International Journal of Assessment Tools in Education
2021, Vol. 8, No. 1, 90–105
https://doi.org/10.21449/ijate.690576
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

Investigation of measurement invariance of PISA 2015 collaborative problem solving skills: Turkey, Norway and Singapore

Yusuf Taner Tekin [1,*], **Derya Cobanoglu Aktan** [1]

[1]Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey

ARTICLE HISTORY

Received: Feb 18, 2020
Revised: Sep. 08, 2020
Accepted: Jan. 06, 2021

Keywords

PISA Collaborative problem solving skills, Measurement invariance, Multi-group confirmatory factor analysis, Cultural comparisons

Abstract: The purpose of this research is to examine measurement invariance of collaborative problem solving skills measured by PISA 2015 Xandar subtest for Singapore, Norway, and Turkey. The research was conducted with 2990 participants' data obtained from Turkey (1032), Norway (923), and Singapore (1035) on PISA 2015 collaborative problem solving study. In the first part of the study, exploratory factor analysis was performed to obtain the factor structure of the Xandar subtest. Then, the model data fit was checked by confirmatory factor analysis via χ^2 / *df* (3.127), RMSEA (0.027), CFI (0.987) and TLI (0.979) values. Multi-group confirmatory factor analysis was used in invariance analyses. The findings show that the collaborative problem solving model met only configural invariance across the countries and has not met the metric, scale, and strict invariance stages. The results show that meaningful comparisons cannot be made between the countries, because the factor loadings, variances, error variances, and covariances differ among countries.

1. INTRODUCTION

Global developments, demographic changes, and technological progress require certain changes in individuals' lives and and specific skills are needed in every field. These skills include communication, teamwork, leadership, taking initiative, literacy in mother tongue and a foreign language, competence in science, mathematics, and problem solving. Having these skills will enable individuals to be more successful in their daily, business and social lives. The acquisition of the mentioned skills can occur spontaneously in social life and is also acquired through education. However, these skills to be acquired through educational institutions should be transferred to daily life situations. At the same time, it is necessary to measure the level of acquisition of these skills and to plan educational policies according to these measures.

International exams such as the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMMS) aim to assess the transfer of acquired knowledge and skills in the fields of science, mathematics, and reading to daily life situations. PISA which is implemented by the Organization for Economic Cooperation and Development (OECD), is a pioneer test in this field. Moreover, PISA has developed different

CONTACT: Yusuf Taner Tekin ✉ tanerrr.tekin@gmail.com Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Turkey

ISSN-e: 2148-7456 /© IJATE 2021

assessment applications in recent years. In 2012, PISA started to assess individuals' financial literacy skills and in 2015 collaborative problem solving skills. The reason for that is nowadays especially the labor market requires individuals who are in dialogue with others, can communicate and solve problems collaboratively. The increasing demand for highly qualified individuals also emphasizes those who have these skills. With this in mind, the results obtained from PISA, also provide resources for developing specific policies for the countries on the quality of their education and their students' collaborative problem solving skills.

1.1. Collaborative Problem Solving Skills

PISA 2015 defines collaborative problem solving as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution." (OECD, 2017, p.34). According to Demir and Seferoglu (2017) subjective structuring and transfer of knowledge, increasing emphasis on authentic learning and producing knowledge have led to the emergence of collaborative problem solving as well as problem solving skills. Nelson (2009) argued that collaborative problem solving has two structural components such as cooperative learning and problem-based learning. The author also states that collaborative problem solving provides students with experiences that create an intrinsic motivation for learning, questioning, collaborating and problem solving (Nelson, 2009). The nature of collaborative problem solving goes back to the work that O'Neil and his colleagues started in the 1990s to evaluate concepts in the best way and develop a theoretical framework and methodology. O'Neil, Chuang, and Chung (2003) have defined competencies similar to those used by PISA today. These competencies are grouped under five categories and expressed as the use of resources, interpersonal relations, information, systems, and technology. Thus, collaborative problem solving is process based on the contribution of both the cognitive and social skills of individuals involved in an activity (Hesse, 2017). In light of such developments, PISA implemented collaborative problem solving in 2015 and focused on solving the problem situations presented to individuals in a computer-assisted environment on a common understanding with one's teammates. It is meant that the computer accompanies the people participating in the application as virtual individuals (OECD, 2017). In the process of collaborative problem solving, PISA defined the following competences;

- establishing and maintaining shared understanding,
- taking appropriate action to solve the problem,
- establishing and maintaining team organization.

Furthermore, the capacity of the individual, doing the work with at least two or more people, and attempting to solve problems were identified as key competences (OECD, 2017). The theoretical development of the competencies identified by PISA is based on the topics of "computer-assisted collaboration, team discourse analysis, information sharing, individual problem solving, organizational psychology, and business context assessment".

Collaborative problem solving research gained popularity in recent years. The recent examination of the concept is closely related to the fact that it is one of the skills sought after today. A recent study (Erkoç, 2018) investigated the effects of collaborative game design on various skills (-settings-) such as critical thinking, problem solving and algorithm development. Erkoç (2018) found that there is a significant difference in terms of the problem solving skills in favor of the group in which the collaborative game development approach is applied. At the same time, it was observed that there was also a significant difference in favor of the collaborative group between self-control factor, which is one of the problem solving skills. Uzunosmanoğlu (2013) conducted a study on the computer-assisted collaborative problem solving processes with dual eye-tracking. The study was conducted with 18 university students and focused on the participants' ability to discuss geometry problems with their teammates

using a collaborative approach. When the results are obtained, it was seen that the team members who collaborate more often achieve better results than the team members who collaborate less. In another study conducted by Özdemir (2005), the effects of individual and collaborative problem-based learning on critical thinking skills, academic achievement and attitude towards internet use were examined among 70 university students. It was found that there was a significant difference between the scores of using critical thinking skills according to the students' groups, and the researcher reported that this difference was in favor of the collaborative group. The results of these and similar studies show that collaborative problem solving is important for solving complex problems and critical skill for individuals to have.

