



## International Journal of Contemporary Educational Research (IJCER)

[www.ijcer.net](http://www.ijcer.net)

### Examining the Invariance of a Measurement Model by Using the Covariance Structure Approach

Şeref Tan<sup>1</sup>, Sami Pektaş<sup>2</sup>

<sup>1</sup>Gazi University

<sup>2</sup>Niğde Ömer Halisdemir University

#### To cite this article:

Tan, Ş. & Pektaş, S. (2020). Examining the invariance of a measurement model by using the covariance structure approach. *International Journal of Contemporary Educational Research*, 7(2), 27-39. DOI: <https://doi.org/10.33200/ijcer.756865>

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

## Examining the Invariance of a Measurement Model by Using the Covariance Structure Approach

Şeref Tan<sup>1\*</sup>, Sami Pektaş<sup>2</sup>

<sup>1</sup>Gazi University

<sup>2</sup>Niğde Ömer Halisdemir University

### Abstract

The primary aim of the present study is to examine the measurement invariance of the structural equating model constructed on the numerical and verbal abilities test for sixth grade students across gender, amount of weekly pocket money and students' perceptions of the sufficiency of their pocket money. The secondary aim is to illustrate the use of the IBM AMOS-24 software package step by step with examples to address invariance using the covariance structural analysis approach. The research data were collected from 2304 sixth grade students enrolled in public schools within the Keçiören and Pursaklar suburbs in Ankara. The covariance structures analysis approach was employed during the examination of the measurement model invariance. The study revealed that invariance was achieved in terms of configural, measurement (in both measurement weights and measurement residuals) and structural invariance with respect to all subgroups.

**Key words:** Configural invariance, Measurement invariance, Structural invariance, Structural equation modeling, Ability.

### Introduction

In studies where measurements are used to compare different groups, it is important to ensure measurement invariance. If there is a situation of obtaining biased measurement results for a subgroup, determining this situation allows the research to be interpreted the findings more accurately. In this study, an example of how the measurement invariance study can be performed by using IBM AMOS-24 for an ability test scores is presented. For this purpose, first theoretical information about ability test and then some theoretical information about measurement invariance analysis are explained in this section.

The concept of ability has been defined in various ways. The concept of ability, which refers to mental computation activities, can be categorized based on the common properties of separate factors essential for individuals to carry out mental operations. The mental power required for each of these identified groups is called ability. It is defined as being inherited, a boundary enveloping learning, and the power to accommodate the effect originating from external factors (Turkish Language Association [TDK], 2016). As ability conveys differences among individuals and reflects the development of a process, the measurement of abilities is considered important. Thus, tests that could create the opportunity to examine individuals' abilities and identify differences among individuals were developed. Ability tests typically measure knowledge and skills acquired over long periods of time whereas so-called achievement tests are often "subject/topic" specific and may require more recent targeted study to perform well (Benson, 2008; Ghanizadeh, 2017; Kaufman, 2018). Another stream of literature shows that noncognitive skills are important determinants of performance in achievement tests (McGrew, 2005; Borghans, Golsteyn, Heckman & Humphries 2016). Researchers consider that the tests, which measure cognitive skills, abilities and intelligence, will show a positive correlation with achievement points. In this case, the scores obtained by achievement tests are used as the points that represent cognitive ability (Berkowitz & Stern, 2018; Kyllonen & Kell, 2018). In addition to ability tests that can measure more than one ability, there are also ability tests that measure specific characteristics (R. Atkinson, Atkinson, & Hilgard, 1995). Ability tests are categorized into two groups based on the distinction between individuals' abilities as general and specific. While ability tests are observed to have a homogeneous structure when intelligence tests are examined, they also appear to display a heterogeneous structure since they also measure such characteristics as language, number and reasoning as well as all the characteristics of intelligence (Özğüven, 2007).

\* Corresponding Author: Şeref Tan, [sereftan4@yahoo.com](mailto:sereftan4@yahoo.com)

Specific ability is defined as the power individuals embody to realize behaviors with specific characteristics. It is individuals' power to benefit from this condition since it gives them the opportunity to achieve certain things in different areas of ability (Yeşilyaprak, 2007). On the other hand, general ability can be defined as the general capacity that includes such abilities as solving problems, solving arithmetic computations, thinking abstractly, reasoning, establishing association between words, finding synonyms, and can more or less affect all the behaviours of an individual. This ability is equivalent to the general intelligence, also known as g factor (g), proposed by Charles Spearman and, hence, it can also be defined as general mental power (Sak, 2014, p. 42). Cattell-Horn-Carroll theory classifies cognitive abilities into three levels, which are narrow abilities, broad abilities and general ability. Cognitive ability refers to the mechanisms of mental capacity such as remembering, problem solving than actual knowledge (McGrew, 2005; Floyd, Keith, Taub & McGrew, 2007)

Measurement of the same characteristic across different groups during the construction stage of the tests developed is examined under the heading of measurement invariance. As Millsap & Tein (2004) pointed out that, the extension of the analysis to the multiple-population case is less well known especially for ordered-categorical data in the literature on factor analysis. When we want to compare different groups, it should be proved that the scores obtained from the scale are not biased. As Camilli (2006) pointed out that measurement invariance contributes to validity evidence in that scores from a tool are subject to issues of bias and lack of fairness if invariance does not hold. Because of this reason, as Chung and others, (2016) mentioned, further investigation is needed to answer the question of whether the scale items perform similarly across subgroups and one way to examine this question is through assessing the measurement invariance of a scale. There are studies conducted on measurement invariance to decide whether or not items on a test express the same meaning across different groups (Arana, Rice & Ashby, 2018; Başusta & Gelbal, 2015; Camerota, Willoughby, Kuhn & Blair, 2018; Chavez, ShROUT, Garcia, Forno & Celedón, 2018; Gaddy, Casner & Rosinski, 2018).

