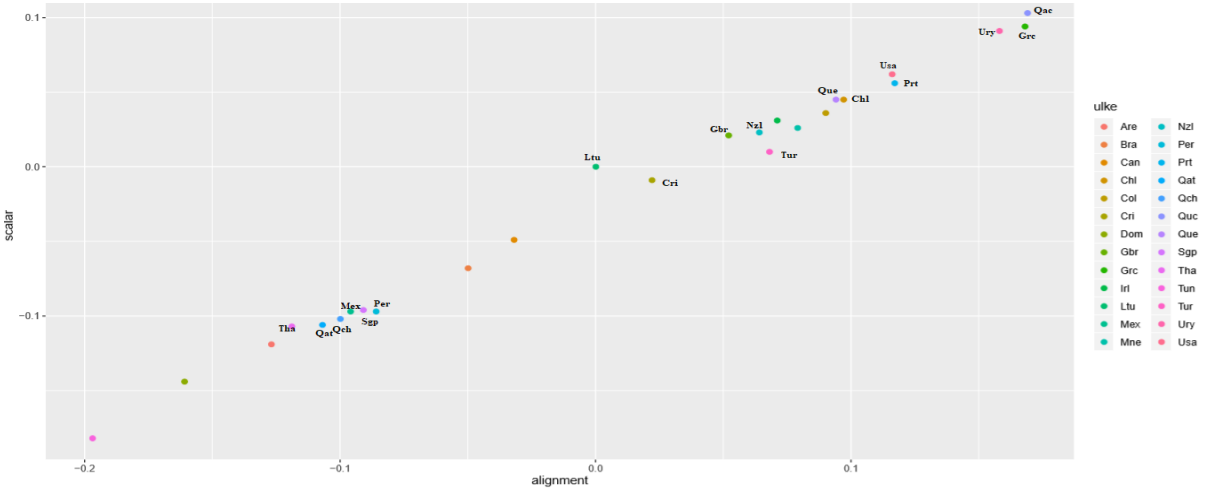

Figure 2. *Factor means obtained from the alignment method versus scalar model (26 countries)*

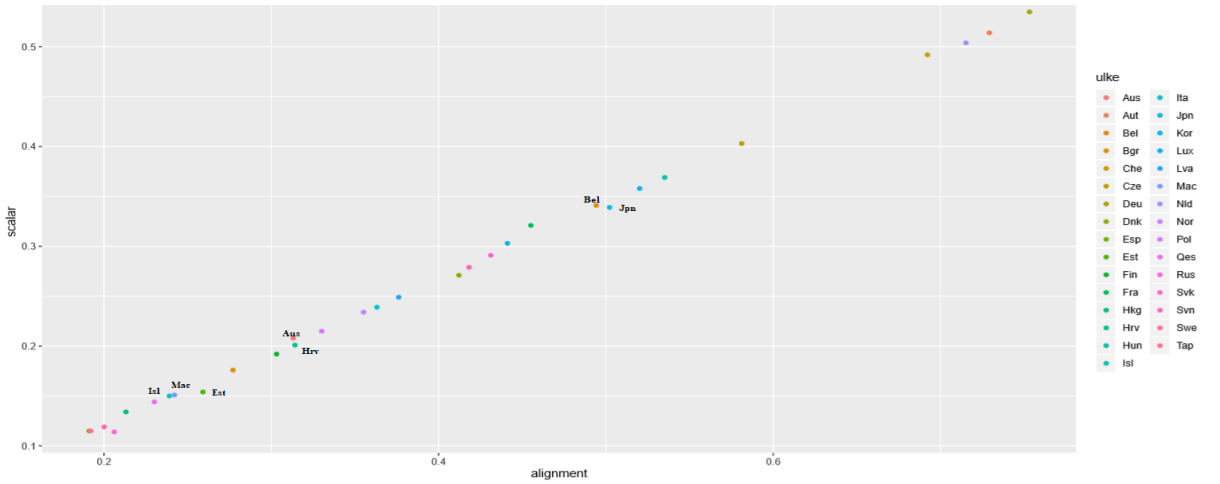

Figure 3. *Factor means obtained from the alignment method versus scalar model (31 countries)*

Figure 1 shows the scatter diagram between the factor means obtained in the alignment method of 57 countries and the factor means obtained from the scalar invariance model of MG-CFA. The fit is most impaired for Germany (Deu, 19); least impaired for Mexico (Mex, 32) and China (Qch, 57) in both methods. The correlation between the factor means obtained from the alignment method and scalar invariance was calculated as 0.999. Despite this high correlation, there are some differences between the two methods. For example; although there are no significant differences between the factor means of Turkey-Lithuania and Turkey-Costa Rica in scalar invariance, Turkey's factor mean calculated in the alignment method is significantly higher than factor means of Lithuania, and Costa Rica. Similarly, in the studies conducted by Asparouhov and Muthén (2014), Muthén and Asparouhov (2018) and Marsh, Guo, Parker, Nagengast, Asparouhov, Muthén and Dicke (2018), there was a high correlation between factor means predicted by strong invariance and the alignment method. But, there was no significant difference for factor means of some countries in scalar invariance while there was a significant difference between the factor means of these countries in the alignment method.