1.2. Measurement Invariance

In the PISA studies, it was found that students' achievements are associated with certain variables. These are the variables that can directly or indirectly affect the achievement of individuals such as their socio-economic status, equality of opportunity, time devoted to learning, future academic expectations, and pre-school education. However, when the results obtained from these variables are compared, it is not right to attribute the differences that arise only to the characteristics of individuals and to environmental factors. Because these differences among individuals may not stem from the individuals themselves, but also the measurement tool too. Even though the language experts in different countries have made efforts to eliminate language-related differences, it is not guaranteed that the measurement tool will have the same meaning and be interpreted by individuals in different countries (Başusta, and Gelbal, 2015). Hence, this situation will make it impossible to carry out generalizability studies on the groups for the measurement results.

It is not desirable that the other traits interfere with the measurement results other than the trait that is intended to be measured. Otherwise, this can cause validity problem for the measurement results. The items in measurement instruments are expected to be interpreted in the same way without being affected by the other variables. When the studies conducted in Turkey were examined, it was observed that the studies on measurement invariance have increased in recent years. Invariance means that measurements administered to the different groups show equivalent or similar psychometric properties (Başusta, Gelbal, 2015). Uyar (2011) conducted a measurement invariance study on gender, statistical area, and school types by using the learning strategies model for PISA 2009 Turkey data. Bahadır (2012), on the other hand, used structural equation modeling (SEM) to examine the differences among the seven geographical regions of the reading skills model of PISA 2009 and concluded that the model was in good agreement with the data. In another study, Başusta and Gelbal (2015) examined the factor structure of the science technology-related items in the PISA 2009 student questionnaire and they tested these factors for measurement invariance in terms of gender. Research on the measurement invariance (cultural and country invariance) studies for PISA tests were also reported in the literature. For instance, Kıbrıslıoğlu (2015) examined the measurement invariance based on culture and gender for PISA 2012 mathematics test for Turkey, China (Shanghai) and Indonesia data. The results of the research showed that the model holds the configural invariance stage among countries but does not hold the metric, scalar and strict invariance stages. Lately, Karakoç Alatlı (2016) studied the measurement invariance for Australia, France, Shanghai-China, and Turkey for PISA 2012 literacy test.

In terms of the measurement invariance studies that were carried out of Turkey, Greif, Wüstenberg, Molnar, et al (2013) studied the measurement invariance of complex problem-solving skills models over the grade level by using the Hungarian students' data. Oliden and Lizaso (2013), on the other hand, examined the measurement invariance of four different language forms, Spanish, Galician, Catalan, and Basque, on the data from the Spanish sample of PISA 2009 reading skills. Findings showed that the scores obtained from different language

forms do not exhibit invariance property. Another study by Wu, Liuand Zumbo (2007) tried to explain why the strict invariance stage is necessary for measurement invariance. For this purpose, the authors examined the countries such as the United States of America, Canada, Australia, New Zealand, and also the countries with similar cultures like Taiwan, Korea, and Japan by using TIMSS 1999 math tests. In this context, the researchers examined the measurement invariance by making 21 comparisons among and within various cultures.

The review of the previous studies shows that the measurement invariance is not always met and therefore before making comparisons, invariance studies should be performed. In particular, it is suggested and important to examine invariance if the results of the research are going to be/expected to be used in shaping educational policies. In addition, as in the case of PISA in the international arena, it is also necessary to show how different groups interpret the test applied to collaborative problem solving skills, which are among the essential critical skills of our time.

When a measurement tool is applied to groups with different characteristics, errors can be encountered in interpreting the results obtained if the effects of the demographic characteristics cannot be eliminated. However, errors encountered here cannot be attributed to only a single group membership. This could originate from the measurement tool. Cheung and Resvold (2002) state that differences can be explained not only by individual characteristics but also by measurement tools. The basic problem that is desired to be solved in the measurement invariance is whether the measurements of the same properties, measured with the same measurement tool, could change in different observations and working conditions of a given situation. If there is no such evidence of measurement invariance, it would not be right to make a scientific inference. In such a case, hence, it would not be correct to interpret the findings of differences between individuals and groups clearly (Mark and Wan, 2005). To make a comparison between groups on measurement results, measurement invariance must exist. To have a measurement invariance, the relationship between observed and latent variables must be the same in different groups (Karakoç Alatlı, 2016). According Millsap and Kwok (2004), to meet the invariance, the likelihood of getting a certain score is equal for individuals belonging to different groups whose similar characteristics are measured in the test. However, the most important feature sought in a measurement tool is validity and validity evidence. Therefore, accurate evidence on the validity of the scores obtained from measurement tools also necessitates measurement invariance studies (Yandı, Köse& Uysal, 2017).