Wicherts (2016) states that measurement invariance is a fundamental problem in identifying whether or not population norms are valid for sub-groups as well. In more than half of the studies he examined, he concluded that in intelligence tests, there was no measurement invariance across the groups by ethnic origin, gender, educational condition, and age. He underlined the fact that measurement invariance is very important for the validity of neurocognitive tests in clinical, educational and professional practices. Wicherts and Dolan (2010) stated that the fair use of measurement invariance across groups was very important in intelligence tests and other psychological tests. They mentioned that there is common belief that the factor loadings in confirmatory factor analysis are sufficient to ensure measurement invariance. At this point, they indicate that in constructing measurement invariance by means of confirmatory factor analysis, there is a need for a statistical test for the equation of the groups measured. Blankson and McArdle (2013) used cognitive performance tests measuring episodic memory and mental status and tested these by using multilevel modeling for two-factor structural invariance, confirmatory factor analyses and longitudinal data. In a study they conducted with 244 undergraduate students, Bailey, Neigel, Dhanani and Sims (2018) applied two spatial ability (spatial visualization and projection) measurements on a computer- and a paper-based format. By ensuring the measurement invariance in both paper- and computer-based tests to measure spatial ability, they aimed to ensure reliability. It was found that based on the type of test implemented, measurement invariance could not be ensured and that the way the test was implemented had an impact on different types of errors. Furthermore, when compared to the computer-based tests, the paper-based tests were found to be more reliable. Since the existing proof cannot reach the same structure in a reliable way, in such tests as ability tests, they mentioned the necessity of conducting measurement invariance.

During the process of testing structures that want to be measured, the condition where individuals from different subgroups have equal chance of achieving a certain score is referred to as measurement invariance (Watson, Thompson & Adam, 2007). For equal measurements, the connection between the observed and latent variable should be the same (Drasgow & Kanfer, 1985). For a measurement model to have the same structure across different groups, the factor loadings of the items in a scale, and the correlations and variances among the identified factors, should be the same. During the examination of between-groups measurement invariance of the measurement model constructed, the model constructed in each phase is built on the model constructed in the previous phase. Accordingly, the measurement invariance examined at a certain phase is examined based on the model in the previous phase, in with fewer restrictions are placed, by using the research data to test whether a significantly lower level of model fit is displayed. If it displays a good level of model fit with the data – as good as that of the previous model in which more restrictions were placed – then, it is believed that the more complex model can explain the data. The examination conducted shows that the measurement invariance in that phase was realized (Cheung & Rensvold, 2002). Measurement invariance proposed by Milfont and Fischer (2010) is addressed under seven titles, namely configural, metric, scalar, error variances, factor variances, factor

covariances and invariance of factor means. Vandenberg and Lance (2000) also address measurement invariance under seven terms: configural, metric, scalar, uniqueness, factor invariance, factor covariance and invariance of factor means. On the other hand, in some research, measurement invariance is addressed under three titles: configural, metric and scalar invariance (Campbell, Berry, Joe & Finney, 2008; Xu & Tracey, 2017). In some other studies, generally measurement invariance is addressed under four headings, namely configural, metric, scalar and strict invariance (Meredith, 1993; Wu, Li and Zumbo, 2007).

In this regard, within the scope of research in the literature defined as configural, metric, scalar and strict invariance, measurement invariance is addressed under four headings; namely unconstrained model (configural invariance), measurement weighted model (metric invariance), structural covariances model (scalar invariance) and measurement errors model (strict invariance).

As stated by Byrne (2016: 227-228), *“In seeking evidence of multigroup equivalence, researchers are typically interested in finding the answer to one of five questions. First, do the items comprising a particular measuring instrument operate equivalently across different populations? In other words, is the measurement model group-invariant? Second, is the factorial structure of a single instrument or of a theoretical construct equivalent across populations? Third, are certain paths in a specified causal structure equivalent across populations? Fourth, are the latent means of particular constructs in a model different across populations? Finally, does the factorial structure of a measuring instrument replicate across independent samples drawn from the same population? This latter question addresses the issue of cross-validation.”*

Configural invariance refers to whether or not the constructed model is the same across all groups. As Chung and others (2016) pointed out, configural invariance is the fact that factor structures between groups are equivalent. In other words, configural invariance tests that the same pattern of item-factor loadings exists across the groups being compared, which requires that the same items have nonzero loadings on the same factors (Chung and others, 2016). The model consisting of constant and free parameters is equal across the groups in the model at the step. Since it is the most fundamental structure, it is also referred to as unconstrained model; it is also known as the initial model in measurement invariance analyses. To observe whether the other steps of invariance are ensured, comparisons are made based on the configural invariance values (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000).

Metric invariance refers to the equalivance among regression coefficients, that is, factor loadings. As Chung and others (2016) mentioned, metric invariance, additionally requires that unstandardized factor loadings are the same across groups. Metric invariance identifies the invariance of the factor loadings across the groups; that is, it determines whether or not the responses given to the latent variables are equivalent. In addition to the factor loadings, it also refers to the equivalence across the factor loadings (Byrne, 2016; Meredith, 1993).