4. DISCUSSION and CONCLUSION

Traditional measurement invariance methods, also known as exact measurement invariance, are inadequate especially in studies comparing large numbers of groups. When the literature is reviewed, it is seen that MG-CFA is frequently preferred in the determination of measurement invariance based on structural equation modeling. However, in studies comparing many groups by MG-CFA, there is almost no study in which full measurement invariance is provided. In these studies, it is generally reported at which stage the measurement invariance is impaired, and the situations where partial measurement invariance is provided with the proposed modifications. While in some studies, no invariance was provided in any model (Gülleroğlu, 2016), in some it was observed that only configural invariance was provided (Hansson & Gustafsson, 2013; Sırgancı & Çakan, 2020). In some of them, it was determined that metric invariance was hold (Asil & Brown, 2015; Pauwels, 2018; İmrol, 2017; Luo, 2010). It has been reported that scalar invariance is held in a few studies (Uzun & Öğretmen, 2010) and rare of the studies, strict invariance is provided (Wu, Li & Zumbo (2007). The presence of many modifications in the MG-CFA both prolong the analysis time and there is no guarantee to determine the best-fit model. Data was collected from many countries with measurement applications such as PISA, TIMSS (The Trends in International Mathematics and Science Study), ESS (The European Social Survey). Due to the reasons mentioned above, in the studies comparing many cultures in large data, it is quite difficult to determine the exact measurement invariance or to determine in which cultures the measurement invariance is provided. Therefore, it is concluded that cross-cultural comparisons will not be valid since measurement invariance is not provided in the studies. However, the alignment method proposed in this study determines the approximate measurement invariance without requiring strict measurement invariance by estimating the group-specific factor mean and variance. The Alignment method is a powerful method to predict the model for multidimensional structures or multiple indicators (items). In this respect, it has essential advantages against MG-CFA method.

Alignment method has main three advantages compared other MI methods in the studies by using multicultural database such as TIMSS, PISA. Firstly, the maximum number of groups in which the measurement invariance is ensured at both item and scale levels can be determined by a single analysis. Secondly, in item-based analysis, the factor load and the intercept of each item can be pairly compared across countries in a one step analysis and also how much each item has contributed to the measurement invariance is determined. Thirdly, it is determined whether the factor means show a significant difference between which countries, and thus, cross-country comparability at the scale level is determined.

In this study, which was carried out using the data of 406.961 participants from 57 countries

participating in PISA 2015, it was found that only configural invariance was provided in the investigations analyzed with MG-CFA. Uyar and Uyanık (2019) examined the measurement invariance of the science learning model constituted of measurement instruments in PISA 2015 questionnaires for Turkey and Singapore. As a result, it was found that only configural invariance was provided among these countries. It was concluded that the comparison of the item-scores obtained from the groups may be biased because they did not respond similarly to the items, thus the relationship between measured properties and dimensions of scale is not similar when compared Turkey and Singapore.In this study, the alignment analysis findings show that the intercept and load parameter of the items coded as (ST113.3) and (ST113.4) of the Instrumental Motivation Scale in science learning were equal between Turkey (1) and Singapore (42). In framework of exact MI, it means that the metric invariance is provided for these two items between Turkey and Singapore. Besides, there is no significant difference between the factor means of these two countries. This finding shows that scalar invariance is provided within the framework of exact measurement invariance between these two countries on the scale. Therefore, contrary to the findings of Uyar and Uyanık (2019), the measurement invariance results with the alignment method showed that the scale scores were comparable between these two countries. The findings of MG-CFA of this study showed that only configural invariance was provided for 57 countries. This finding is consistent with the study findings. However, the alignment analysis findings show that the metric invariance is also provided for in the third (ST113.3) and the fourth item (ST113.4) of the Instrumental Motivation Scale in science learning for Turkey (1) and Singapore (42). On the other hand, there is no significant difference between the factor means of these two countries.