Different definitions of measurement invariance can be found in the literature. Bryne and Watkins (2003) define invariance as the interpretation and perception of the scale by individuals of different groups in the same way. On the other hand, Raju, Laffittleand Byrne (2002) define invariance as getting the same score by different groups in terms of the characteristics measured by the scale. In other respects, measurement invariance can be realized in different cases or comparison of sub-sample groups of the same population. That is, the measurement invariance shows the comparability of the same structure in different cultural groups, the variance of the variables can be estimated independently from the group and the comparability of latent mean, variance and covariances of different groups (Bahadır, 2012). The comparisons here test the hypothesis of intergroup differences rather than the intra-group invariance of the model (Lance and Vandenberg, 2000). The main purpose of such studies is to use the measurements based on equality between groups. However, the measurement tools are prepared with the assumption of ‘different groups measure the same property’. If this assumption is confirmed, the accuracy of scoring and analysis will be meaningful. If this assumption cannot be verified, the analysis and the results obtained will lose their significance. In other words, the measurement model shows the same structure in more than one group. This means that the factor loadings of the scale

items, the correlations between the factors and the error variances are the same (Bollen, 1989; Byrne, 2004; Jöreskog & Sörbom, 1993).

Collaborative problem solving skills were highlighted and stated as critical skills for today's well educated students in the 2015 PISA assessment. Thus, collaborative problem solving skills were important components of the PISA 2015 collaborative problem solving test. From this point of view, it is critical to test the validity of the results and the comparability of the measurement model formed by the collaborative problem solving skills in the light of the PISA data, which offers a large sample and cross-country comparisons. The countries (Singapore, Norway, Turkey) in this study were selected based on their high, medium and low scores respectively in the PISA 2015 collaborative problem solving test. To solve the problems and sub-problems determined within the scope of this research, the steps of measurement invariance by Multi Group Confirmatory Factor Analysis (MGCFA) method were examined for the paired country groups respectively. So, we aimed to answer the problem of "do the collaborative problem solving PISA 2015 data hold the measurement invariance for the countries (Singapore, Norway, Turkey)?" Moreover, the following sub-problems were also examined in this study.

1) Do Singapore - Norway, Singapore - Turkey, and Norway - Turkey measurement models show;

(a) configural invariance,
(b) metric invariance,
(c) scalar invariance, and
(d) strict invariance?

2) If the invariance cannot be achieved, what are the relevant parameters for the invariance stages?

2. METHOD

This study is carried out to examine whether measurement invariance for collaborative problem solving PISA 2015 Xandar subtest data is met for Singapore, Turkey, and Norway groups. In this study, the data obtained from the OECD official website (https://www.oecd.org/pisa/data/2015database/) were used and no data collection was performed. According to the data characteristics, the research is a quantitative and a correlational study because it examines the relationship of observed variables with latent variables.

2.1. Data Characteristics

PISA 2015, the sixth round of PISA, was implemented in 2015 with the participation of approximately 540,000 students, representing approximately 29 million students in 72 countries and economies. 35 of these participating countries are OECD members. Within the scope of the study, the Xandar subtest, which is one of the six different subtests in which the cooperative problem solving skills are measured, was selected by purposeful sampling method as one of the non-probable sampling methods. As seen in Table 1 the number of individuals who answered the Xandar subtest was 1035 for Singapore, 923 for Turkey, and 1032 for Norway, and a total of 2990 individuals. Testing whether the measurement model created with collaborative problem-solving skills in the light of PISA data has the same structure for different countries will ensure the validity of the results and the significance of the comparisons. Here, we mean the countries with high, medium and low scores in the collaborative problem solving test scores of the PISA 2015 application. Therefore Singapore, Norway and Turkey have been selected.

Table 1. *Number of PISA 2015 and Xandar Subtest Participants by Country.*

	PISA 2015		Xandar Subtest	
Country	Number of participants	Percentage	Number of participants	Percentage
Norway	5.456	31.2	923	30.9
Singapore	6.115	35.0	1.035	34.6
Turkey	5.895	33.8	1.032	34.5
Total	17.466	100.0	2.990	100.0

In this study, the Xandar subtest was selected because its questions were published as examples, explanations were made according to the proficiency levels of these questions and had a sufficient sample size of data. The Xandar section starts with a general explanation. In this explanation, it is stated that each person will work with three teammates. However, the teammates expressed here are virtual persons. At the next stage, it is aimed to determine how individuals understand and solve the problem together with their team members in the face of three different situations. Following the instructions in the introduction of the test, participants are expected to proceed to the next stage by selecting one of the possible answers that appear on the screen, based on the comments they made by the virtual teammates. Here, one of the expressions chosen from the possible answers is correct (1) and the others are incorrect (0). The second screen, according to the individual's response to the event on the first screen and the views of the virtual persons about the event, appears on this screen. Thus, the individual completes the section each time by selecting one of the possible answers to continue the plot. More information about the Xandar subtest can be found on the OECD's website. (https://www.oecd.org/pisa/test/cps-xandar-scoring-guide.pdf).

Table 2. *Collaborative Problem Solving Competencies for Items.*

Item	CPS Skills	Description
m1	C3	Following rules of engagement (e.g., prompting other team members to perform their tasks)
m2	C1	Communicating with team members about the actions to be/being performed
m3	B1	Building a shared representation and negotiating the meaning of the problem (common ground)
m4	B1	Building a shared representation and negotiating the meaning of the problem (common ground)
m5	B3	Describing roles and team organization (communication protocol/rules of engagement)
m6	A1	Discovering perspectives and abilities of team members
m7	B3	Describing roles and team organization (communication protocol/rules of engagement)
m8	C3	Following rules of engagement (e.g., prompting other team members to perform their tasks)
m9	D1	Monitoring and repairing the shared understanding
m10	D2	Monitoring results of actions and evaluating success in solving the problem
m11	D3	Monitoring, providing feedback and adapting the team organization and roles

The Xandar subtest includes 12 items, but one of the items is scored differently from the "1-0" form and therefore it was not included in the study. The study was performed with 11 items. Table 2 contains the levels and descriptions of these items. They were coded as CC100101 in the original data set and these codes were changed to m1, m2, … and m11 for the convenience of the analysis and interpretation of the data. The collaborative problem solving competencies of these items are as follows:

- At level 1, the items (m2, m3, m4, m6, and m9), establishing and maintaining shared understanding
- At Level 2, item 10 (taking appropriate action to solve the problem)
- At level 3, the items (m1, m5, m7, m8, and m11), establishing and maintaining team organization.