Scalar invariance is based on the equivalence of factor covariances across groups. It is the model where all factor loadings, factor variances and factor covariances are constrained. It is a kind of invariance where factor covariances are equalized across the groups after configural and metric invariances are ensured (Cheung & Rensvold, 2002; Meredith, 1993).

Strict invariance is also referred to as invariant uniqueness. It is based on the principle that the error terms are equivalent across the comparative groups. It is based on the error variance equivalence after configural, metric, and scalar invariances are ensured. It is a type of invariance where all factor loadings, factor variances, factor covariances and error variances are constrained (Cheung & Rensvold, 2002; Meredith, 1993; Vandenberg & Lance, 2000; Wu, Li & Zumbo, 2007).

As we mentioned before, if measurement invariance is not provided, the results would be biased. For this reason, measurement invariance is an important property that should be examined when developing or using a measurement tool. Measurement invariance can be used with tests and scales based on two fundamental theories such as the measurement invariance structural equating model and item response theory. In the present article, the stage of identifying measurement invariance was realized by utilizing the covariance structural analysis (COVS) approach. Only the variances in observed variables and the covariances among observed variables are made use of in covariance structural analysis.

## Aim of the Study

The present study aimed to examine the measurement invariance of the structural equating model constructed on the numerical and verbal abilities tests for sixth grade students across gender, amount of weekly pocket money, and students' perceptions of the sufficiency of their pocket money by using the covariance structural analysis (COVS). This study also aims to illustrate the use of the IBM AMOS-24 software package step by step with examples to address invariance using the covariance structural analysis approach.

## Method

### The Research Model

The aim was to construct a structural model of the students' numerical and verbal ability tests and the sub-factors comprising these tests. Thus, in this respect, the research is based on a correlational study. It is descriptive in nature as it examines the measurement invariance aspect of the structural equating model based on the groups formed based on student variables.

### The Study Group

The study group was comprised of 2304 sixth grade students enrolled in public schools within the Keçiören and Pursaklar suburbs in 2016-2017 academic year, Ankara. Of these students, 1146 (49.7%) were female students, while 1158 (50.3%) were male students. With respect to amount of pocket money, 608 (26.4%) students had a low (0-5 TL), 1118 (48.5%) students had a moderate (6-20 TL), and 578 (25.1%) students had a high amount of pocket money (21 TL and above). In terms of students' perceptions regarding the sufficiency of their pocket money, 1914 (83.1%) students stated that it was sufficient, whereas 390 (16.9%) students claimed that it was not sufficient.

### Data Collection Instruments

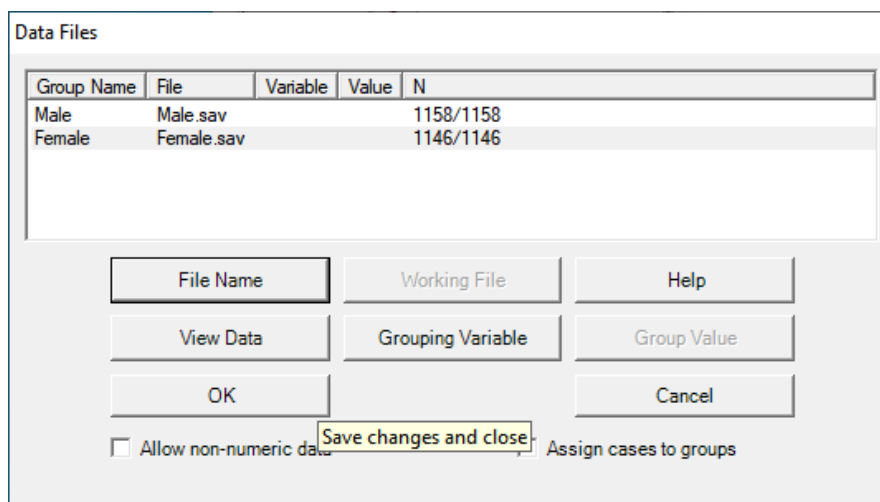
To measure students' numerical and verbal abilities within the scope of the present study, the data were collected by means of the "Numerical Ability Test" and "Verbal Ability Test", developed by Pektaş (2018) within the scope of his Ph.D. dissertation study. This test is used to show how to perform measurement invariance analysis. In this test, the fact that it contains the actual data set is important for more realistic analysis compared to the simulation data. The numerical test consisted of 45 multiple choice test items and four sub-factors: patterns (*pattern*), finding the four operations based on the symbols (*operation's symbol*), finding the symbols used in the four operations (what is the *symbol*) and problem solving (*problem*). The verbal ability test consisted of two sub-factors, vocabulary and inter-word relationship, and of 45 multiple-choice items. When the test statistics of the pilot study for the numerical ability test were examined, it was observed that the difficulty level of the test was at moderate level with a difficulty index of 0.51, and that it was found to sufficiently discriminate between students with high and low levels of numerical ability with an average upper-lower group item discrimination index value of 0.43. The KR-20 reliability coefficient of the numerical ability test was 0.90, which shows that the measurements had a high degree of reliability. When the test statistics of the pilot study for the verbal ability test was examined, it was observed that the difficulty level of the test was again at a moderate level with a difficulty index of 0.58, and that it was found to sufficiently discriminate between students with high and low levels of verbal ability with an average upper-lower group item discrimination index value of 0.40. The KR-20 reliability coefficient of the verbal ability test was 0.85, which shows that the measurements had a high degree of reliability. During the confirmatory analysis phase of the four-observed variables of numerical ability test, the fit indices of the first level with these 4-observed variable, the fit indexes of this CFA model were found to be as follows:  $\chi^2_{937} = 3018.356$ ;  $\chi^2/df = 3.22$ ; CFI=.971, GFI=.932, AGFI=.924 and RMSEA=.033. In the confirmatory analysis phase of the two-observed variables of the verbal ability test, the fit indices of this CFA model were observed as the following:  $\chi^2_{940} = 2838.217$ ;  $\chi^2/df = 3.02$ ; CFI=.976, GFI=.939, AGFI=.933 and RMSEA=.030. When these are compared to the criteria defined by Byrne (2013) and Schermelleh-Engel, Moosbrugger & Müller (2003), they are considered acceptable and they display a perfect fit.