In the study conducted by Tiryaki (2019), the measurement invariance of the scales measuring students' attitudes towards science for Turkey and the USA were investigated. When the model fit indexes for ST113 scale were examined, it was found that all invariance stages (configural, metric, scalar and strict invariance) were provided. In this study, the researchers reported that the intercepts and factor loadings were invariance for Turkey and the USA. Besides it was stated that the responses of the items were similar in terms of these two cultures, and the difference in the scores was due to the subgroups. In the same study, it was found that all the items in ST113 had DIF according to the Likelihood Ratio Test based on IRT. However, it was found that factor loadings of the item coded as ST113.3 is invariance for Turkey and the USA in this study differently from Tiryaki (2019). This means that the metric invariance is provided for this item. For other items in ST113, the measurement invariance is not provided for Turkey and the USA. It has also determined the factor means of the USA is significantly higher than Turkey's.

Gür (2019) compared in respect to measurement invariance England-Ireland, England-USA and England-Turkey by using generalized Mantel-Haenszel, poly-SIBTEST and ordinal logistic regression. For England-Ireland (same language-similar culture), the first item had DIF according to GMH, the first and second items had DIF according to OLR and the first, the second and the fourth items had DIF according to poly-SIBTEST. For England-USA (same language-different culture) all of the items had DIF according to OLR, the second, the third items had DIF according to poly-SIBTEST, and the first, the second and the third items had DIF according to GMH. For England-Turkey (different language and culture), it was detected that all items had DIF according to OLR and the second, the third and the fourth items had DIF according to poly-SIBTEST and GMH. In this study, it is concluded that the only factor loadings of the fourth item in ST113 is comparable for England-Ireland and England-Turkey. It means that the metric invariance is provided for this item for these samples. In addition, the factor loadings and measurement intercepts of the all items significantly differentiate for England-USA and the factor means of USA is significantly higher than England's. Besides scalar invariance is provided for the first item and metric invariance is provided for the fourth item for Turkey-Ireland. The factor means of these countries do not differ significantly.

When the findings of this study and the studies mentioned above are evaluated together, it was seen that the alignment analysis presented information that is more detailed as opposed to the traditional methods about measurement invariance between countries. Generally, the configural invariance is provided in studies where full measurement invariance is tested. This and other studies show that when the cultural differences increase, the measurement invariance is impaired (Asil & Gelbal, 2012; Ercikan & Koh, 2005; Kıbrıslıoğlu, 2015; Yandı, Köse & Uysal, 2017). In these studies, pairwise comparisons of the groups predicted to reflect cultural differences made by the researchers. However, it is discovered between which groups the measurement invariance is provided with an exploratory approach with the alignment method. In addition, information is provided on how much each item in the measurement instrument contributes to the invariance between comparison groups. In the case of exact measurement invariance studies, only in case of scalar invariance, which is very rare, factor means can be compared between groups. In the alignment method in which the approximate measurement invariance is tested, it is calculated between which groups the factor means differ statistically, so it is possible to compare factor means between the groups. Thus, the problem of comparability of factor means encountered in scalar invariance in traditional MG-CFA is solved with the alignment method thanks to the factor means estimated for all groups.

Acknowledgements

A part of this study was presented as an oral presentation at the 11th International Statistics Congress (4-8 October 2019, Turkey, Bodrum) with the title "Alignment Method: Concept and Application".

Declaration of Conflicting Interests and Ethics

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

ORCID

Gözde SIRGANCI https://orcid.org/0000-0003-4824-5413
Gizem UYUMAZ https://orcid.org/0000-0003-0792-2289
Alperen YANDI https://orcid.org/0000-0002-1612-4249

5. REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Asil, M., & Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim, 37*(166), 236-249.

Asil, M., & Brown, G. T. L. (2015). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing, (16)*1, 71-93. https://doi.org/10.1080/15305058.2015.1064431

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508. https://doi.org/10.1080/10705511.2014.919210

Byrne, B.M., Shavelson, R.J., & Muthén, B.O. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items* (4th ed.). Thousand Oaks, California, USA: Sage Publications.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*(4), 662-80.

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

Fan, X., & Sivo, S. A. (2009). Using Δ-goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(1), 54-69, https://doi.org/10.1080/10705510802561311

Gülleroğlu, H. D. (2017). PISA 2012 matematik uygulamasına katılan türk öğrencilerin duyuşsal özelliklerinin cinsiyete göre ölçme değişmezliğinin incelenmesi, *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 37*(1), 151-175.

Gür, E. (2019). *PISA 2015 uygulamasındaki maddelerin kültüre göre değişen madde fonksiyonu acısından incelenmesi* (Master's thesis). Hacettepe University, Graduate School of Educational Sciences, Ankara, Türkiye.

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

İmrol, F. (2017). Pisa 2012 *Türkiye örnekleminde matematiğe yönelik motivasyon ve öz-inanç yapılarının ölçme değişmezliğinin incelenmesi* (Master's thesis). Ankara University Graduate School of Educational Sciences, Ankara, Türkiye.

Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika, 71*(1), 173-191. https://doi.org/10.1007/s11336-003-1136-B

Jöreskog, K.G., Sörbom, D., Du Toit, S.H.C., & Du Toit, M. (2001). *LISREL 8: New statistical features* (3rd ed.). Lincolnwood, IL: Scientific Software International.

Kıbrıslıoğlu, N. (2015*). PISA 2012 matematik öğrenme modelinin kültürlere ve cinsiyete göre ölçme değişmezliğinin incelenmesi: Türkiye, Çin (Şangay)-Endonezya örneği.* (Master's thesis, Hacettepe University, Graduate School of Educational Sciences, Ankara, Türkiye). Retrieved from http://openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/1843/70bd9399-b70f-414a-a11f-b81218adb77c.pdf?sequence=1

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: a comparison of five approaches, *Structural Equation Modeling: A Multidisciplinary Journal, 24*(4), 524-544. https://doi.org/10.1080/10705511.2017.1304822

Luo, C. (2010). *Measurement invariance of Rosenberg Self-Esteem Scale between British and Chinese college students.* (Master's thesis The University of Edinburgh, Edinburgh, Scotland). Retrieved from https://pdfs.semanticscholar.org/274b/015fedaa0c835d94680df2a3c153aff36e5a.pdf

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). *What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. Psychological Methods, 23*(3), 524-545. https://doi.org/10.1037/met0000113

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *ETS Research Report Series, 2*, i-28. https://doi.org/10.1002/j.2333-8504.1994.tb01618.x

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: alignment and random effects. *Sociological Methods & Research, 47*(4) 637-664. https://doi.org/10.1177/0049124117701488

Pauwels, J. (2018) Do you really measure the same? (Doctoral dissertation, Ghent University, Gent, Belgium). Retrieved from: https://lib.ugent.be/fulltxt/RUG01/002/481/799/RUG01-002481799_2018_0001_AC.pdf

OECD (2016). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264267510-en

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

Sırgancı, G., & Çakan, M. (2020). Sirali lojistik regresyon ve poly-sibtest yöntemleri ile değişen madde fonksiyonunun belirlenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *20*(1), 705-717.

Sokolov, B. (2019). *Sensitivity of goodness of fit indices to lack of measurement invariance with categorical indicators and many groups*. Higher School of Economics Research Paper No. WP BRP 86/SOC/2019. Retrieved from https://wp.hse.ru/data/2019/07/09/1480015921/86SOC2019.pdf

Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371–384.

Tiryaki, F. (2019). *PISA 2015 öğrenci tutum anketlerinin değişen madde fonksiyonu ve ölçme değişmezliğinin incelenmesi* (Master's Thesis). Ankara University Graduate School of Educational Sciences, Ankara, Türkiye

Uyar, Ş. (2011). PISA 2009 Türkiye örnekleminde öğrenme stratejileri modelinin farklı gruplarda ölçme değişmezliğinin incelenmesi (Master's Thesis). Hacettepe University Graduate School of Educational Sciences, Ankara, Türkiye

Uyar, Ş., & Uyanık, G. K. (2019). Fen bilimlerine yönelik öğrenme modelinin ölçme değişmezliğinin incelenmesi: Pisa 2015 örneği. *Kastamonu Eğitim Dergisi, 27*(2), 497-507. https://doi.org/10.24106/kefdergi.2570

Uzun, B., & Öğretmen, T. (2010). Fen başarısı ile ilgili bazı değişkenlerin timss-r türkiye örnekleminde cinsiyete göre ölçme değişmezliğinin değerlendirilmesi. *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*, 4-70. https://doi.org/10.1177/109442810031002

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* (p. 281–324). Washington DC, USA: American Psychological Association. https://doi.org/10.1037/10222-009

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial ınvariance and updating the practice of multigroup 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. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 13*(1), 243-253. https://doi.org/10.17860/mersinefd.305952

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (Ordinal) item scores.* Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

6. APPENDIX

The Mplus script of the alignment analysis:

```
TITLE: align MODEL
DATA: file is C:\Users\Lenovo\Documents\GGA\2- ST113\st113.dat;
VARIABLE:
variable:
    NAMES ARE country clus u1 u2 u3 u4;
    USEVARIABLES ARE u1 u2 u3 u4;
    MISSING=ALL (999);
    classes = c(57);
    knownclass = c(country = 1 2 3 4 5 6 7 8 9 10 11 12
    13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
    31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
    49 50 51 52 53 54 55 56 57);
ANALYSIS:
    type = mixture;
    estimator = ml;
    alignment = fixed(29);
    astarts=100;
model:
    %overall%
    st113 by u1-u4;
output:
    tech1 tech8 align;
plot:
    type = plot2;
```