Students who respond correctly to the items in level 1 are expected to explore the perspectives and abilities of their teammates, discuss a problem on shared ground, and communicate with

team members about the actions to be taken, and monitor and evaluate the actions they take in this direction. Students who answer the items in level 2 are expected to discover the type of communication they will perform to solve the problem, define the tasks to be completed, and monitor and evaluate the actions they perform as in the first level. Students who respond to the top 3 and top-level items correctly are expected to understand the roles for solving the problem, define the roles, follow the agreement rules set out in this direction, and follow and evaluate the team organization and roles, and give feedback.

2.2. Data Analysis

Measurement invariance is analyzed in stages. Four stages need to be tested to ensure that the invariance is fully achieved. These stages are configural invariance, metric invariance, scalar invariance, and strict invariance (Meredith,1993). Configural invariance; tests for identical factor structures for different groups; metric invariance checks equality of the factor loadings; scalar invariance tests equality of intersection points at regression equation; strict invariance refers to the invariance of residual load variance (Brown, 2015). The invariance stages were tested with the Mplus 7 analysis program and it was decided whether the invariance stages were achieved by taking the fit indices χ^2, RMSEA, CFI and TLI as reference. While conducting the MGCFA, one of the groups was taken as a basis and the values of the group were fixed at each stage, and the level of adaptation of the values of the other group to the fixed group was examined. The group whose values are kept constant is called the reference group, and one of the countries was chosen as the reference group in each analysis for the paired groups in the study. In addition to examining whether the fit indices are within the accepted range, the difference of CFI and TLI values compared to the less constrained model in the invariance stages were examined. If this difference is between -0.01 and 0.01, it has been taken into account that it is acceptable level for transition to the next stage (Cheung & Resvold, 2002). The invariance phases start with the structural invariance phase and if the fit indices are at an acceptable level, the next analyses were done. The level of change in chi-square, CFI and TLI values compared to the previous stage is discussed in the next stages after structural invariance. Before doing these analyses, the assumptions necessary for the analyses were checked. After that, the factor structure of the problem solving data were examined. After analyzing the factor structure, the collaborative problem solving model was confirmed by confirmatory factor analysis, and finally, the measurement invariance of the model was tested through Multiple Groups Confirmatory Factor Analysis (MGCFA).

In terms of assumptions, the missing values and multicollinearity were examined. For multicollinearity, tolerance and variance inflation were examined. After that exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed Kaiser-Meyer Olkin (KMO) and Barlett Sphericity Tests were used to investigate the suitability of the data set to EFA. In the CFA MGCFA model and data fit levels were examined by χ^2 / *df*, RMSEA, CFI and TLI indices.

Before analyzing the data, it is important to consider whether the data set is suitable for the analysis and whether missing data and multicollinearity are affecting the data set. The analysis of CFA and EFA were done by using MPLUS packages (WLSMV) which are employed with dichotomous (1-0) data. There were no missing data. For the multicollinearity assumption, tolerance values and variance inflation factor values (VIF) were examined, separately for each factor. These values are given in Table 3. When Table 3 was examined, it is seen that all tolerance values are greater than 0.01, and variance inflation factor values are less than 10, which shows that there was no multicollinearity.

After checking assumptions, EFA was employed with 11 items of the Xandar subtest of collaborative problem solving skills. The distribution of the items to the factors and the

corresponding collaborative problem solving competencies indicated in the PISA Final Report were examined with EFA analysis.

Table 3. *Tolerance and Variance Inflation Values.*

Factor	Item	VIF	Tolerance
f1	m2	1.193	0.838
	m3	1.122	0.891
	m4	1.119	0.894
	m6	1.103	0.907
f2	m5	1.072	0.933
	m7	1.087	0.920
	m8	1.042	0.960

KMO and Barlett Sphericity Tests were used to determine the suitability of the data set for the EFA. The KMO value indicates whether the data matrix is suitable for factor analysis and is expected to be greater than 0.60. The Barlett sphericity test examines whether there is a relationship between variables based on partial correlations and the chi-square value calculated here is expected to be significant (Çokluk, Şekercioğlu, & Büyüköztürk, 2015). KMO and Barlett's values indicate that the data set is appropriate for EFA. EFA is an analysis based on correlation or covariance matrix. For this reason, when the EFA with 1-0 data patterns is desired, the correlation matrix should be tetrachoric. Since the data characteristics in this study were of 1-0 structure, an analysis was performed by the tetrachoric correlation matrix. EFA analysis started with 11 items, but the items (m1, m9, m10 and m11) with low factor loadings (<0.3), were excluded from the analysis. The analysis was continued with the remaining seven items. The analysis results in Table 4 show that seven items were collected in two factors. The items in the first factor (f1) are m2, m3, m4, and m6. The items in the second factor are m5, m7, and m8. The item distributions obtained in the factors also align with the competencies in the PISA final report. The PISA report is also used for naming the factors. Accordingly, f1 is called as "Common Understanding", and f2 is "Team Organization. Factor loadings of the items collected under the Common Understanding factor and the Team Organization factor are presented in Table 4.