## Data Analysis

The data collected from the 6<sup>th</sup> grade students were entered into the IBM SPSS-25 package program. In the examination of the measurement model invariance, the covariance structural analysis approach was utilized. The data were examined for missing data and outliers, and the mahalanobis distance values were examined. Tests for normality and multicollinearity were also conducted. The covariance matrix, the asymptotic covariance matrix and average vectors were computed for each subgroup. In the study, maximum likelihood approach was used as the parameter estimation method. During the construction of the measurement model and the invariance test phase, the IBM AMOS-24 package program was utilized. For the comparison of the measurement invariance in the measurement model, the changes in the criterion of CFI ( $\Delta$ CFI) values were taken into consideration. In the first stage of the measurement model, the multivariate normality assumption was tested for each subgroup, and each kurtosis value obtained for each group was observed to be below 1.00. The difference between the CFI value in the configural model and the CFI value of the models tested in the later stages was less than .01. Based on the conditions for ensuring measurement invariance, this has been accepted as proof for the presence of measurement invariance (Cheung & Rensvold, 2002). Also a difference of less than .01 in the  $\Delta$ CFI index supports the less parameterized model (Chung and others, 2016). The measurement invariance approach and the interpretations explained in the present article can be analyzed in models constructed with such measurement tools as ability tests, achievement tests, scales, psychological tests with the aim of comparing different groups. During the analyses, the operations done via the IBM AMOS-24 package program are explained as follows:

*IBM AMOS-24 operations for configural invariance.* In the first step, the groups are defined by selecting the *Manage Groups*. Function from the *Analyze* menu in the *AMOS* program: (*Analyze, Manage Groups*, In Group Name Box **type** the name of the first group as **Male**, click **New** icon and **type** the name of the second group as **Female**, then click **Close**). In the second step, the data files for male and female sub-groups are assigned by using the *Select Data File(s)* icon or by using function from the *File, Data Files* menu.



In third step, the *Emulisrel6* box is clicked by selecting *Estimation* from *Analysis Properties* in the *View* menu. In final step, the analysis is run by selecting *Calculate Estimates* from the *Analyze* menu.

*IBM AMOS-24 operations for configural and structural invariance.* Until the stage of making the predictions, as an addition to the operations mentioned above, the parameters to be predicted in the model are labeled manually or automatically. For automatic labeling, the *Multiple Group Analysis* function is selected from the *Analyze* menu. Then, the parameters to be constrained are selected in the *Multiple-Group Analysis* dialog box.

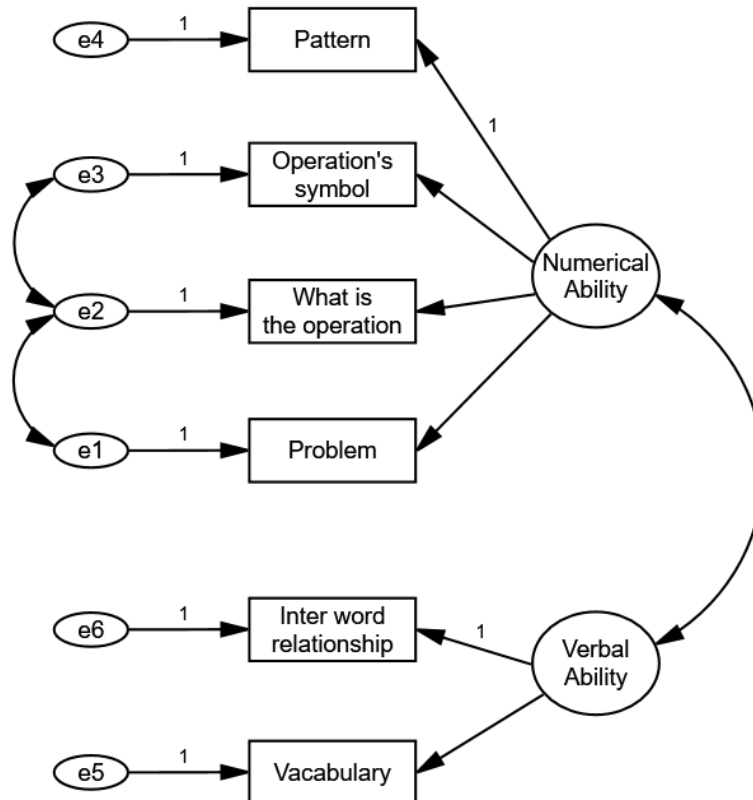
Multiple-Group Analysis		Models							
Parameter Subsets		1	2	3	4	5	6	7	8
Measurement weights	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Measurement intercepts	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural weights	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural intercepts	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural means	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural covariances	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structural residuals	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Measurement residuals	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

In final step, the analysis is run by selecting *Calculate Estimates* from the *Analyze* menu.