Table 4. *Item Factor Loadings.*

	Factor Loadings	
Item	f1	f2
m2	**0.703**	0.285
m3	**0.502**	0.277
m4	**0.537**	0.168
m6	**0.512**	0.153
m5	0.198	**0.503**
m7	0.269	**0.655**
m8	0.129	**0.424**

Collaborative problem solving model, which was put forward by EFA, was confirmed by CFA. The obtained model is shown in Figure 1.

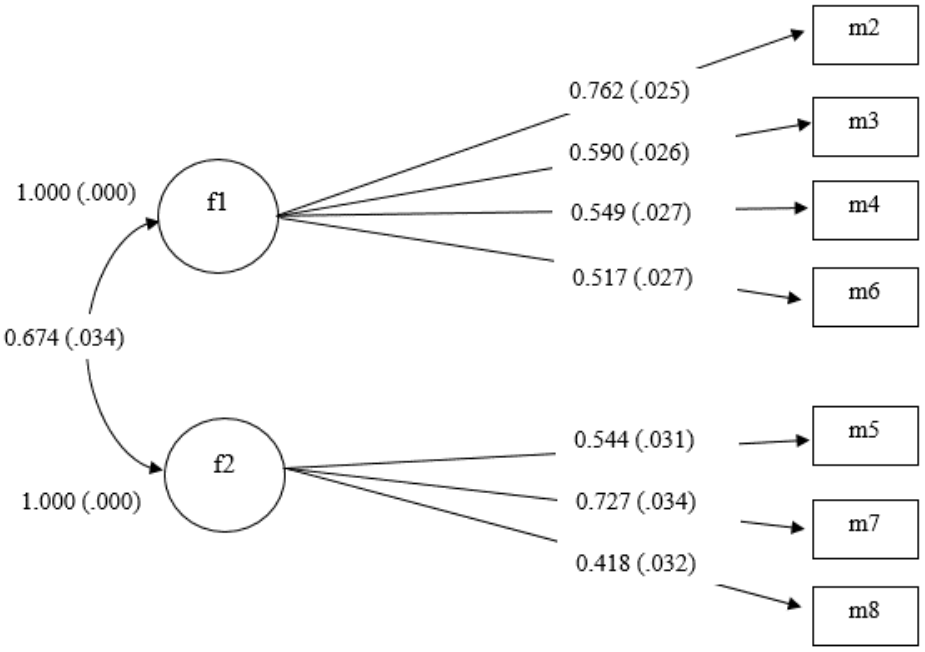

Figure 1. *Collaborative Problem Solving Model.*

The CFA was performed with Mplus 7 program and the model data fit was examined by referring to the indices indicated in Table 5.

Table 5. *Acceptable Levels of Goodness of Fit Indices.*

Fit Indices	Acceptable fit	Good fit
χ^2	$2df < \chi^2 \leq 3df$	$0 \leq \chi^2 \leq 2df$
χ^2/df	$2 < \chi^2/df \leq 8df$	$0 \leq \chi^2/df \leq 2$
RMSEA	0.05<RMSEA≤0.08	0≤RMSEA≤0.05
TLI	0.95≤NNFI<0.97	0.97≤NNFI≤1.00
CFI	0.95≤CFI<0.97	0.97≤CFI≤1.00

(Schermelleh and Moosbrugger, 2003; Tabachnick and Fidell, 2007)

Collaborative problem solving skills model and tested CFA model results for Singapore, Norway and Turkey subgroups are listed in Table 6.

Table 6. *Collaborative Problem Solving Model and Model Fit Indices of Subgroups*.

Models	χ^2	χ^2 (p)	χ^2/df	RMSEA	CFI	TLI
Collaborative Problem Solving Model	40.657	0.000	3.127	0.027	0.987	0.979
Singapore	20.509	0.083	1.577	0.024	0.987	0.979
Norway	18.743	0.131	1.441	0.022	0.990	0.984
Turkey	22.363	0.050	1.720	0.026	0.961	0.936

When Table 6, which includes model fit indices for collaborative problem solving model, is examined, it can be said that model data fit level shows a good fit for $p = 0.05$ significance level. When subgroups elaborated separately, chi-square value $p = 0.05$ level of significance for Singapore is 0.083> 0.05 for Norway is 0.131> 0.05 not meaningful, but Turkey = 0.05 = 0.05 is significant. In addition, to control the effect of sample size χ^2/df, and goodness of fit indices RMSEA, CFI and TLI were also examined. While each of the discussed indices showed

a good fit for Singapore and Norway, for Turkey χ^2/df and RMSEA showed a good fit, CFI and TLI values indicated acceptable fit.

In the next step of the study, MGCFA was used to reveal the effects of unobservable structures on observable variables. One of the groups was considered as a reference at the MGCFA, and the values of this group were fixed at each stage and the level of fit of the values of the other groups was examined accordingly. The group whose values were constant is called the reference group, and in each analysis, one of the countries was selected as the reference group for the binary groups. In addition to examining whether the fit indices were within the acceptable range, the differences between CFI and TLI values were examined according to the less restricted model at the invariance stages. If this difference is between -0.01 and 0.01, it is considered to be an acceptable level for the transition to the next stage (Cheung and Resvold, 2002).

3. RESULT / FINDINGS

In this section, the findings are presented in the order of research problems. Findings related to the configural, metric, scalar, and strict invariance of PISA 2015 collaborative problem solving model are presented respectively. Firstly, the findings for the related countries regarding configural invariance are shown in Table 7.