## Findings and Interpretations

### Measurement Model

Measurement invariance was tested both for the observed 6 sub-scale scores (For the numerical latent variable: *pattern*, *operation's symbol*, *what is the operation* and *problem* observed variables; for the verbal latent variable: *vocabulary* and *inter word relationship* observed variables) and for the latent structure related to these subscales (the relationship between numerical ability and verbal ability). The baseline model was displayed in Figure 1 below.



**Figure 1.** The baseline model for the multiple-group invariance of the numerical and verbal ability measurement model.

## Measurement Invariance by Gender

### *The Baseline Model for the Measurement Invariance*

The first stage, of the measurement invariance analysis to be conducted in various stages, is the identification of an appropriate baseline measurement model for each group. The baseline model is portrayed in Figure 1. In the baseline model of the measurement model based on 6<sup>th</sup> grade students' gender, the fit indices for female students were found to be as follows:  $\chi^2_6 = 4.351$ ;  $\chi^2/df = .725$ ; GFI=.999; CFI=1.000 and RMSEA=.000. As for male students, they were found to be as follows:  $\chi^2_6 = 15.335$ ;  $\chi^2/df = 2.556$ ; GFI=.996; CFI=.997 and RMSEA=.037. In the baseline model, theoretically reasonable two covariance links (e1 with e2 and e2 with e3) between residuals of the same factor were used in order to have better-fit indexes for all subgroups in the study. In conclusion, the baseline model in Figure 1 displayed a high level of model fit for both female and male students.

### *Configural Invariance of the Measurement Model for Gender*

As stated by Byrne (2016), to ensure configural invariance, factor loading patterns and the number of factors should be similar for each group. The measurement model based on students' gender has ensured configural invariance:  $\chi^2_{12} = 19.686$ ;  $\chi^2/df = 1.641$ ; GFI=.997; CFI=.999 and RMSEA=.017. That is, in this unconstrained measurement model, the factor structure for these two populations based on gender was found to be similar.

The unstandardized estimated parameters (regression weights, covariances, and variances) of Male and Female groups for configural invariance are given in Tables 1a, 1b and 1c, below.

**Table 1a.** Regression weight estimates of gender for configural model

Regression Weights			Male	Female
Problem	<---	Numerical Ability	.981**	1.093**
Pattern	<---	Numerical Ability	1.000	1.000
Vocabulary	<---	Verbal Ability	.651**	.786**
Inter word relationship	<---	Verbal Ability	1.000	1.000
What is the Operation	<---	Numerical Ability	.509**	.574**
Operation Symbol	<---	Numerical Ability	.690**	.698**

\*\* : p < .01

**Table 1b.** Covariance estimates of gender for configural model

Covariance			Male	Female
Numerical Ability	<-->	Verbal Ability	9,308**	7,963**
e1	<-->	e2	1,013**	,720**
e2	<-->	e3	,501**	,445**

\*\* : p < .01

**Table 1c.** Variance estimates of gender for configural model

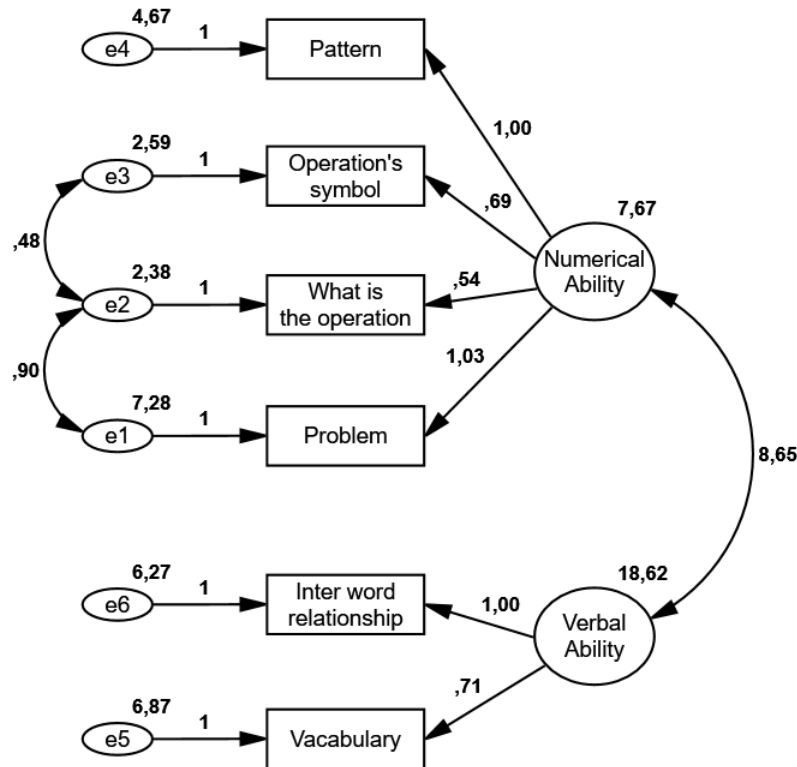
Covariance	Male	Female
Numerical Ability	7,985**	7,367**
Verbal Ability	21,469**	16,005**
e1	7,233**	7,240**
e2	2,385**	2,337**
e3	2,295**	2,885**
e4	4,969**	4,365**
e5	7,890**	5,803**
e6	5,642**	6,645**

\*\* : p < .01

### *Configural, Measurement and Structural Invariance of the Measurement Model for Gender*



As it was given by Byrne (2016), in the measurement and structural invariance test, the focus is more on which parameters in the measurement model and its structural constituents are equivalent in both groups. In this part of the analysis, progressively, first *measurement weights* are constrained, then *structural covariances* are constrained and finally *measurement errors* are constrained. The IBM AMOS-24 output path diagram of the unstandardized estimated parameters when all the model parameters are constrained equal is given in Figure 2. The results obtained for the measurement invariance by gender in terms of factor loadings (measurement weights), structural covariances and measurement errors are presented in Table 2, below.



**Chi square = 63,715, Df= 27, GFI= ,991, CFI= ,994, RMSEA= ,024**

**Figure 2.** Output path diagram for configural, measurement weight, structural covariance, and measurement error invariance of the measurement model for gender.

**Table 2.** Measurement and structural invariance results by gender.

Model	Number of Parameters	$\chi^2$	df	$\chi^2/df$	CFI	$\Delta CFI$	RMSEA
1. Unconstrained (Configural)	30	19.686	12	1.641	.999		.017
2. Measurement Weights	26	33.334	16	2.083	.997	<b>0.002</b>	.022
3. Structural Covariances	23	37.196	19	1.958	.997	<b>0.002</b>	.020
4. Measurement Errors	15	63.715	27	2.36	.994	<b>0.005</b>	.024

Note:

Unconstrained Model: All the parameters are predicted freely.

Measurement Weights Model: All *Factor loadings* are constrained (equated).

Structural Covariances Model: All *Factor loadings + factor variances and covariances* are constrained (equated).

Measurement Errors Model: All *Factor loadings + factor variances + factor covariances + error variances* are constrained (equated).

According to the unconstrained model used for configural invariance, as initially only the four factor loadings (measurement weights) predicted for the measurement model are to be defined as being equivalent for the two groups, the number of parameters estimated by the measurement model in which the factor loadings are equated is reduced by 4 when compared to the configural model and, hence, the predicted number of parameter reduced to 26. In addition, owing to structural variances and covariances, numerical and verbal latent variables are to be defined as being equivalent for the two latent variances and 1 covariance; thus, the number of predicted parameters is reduced by 3, yielding 23 parameters. Finally, since 6 error variances and two error covariances are to be predicted once for each group, the parameters to be predicted are reduced by 8, yielding 15 parameters. As can be seen from Table 2, there is an increase in degrees of freedom as much as a decrease in the number of parameter predicted in the model. As can be observed in Table 2, according to the unconstrained (used in configural invariance) model, the changes in CFI in the models obtained by constraining, in sequence, measurement weights, structural covariances and error variances, were less than .01. Hence, the measurement model has ensured configural, measurement and structural invariance across gender.

### Measurement Invariance by Amount of Weekly Pocket Money

#### *The Baseline Model for Measurement Invariance*

Based on students' amount of weekly pocket money, the constructed baseline model yielded the following fit indices for the group with low amount of pocket money:  $\chi^2_6 = 6.334$ ;  $\chi^2/df = 1.056$ ; CFI=1.000 and RMSEA=.010; for the group with a moderate amount of pocket money:  $\chi^2_6 = 10.924$ ;  $\chi^2/df = 1.821$ ; CFI=.998 and RMSEA=.027; and for the group with a high amount of pocket money:  $\chi^2_6 = 5.616$ ;  $\chi^2/df = 0.939$ ; CFI=1.000 and RMSEA=.000. Hence, the measurement model has ensured a high level of model fit for all three groups- a low, moderate, high amount of pocket money- in the baseline model.

#### *Configural Invariance of the Measurement Model for Amount of Weekly Pocket Money*

The measurement model has ensured configural invariance with respect to students' amount of weekly pocket money:  $\chi^2_{18} = 22.873$ ;  $\chi^2/df = 1.271$ ; CFI=.999 and RMSEA=.011.

#### *Configural, Measurement and Structural Invariance of the Measurement Model for Amount of Weekly Pocket Money*

The results obtained for the measurement invariance by students' amount of weekly pocket money in terms of factor loadings, structural covariances and measurement errors are presented in Table 3 below.

**Table 3.** Measurement and structural invariance results by amount of weekly pocket money.

Model	Number of Parameters	$\chi^2$	df	$\chi^2/df$	CFI	$\Delta$ CFI	RMSEA
1. Unconstrained (Configural)	45	22.873	18	1.271	.999		.011
2. Measurement Weights	37	30.314	26	1.166	.999	<b>0.000</b>	.008
3. Structural Covariances	31	36.276	32	1.134	.999	<b>0.000</b>	.008
4. Measurement Errors	15	57.596	48	1.2	.998	<b>0.001</b>	.009

As can be observed in Table 3, according to the unconstrained model, the changes in CFI in the models obtained by constraining, in sequence, measurement weights, structural covariances and error variances, were less than .01. Hence, the measurement model has ensured configural, measurement and structural invariance based on amount of weekly pocket money.

### Measurement Invariance by Sufficiency of the Amount of Weekly Pocket Money

#### *The Baseline Model for Measurement Invariance*

The fit indices that the baseline model of the measurement model based on whether students' amount of weekly pocket money was sufficient yielded the following fit indices for students who claimed it was insufficient:  $\chi^2_6=3.148$ ;  $\chi^2/df =.525$ ; CFI=1.000 and RMSEA=.000; for students who claimed it was sufficient:  $\chi^2_6=10.560$ ;  $\chi^2/df =1.76$ ; CFI=.999 and RMSEA=.020. Hence, the baseline model in Figure 1 displayed a high level of model fit for both groups who found the amount of their weekly pocket money either sufficient or insufficient.