Table 7. *Configural Invariance Findings.*

Configural invariance	χ^2	*df*	RMSEA	CFI	TLI
Singapore-Norway	1224.328	42	0.025	0.986	0.979
Norway-Turkey	882.754	42	0.026	0.978	0.967
Singapore-Turkey	879.747	42	0.025	0.978	0.967

The configural invariance of the collaborative problem solving measurement model was tested at this stage. When Table 7 is examined, it is seen that for Singapore and Norway RMSEA = 0.025 <0.05. 0.97 <CFI = 986 <1 and 0.97 <TLI = 0.979 <1 and these values show good fit levels. For Norway and Turkey while RMSEA = 0.026 <0.05. 0.97 <CFI = 0.978 <1 and 0.95 <TL = 0.967 <0.97 values show a good fit for RMSEA and CFI, for TLI index, the fit is considered acceptable. Lastly for Singapore and Turkey RMSEA = 0.025 <0.05. 0.97 <CFI = 0.978 <1 and 0.95 <TLI = 0.967 <0.97, RMSEA, and CFI show a good fit. However, for TLI index, it is only at the acceptable level. These findings for Singapore-Norway, Norway-Turkey and Singapore-Turkey groups demonstrate that the model met the configural invariance. Since the configural invariance is a prerequisite for metric invariance, then the next stage for metric invariance has been tested for all three groups. The fit indices are presented in Table 8 for this purpose.

In order to obtain evidence of metric invariance, item factor loadings were examined in addition to item factor structures. Singapore-Norway, Norway-Turkey, and Singapore-Turkey group analyses results were presented separately. As indicated in Table 8, for Singapore and Norway, RMSEA = 0.031 <0.05. 0.97 <CFI = 0.974 <1, 0.95 <TLI = 0.967 <0.97, RMSEA and CFI indices show a good fit and an acceptable fir for TLI. For Norway and Turkey RMSEA = 0.044 <0.05. CFI = 0.924 <0.95. TLI = 0.903 <0.95 while RMSEA show a good fit. CFI and TLI indices are only at acceptable level. For Singapore and Turkey, as indicated in Table 8 RMSEA = 0.028 <0.05, 0.95 <CFI = 0.967 <0.97, 0.95 <TL = 0.958 <0.97 as in the previous comparison RMSEA showed a good fit but CFI and TLI indices were only at acceptable level.

Table 8. *Metric Invariance Findings.*

Scalar invariance	χ^2	*df*	RMSEA	CFI	TLI	χ^2 diff. test	Δ*df*	ΔCFI	ΔTLI
Singapore-Norway	1224.328	42	0.031	0.974	0.967	15.691 (*p*=0.0078)	0	-0.012	-0.012
Norway-Turkey	882.754	42	0.044	0.924	0.903	38.078 (*p*=0.000)	0	-0.054	-0.064
Singapore- Turkey	879.747	42	0.028	0.967	0.958	13.504 (*p*=0.0191)	0	-0.011	-0.009

For Singapore-Norway, although the fit indices were found to be a good fit for RMSEA and CFI, and acceptable for TLI index, chi-square ($\Delta\chi^2$) difference test results between the two models were found to be $p = 0.0078 < 0.05$. In other words, models for Singapore and Norway groups differ significantly from each other. In addition, when ΔCFI and ΔTLI values are examined, it is observed that they are not in the range of -0.01 to 0.01, which is accepted for the transition to the next stage (scalar invariance). The obtained ΔCFI and ΔTLI values were found to be the same and -0.012. For Norway-Turkey RMSEA, although they present a good level of fit CFI and TLI has presented index values outside the acceptable range. Additionally, chi-square ($\Delta\chi^2$) $p = 0.000 < 0.05$ of the difference test is significant thus the models for Norway and Turkey have been found to significantly differ from each other. When ΔCFI and ΔTLI values were examined, it was observed that they were not in the range of -0.01 to 0.01. The obtained ΔCFI and ΔTLI values are -0.054 and -0.064 respectively. For Singapore and Turkey RMSEA showed a good level of fit, but CFI and TLI indices are only at the acceptable level. However, as is clear from Table 8, chi-square ($\Delta\chi^2$) $p = 0.0191 < 0.05$ of the difference test result is significant therefore; the model for Singapore and Turkey group has been found to significantly differ from each other. When ΔCFI and ΔTLI values were examined, it was found that these values are not in the specified range of -0.01 and 0.01. The obtained ΔCFI and ΔTLI values are -0.011 and -0.009 respectively.

The chi-square difference test results presented in the findings were obtained by a two-step approach using the DIFFTEST option in the Mplus analysis program (Wang and Wang. 2012). The findings show that models for Singapore-Norway. Turkey-Norway Singapore-Turkey groups did not show the metric invariance step. This reveals that the PISA 2015 collaborative problem solving test might have been affected by the other variables for these countries.

Table 9. *Item Factor Loadings and Thresholds for Singapore and Norway (Configural).*

Factor Loadings			Thresholds		
Item	Singapore	Norway	Item	Singapore	Norway
M2	0.703	0.707	M2$1	-0.683	-0.160
M3	0.423	0.625	M3$1	-0.755	-0.254
M4	0.496	0.585	M4$1	-0.316	0.151
M5	0.618	0.707	M5$1	-0.423	-0.736
M6	0.498	0.571	M6$1	-0.339	0.384
M7	0.700	0.527	M7$1	-0.928	-1.100
M8	0.368	0.253	M8$1	-0.788	-0.957

Considering that metric invariance is a prerequisite for scalar invariance and the findings are significant at 0.05 level and metric invariance does not hold, the analysis did not proceed to the next stage of invariance. In the second stage of the study, to investigate which items differ from

each other, the factor loadings of the items and the threshold values for country groups were examined and the findings are presented in Tables 9, 10, and 11.