#### *Configural Invariance of the Measurement Model for Sufficiency of the Amount of Weekly Pocket Money*

The fit indices that the measurement model yielded for configural invariance with respect to the status of whether the amount of students' weekly pocket money was sufficient are as follows:  $\chi^2_{12}=13.708$ ;  $\chi^2/df =1.142$ ; CFI=1.000 and RMSEA=.008 Hence, configural invariance has been ensured.

#### *Configural, Measurement and Structural Invariance of the Measurement Model for Sufficiency of the Amount of Weekly Pocket Money*

The results obtained for the measurement invariance by the sufficiency of students' amount of weekly pocket money in terms of factor loadings, structural covariances and measurement errors are presented in Table 4 below.

**Table 4.** Measurement and structural invariance results by sufficiency of the amount of weekly pocket money.

Model	Number of Parameters	$\chi^2$	df	$\chi^2/df$	CFI	$\Delta$ CFI	RMSEA
1. Unconstrained (Configural)	30	13.708	12	1.142	1.000		.008
2. Measurement Weights	26	13.948	16	0.872	1.000	<b>0.000</b>	.000
3. Structural Covariances	23	16.282	19	0.857	1.000	<b>0.000</b>	.000
4. Measurement Errors	15	36.810	27	1.363	.998	<b>0.001</b>	.013

As can be observed in Table 4, according to the unconstrained model, the changes in CFI in the models obtained by constraining, in sequence, measurement weights, structural covariances and error variances, were less than .01. Hence, the measurement model has ensured configural, measurement and structural invariance based on sufficiency of the amount of weekly pocket money.

### Conclusion and Discussion

This study investigates the measurement invariance of the numerical ability and verbal ability tests across some subgroups, explained above. For this purpose, the configural, measurement (both measurement weights and measurement residuals) and structural invariance across the sub-groups was tested for the multiple-group measurement invariance of the constructed measurement model by using the covariance structural analysis approach. In numerous decisions to be taken in the field of education, the variable of ability is used as an important measurement. Using measurement invariance to test whether models constructed with psychological tests like ability operate in the same way in different subgroups and to report the findings are of utmost importance in order to avoid biases in serious decisions to make. As seen in the explanatory examples, the IBM AMOS-24 program is a very practical and user-friendly for measurement invariance and provides very detailed outputs regarding the invariance stages.

Having examined the measurement model invariance with respect to configural, measurement, and structural invariance across the groups in terms of students' gender, the amount of pocket money they received from their family, and their perceptions of whether the amount of pocket money was sufficient, the present study reached at the conclusion that both configural and measurement and structural invariances were ensured. The measurement model constructed with the data obtained from the numerical ability and verbal ability student tests developed by Pektaş (2018), by using the covariance structural analysis and has perfectly met all the possible measurement invariances. Studies that examined different sub-groups regarding such ability tests as

psychological tests can be encountered in the related literature. Wicherts (2016) mentioned that neurocognitive test batteries, such as the up-to-date version of Wechsler's Adult Intelligence Scale (WAIS), were used in population-based norms. The main fundamental question focused on whether the implemented test batteries operated in the same way with different sub-groups as gender, age, educational background, socioeconomic status, ethnicity, mother tongue, and race. Based on the studies they reviewed, Wicherts and Dolan (2010) reported that overlooking group variations based on ethnicity in intelligence tests can lead to a high degree of bias in terms of minorities, and that in comparisons based on group differences in intelligence tests, it is essential to initially analyze measurement invariance. In their study titled „Health and Retirement Study/Asset and Health Dynamics among the Oldest Old“ (HRS/AHEAD), Blankson and McArdle (2013) aimed to test the invariance of cognitive variables across ethnic origin, gender and time. The analyses were done using a selected sub-sample of the HRS/AHEAD data set. Metric invariance was ensured across time. By means of measurement invariance, the invariance of cognitive talent measurements based on HRS/AHEAD was better understood. Since measurement invariances were provided in our examples, partial measurement invariance analyzes or additional analyzes for measurement invariance sources were not included.

As a conclusion, the results of this study provide evidence that the measurement invariance requirement for valid group comparisons has been satisfied. Proving the invariances of errors as well in the measurement model constructed to test numerical and verbal abilities also proves that the reliability of these numerical and verbal ability tests developed for sixth grade students does not vary across the sub-groups examined. The present study on students also proves that the characteristics examined as a group variable or the particular group one had fallen in did not create any bias in terms of sub-scale and scale scores in the measurement of numerical and verbal abilities. Finally, the measurement invariance approach and the interpretations explained in this article can be applied to measurement tools in all studies aiming to compare different groups. Hence, as it explains and exemplifies measurement invariance at the same time, the article also sheds light on an important issue for measurements that will be used as support for important decisions to be taken.