When the factor loadings were examined for Norway and Singapore, it was observed that the differences were large for items m3, m4, m5, m6, m7, and m8. When we consider the content of these items, the participants of these two countries; "discuss the meaning of the problem on a common basis for the solution of an existing problem" (m3 and m4); "establish team organization and team rules" (m5 and m7); "explore different team members' perspectives and abilities" (m6) and "ask other team members to perform their duties" (m8). On the other hand, the item thresholds in Table 9 also contain interesting findings. The magnitude of the negative item thresholds shows that item's easiness and for positives vice versa. In this respect, when the threshold values in the table above are examined, it is observed that the items m2, m3, m4 in the instrument were easier for Singapore and the other items were easier for Norway.

Table 10. *Item Factor Loadings and Thresholds for Norway and Turkey (Configural).*

	Factor Loadings			Thresholds	
Item	Norway	Turkey	Item	Norway	Turkey
M2	0.707	0.683	M2$1	-0.160	0.156
M3	0.625	0.337	M3$1	-0.254	0.367
M4	0.585	0.471	M4$1	0.151	0.316
M5	0.707	0.336	M5$1	-0.736	-0.075
M6	0.571	0.402	M6$1	0.384	0.301
M7	0.528	0.869	M7$1	-1.100	-0.200
M8	0.253	0.252	M8$1	-0.957	-0.160

When Table 10 for Norway and Turkey is examined, it is observed that factor loadings for items m3, m4, m5, m6, and m7 differ from each other in a relatively big magnitude. In terms of content of the items, it was noted that the participants of the two countries differed in their interpretations regarding "discussing the meaning of the problem on a common basis" (m3 and m4); "establishing team organization and team rules" (m5 and m7); and "exploring the perspectives and abilities of different team members for the solution of an existing problem" (m6). At the same time, when the item thresholds are examined items m2, m3, m4, m5, m7, and m8 are is quite easy for the Norwegian participants than participants in the Turkey sample. The only item, which is easy for Turkey sample participants was item m6.

Table 11. *Item Factor Loadings and Thresholds for Singapore and Turkey (Configural).*

	Factor Loadings			Thresholds	
Item	Singapore	Turkey	Item	Singapore	Turkey
M2	0.707	0.683	M2$1	-0.683	0.156
M3	0.422	0.337	M3$1	-0.755	0.368
M4	0.497	0.471	M4$1	-0.316	0.316
M5	0.707	0.336	M5$1	-0.423	-0.075
M6	0.497	0.402	M6$1	-0.339	0.301
M7	0.692	0.870	M7$1	-0.928	-0.201
M8	0.367	0.252	M8$1	-0.788	-0.160

When we compare factor loadings for Singapore and Turkey, significant differences are observed at items m3, m5, m6, m7, and m8. This finding is similar to the findings of the comparisons of two groups (Singapore- Norway, and Norway-Turkey). In addition, in terms of

the item thresholds, all items were easier for the participants of the Singapore than that of Turkey.

When a general evaluation was made on the differences of the items, it was observed that items m3, m5, m6, and m7 differed in all three comparisons. In other words, it can be said that there are differences in terms of discussing the meaning of the problem on common ground, establishing roles and team organization, exploring team members' perspectives and following the rules of the agreement.

4. DISCUSSION and CONCLUSION

According to the findings, while the configural invariance was achieved in all three groups, the metric invariance could not be achieved. Since the metric invariance stage was not achieved, scalar and strict invariance stages were not tested. Therefore, it was concluded that factor structures were the same in all three groups but factor loadings, variances, error variances, and covariances differed. This result shows that the participants of the countries (Singapore. Norway, and Turkey) interpreted the Xandar subtest of the collaborative problem solving skills test differently.

To be able to compare country scores, the established model must hold measurement invariance. However, the findings show that measurement invariance does not hold for the data in this study. The findings show that in country comparisons, factor loadings of m3, m5, m6, and m7 differed from each other. It can be said that these differences can be one of the reasons for not completing all the stages of measurement invariance. These differences in factor loadings may mean that there is a difference in participants' interpretations of these items. The competencies measured in these items are: understanding and discussing the meaning of an existing problem, establishing team rules, and exploring the perspectives and abilities of team members. The differences in the results obtained from the measurement tool show that the participants of this country interpret the items related to these competencies differently.

Considering that, information is globalized and individuals with critical skills are sought after, countries need to become equivalent in this field with other countries. However, the PISA 2015 results show that the scores among the top, middle and low group countries differ significantly from each other in terms of collaborative problem solving skills (OECD. 2017). The fact that the invariance stages cannot be fully achieved is another indication of this. There is also variability between these countries due to unobservable variable(s). This situation leads to the differentiation of the countries in this field due to different reasons and the result that some countries raise competent individuals in terms of collaborative problem solving skills while others are left behind in terms of these skills.

An important contribution of this study to the literature is that its contribution to the collaborative problem solving on the literature. Therefore, there is no measurement invariance on collaborative problem solving research that can be compare to our results with the literature. For the first time in 2015, the OECD conducted a collaborative problem-solving study. Therefore, the results obtained by comparisons of different countries that are made within the scope of this research are of particular importance. On the other hand, although this is the first study in the field, studies are documenting that measurement invariance is not achieved in large scale studies such as PISA and TIMSS. For instance, Kıbrıslıoğlu (2015) found that only configural invariance stage was achieved for mathematics literacy in PISA 2012 Turkey, China-Shanghai, and Indonesia data. Similarly, Karakoç Alatlı (2016) for PISA 2012 mathematical literacy and scientific literacy data of Australia, France, China-Shanghai, and Turkey sample only met the configural stage. As a final example, Wu, Liu, and Zumbo (2007) conducted a study using TIMSS data from the USA, Canada, Australia, New Zealand, Taiwan, Korea, and

Japan and their results showed that only structural and metric invariance stages hold for the data.

Especially in the studies carried out with many countries, the invariance stages must be fully hold for comparisons to be meaningful. For this reason, researchers should examine not only descriptive statistics but also invariance. This study was conducted with the countries in the upper, middle, and lower groups. In addition, invariance studies should be conducted for countries whose scores are not very different from each other. On the other hand, when the literature is examined, the financial literacy test administered in PISA 2012 application is as important as collaborative problem solving skills. In this sense, it is important for the researchers to examine the state of invariance related to financial literacy test on a country-by-country basis and to conduct cross-cultural invariance studies.

Within the scope of this study, only Xandar subtests were examined from six different subtests for collaborative problem solving skills. For this reason, researchers can conduct invariance studies of the other five subtests on different subgroups belonging to different countries and within the same country will contribute to both measurement invariance and collaborative problem solving literature. Another important point is that item bias should be examined in addition to invariance studies. Identifying the factors that cause bias will allow for the purely measurement applications of these factors and to give reliable results.

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).

Authorship contribution statement

Authors are expected to present author contributions statement to their manuscript such as;
Yusuf Taner Tekin: Investigation, Data Analysis, Methodology, Resources, Writing. **Derya Çobanoğlu-Aktan**: Investigation, Data Analysis, Methodology, Resources, Writing, Supervision and Validation.

ORCID

Yusuf Taner TEKİN https://orcid.org/0000-0001-9068-7894
Derya ÇOBANOĞLU AKTAN https://orcid.org/0000-0002-8292-3815

5. REFERENCES

Bahadır, E. (2012). *According programme for ınternational student assessment (PISA 2009), investigation of variables that affect Turkish students' reading skills by regions* [Master Thesis]. Hacettepe University, Ankara.

Başusta, N.B., & Gelbal, S. (2015). Examination of Measurement Invariance at Groups' Comparisons: A Study on PISA Student Questionnaire. *Hacettepe University Journal of Education, 30* (4), 80-90

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley- Interscience Publication.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural equation modeling, 11*(2), 272-300.

Bryne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal Of Cross-Cultural Psychology, 34*, (2), 155-175.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. (Second Edition). The Guilford Press.

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

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2018). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Pegem.

Demir, Ö., & Seferoğlu, S.S. (2017). İşbirlikli Problem Çözmenin Kodlama Eğitimine Yansıması Olarak Eşli Kodlamanın İncelenmesi. *5th International Instructional Technologies& Teacher Education Symposium*. 11-13 October 2017, İzmir.

Erkoç, M.F. (2018). *The effect of collaborative game design on critical thinking, problem solving and algorithm development skills* [Doctoral Dissertation]. İstanbul University.

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts - something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364-379.

Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assesment. *Journal of Educational Measurement*, Spring, *54*(1), 12-35.

Jöreskog, K. G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the simplis command language.* Lincolnwood: Scientific Software International, Inc.

Karakoç Alatlı, B. (2016). *Investigation of measurement invariance of literacy tests in the Programme for International Student Assessment (PISA - 2012)* [Published Doctoral Dissertation]. Ankara University.

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

Lance, C. E., & Vandenberg, R. J., (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods, 3*(1), 4-70.

Mark, B. A., & Wan, T.T.H (2005). Testing measurement equivalence in a patient satisfaction instrument. *Western Journal of Nursing Research, 27*(6), 772-787.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Pyschometrika, 58*, 525-543.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological methods, 9*(1), 93.

Nelson, L.M. (2009). *Instructional-design theories and models: A new paradigm of instructional theory, Volume II*. Routledge Publishing, (ISBN 978-0-8058-2859-7)

OECD, 2017. *PISA 2015 results (Volume 5): Collaborative problem solving*. OECD Publishing.

O'Neil, H. F., Chuang, S. H., & Chung, G. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education: Principles, Policy & Practice*, *10*(3), 361-373.

Oliden, P., E., & Lizaso, J, M. (2013) Invariance levels across language versions of the PISA 2009 reading comprehension tests in Spain, *Psicothema, 25(*3), 390-395.

Özdemir, S. (2005). *The effects of individual and collaborative problem-based learning using an online asynchronized learning tool on critical thinking abilities, academic achievements, and attitudes toward internet use* [Doctoral Dissertation]. Gazi University.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.

Uyar, Ş. (2011). *An investigation of measurement invariance of learning strategies model across different groups in PISA Turkey sample* [Master Thesis] Hacettepe University.

Uzunosmanoğlu, S.D. (2013). *Examining computer supported collaborative problem solving processes using the dual- eye tracking paradigm* [Master's Thesis]. Orta Doğu Teknik Üniversitesi.

Wang, J., & Wang, X. (2012). *Structural equation modelling: Applications using mplus. (First edition).* UK: Wiley Publications.

Wu, D. A., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial ınvariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assesment, Research and Evaluation*, *12*(3), 1-26.

Yandı, A., Köse, İ. A., & Uysal, Ö. (2017). Farklı Yöntemlerle Ölçme Değişmezliğinin İncelenmesi: Pisa 2012 Örneği. *Mersin University Journal of the Faculty of Education, 13*(1), 243-253.