## References

- Arana, F. G., Rice, K. G., & Ashby, J. S. (2018). Perfectionism in Argentina and the United States: Measurement structure, invariance, and implications for depression. *Journal of Personality Assessment, 100*(2), 219-230. <https://doi.org/10.1080/00223891.2017.1296845>
- Atkinson, R., Atkinson, R., & Hilgard, E. (1995). *Psikolojiye giriş II [Introduction to Psychology II]*. İstanbul: Sosyal.
- Bailey, S. K., Neigel, A. R., Dhanani, L. Y., & Sims, V. K. (2018). Establishing measurement equivalence across computer-and paper-based tests of spatial cognition. *Human factors, 60*(3), 340-350. <https://doi.org/10.1177/0018720817747731>
- Başusta, N. B., & Gelbal, S. (2015). Gruplararası karşılaştırmalarda ölçme değişmezliğinin test edilmesi: PISA öğrenci anketi örneği [Testing measurement invariance in intergroup comparisons: PISA student survey sample]. *Hacettepe University Journal of Education, 30*(4), 80-90.
- Benson, N. (2008). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment, 26*(1), 27-41. <https://doi.org/10.1177/0734282907301424>
- Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? Predicting academic achievements in advanced STEM studies. *Journal of Intelligence, 6*(4), 48. <https://doi.org/10.3390/jintelligence6040048>
- Blankson, A. N., & McArdle, J. J. (2013). Measurement invariance of cognitive abilities across ethnicity, gender, and time among older Americans. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 70*(3), 386-397. <https://doi.org/10.1093/geronb/gbt106>
- Borghans, L., Golsteyn B. H., Heckman J. J., & Humphries J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences, 113* (47), 13354–13359. <https://doi.org/10.1073/pnas.1601135113>
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Psychology Press.
- Byrne, Barbara M. (2016) *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.
- Camerota, M., Willoughby, M. T., Kuhn, L. J., & Blair, C. B. (2018). The childhood executive functioning inventory (CHEXI): Factor structure, measurement invariance, and correlates in US preschoolers. *Child Neuropsychology, 24*(3), 322-337. <https://doi.org/10.1080/09297049.2016.1247795>
- Camilli, G. (2006). Test fairness. In R. L. Brennan's (Ed.), *Educational measurement* (pp. 221–256). Westport: Praeger.
- Campbell, H., Barry, C. L., Joe, N. J., & Finney, J. S. (2008). Configural, metric and scalar invariance of the modified achievement goal questionnaire across African American and White university students. *Educational and Psychological Measurement, 68*(6), 988-1007. <https://doi.org/10.1177/0013164408318766>
- Chavez, L. M., Shrout, P. E., García, P., Forno, E., & Celedón, J. C. (2018). Measurement invariance of the adolescent quality of life-mental health scale (AQOL-MHS) across gender, Age and Treatment Context. *Journal of Child and Family Studies, 27*(10), 3176-3184. <https://doi.org/10.1007/s10826-018-1158-5>
- Cheung, G.W. & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Chung, H., Kim, J., Park, R., Bamer, A. M., Bocell, F. D., & Amtmann, D. (2016). Testing the measurement invariance of the University of Washington Self-Efficacy Scale short form across four diagnostic subgroups. *Quality of life research, 25*(10), 2559-2564. <https://doi.org/10.1007/s11136-016-1300-z>
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*(4), 662-680. <http://dx.doi.org/10.1037/0021-9010.70.4.662>
- Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell-Horn-Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*(2), 200-233. <https://doi.org/10.1037/1045-3830.22.2.200>
- Gaddy, M. A., Casner, H. G., & Rosinski, J. (2018). Factor structure and measurement invariance of the Brief Addiction Monitor. *Journal of substance abuse treatment, 90*(7), 29-37. <https://doi.org/10.1016/j.jsat.2018.04.010>
- Ghanizadeh, A. (2017). The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education, 74*(1), 101-114. <https://doi.org/10.1007/s10734-016-0031-y>

- Kaufman, A. S. (2018). *Contemporary intellectual assessment: Theories, tests, and issues*. Guilford Publications.
- Kyllonen, P., & Kell, H. (2018). Ability Tests Measure Personality, Personality Tests Measure Ability: Disentangling Construct and Method in Evaluating the Relationship between Personality and Ability. *Journal of Intelligence*, 6(3), 32. <https://doi.org/10.3390/jintelligence6030032>
- McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (p. 136–181). The Guilford Press.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <http://dx.doi.org/10.1007/BF02294825>
- Millfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111-121.
- Millsap, R., & Tein, J. Y. (2004). Model specification and identification in multiple-group factor analysis of ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479-515. DOI: 10.1207/S15327906MBR3903\_4
- Özğüven, İ. E. (2007). *Psikolojik testler [Psychological tests]*. Ankara: Nobel.
- Pektaş, S. (2018). *The effects of differential item functioning determination methods on test parameters estimates, decision studies, g and phi coefficients*. (PhD Theses). Ankara: Gazi University Graduate School of Educational Sciences Department of Measurement and Evaluation in Education.
- Sak, U. (2014). *Üstün zekâlılar: Özellikleri tanulanmaları eğitimleri [The gifted: Traits identification training]*. Ankara: Vize.
- 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.
- Türk Dil Kurumu (2016). *Yetenek [Talent]*. It is obtained on 26.10.2019 at [http://www.tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.5c27e8237eb1a0.18983524](http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.5c27e8237eb1a0.18983524).
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. doi:10.1177/109442810031002
- Watson, A. M., Thompson, L. F., & Meade, A. W. (2007, April). Measurement invariance of the job satisfaction survey across work contexts. paper presented at the 22nd annual meeting of the society for industrial and organizational psychology, New York. It is obtained at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.583.5341&rep=rep1&type=pdf>.
- Wicherts, J. M. (2016). The importance of measurement invariance in neurocognitive ability testing. *The Clinical Neuropsychologist*, 30(7), 1006-1016. <https://doi.org/10.1080/13854046.2016.1205136>
- Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, 29(3), 39-47. <https://doi.org/10.1111/j.1745-3992.2010.00182.x>
- Wu, D. A., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1-26.
- Xu, H., & Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *The European Journal of Counselling Psychology*, 6(1), 75-82. <http://dx.doi.org/10.23668/psycharchives.2041>
- Yeşilyaprak, B. (2007). *Eğitimde rehberlik hizmetleri [Counseling services in education]*. Ankara: PegemA.