



Received: 30.05.2018
Received in revised version: 25.06.2018
Accepted: 30.06.2018

Şekercioğlu, G. (2018). Measurement invariance: Concept and implementation. *International Online Journal of Education and Teaching (IOJET)*, 5(3). 609-634.

<http://iojet.org/index.php/IOJET/article/view/439/257>

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu 

Akdeniz University, Turkey

guclus@akdeniz.edu.tr

Güçlü Şekercioğlu has had his Ph.D. in Measurement and Evaluation from Ankara University, Turkey. He is a teaching staff at Akdeniz University, Education Faculty. He conducts research in multivariate statistics, measurement invariance, differential item functioning and psychological measurement in culture context.

Copyright by Informascope. Material published and so copyrighted may not be published elsewhere without the written permission of IOJET.

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu

guclus@akdeniz.edu.tr

Abstract

An empirical evidence for independent samples of a population regarding measurement invariance implies that factor structure of a measurement tool is equal across these samples; in other words, it measures the intended psychological trait within the same structure. In this case, the evidence of construct validity would be strengthened within the frame of the scores obtained from the tool. When measurement invariance is not supported, the researchers should consider the possibility of the different factor designs for each group. Ignoring such a situation brings forward the probability about differentiation of the trait(s) measured by measurement tool for that/those group(s), so it causes to suspect the validity of the scores obtained from the tool. The aim of this study is to examine measurement invariance in the context of the conceptual foundations of multi-group confirmatory factor analysis, and discuss the subject through the results from two hypothetical data set that one supports measurement invariance, but the other does not. As a result of analysis performed in this direction, it is determined that the five-factor design derived from the first data set is equal across the groups in the majors of science, health, and social science. It is also concluded that the three-factor design obtained from the secondary data set is not equal for female and male groups. Besides, the exploratory factor analysis performed for female and male groups separately shows that the three-factor design of the tool is valid for females, but the number of factors was four in males. When the factor design for male group is examined, it is determined that the three items in the second factor separate significantly. That leads to the conclusion that it is crucial to test measurement invariance in studies regarding the determination of the psychometric properties of the tool.

Keywords: measurement invariance, equality of factor structures, multi-group confirmatory factor analysis, structural equation modeling

1. Introduction

The major problem in behavioural and educational science studies, which aim developing the psychological measurement tools, cultural adaptation of a tool developed in another culture, using the tool for a different purpose or for a different sample, is to demonstrate the validity of the empirical evidence on the psychometric properties of the tool. In this direction the researchers, within the framework of these fundamental problems, are obliged to question whether the tool measures the trait(s) what it intends to measure properly and precisely. Further examination related to psychometric properties of measurement tool and all other analyses based on the scores obtained from the measurement tool (ANOVA, regression, etc.) has been carried out after the validity of the evidence put forward and decision are taken in this direction. According to Nunnally and Bernstein (1994), the validity of each usage must be documented by empirical evidence even though a measurement result may be valid for more than one purpose. Therefore, the test authors and users should not assume that the validity of evidence cannot change (Crocker and Algina, 1986).

One of the most important dimensions of the validity of scores obtained from psychological measurement tools is the construct validity. In the report of testing standards published in 1954 it was discussed that the concept of validity, actually all types of validity should be assembled under the roof of construct validity (Cronbach and Meehl, 1955; Jonson and Plake, 1998; Urbina, 2004; Westen and Rosenthal, 2005). Similarly, Kline (2000) states that the construct validity includes other approaches as well, thus all types of validity are related to the assessment of construct validity. The factor analysis is one of the most commonly used techniques in the studies which aim to determine the psychometric properties of a measurement tool in behavioural and educational science, in order to obtain evidence of construct validity. According to Büyüköztürk (2002, 2014) the factor analysis is a multivariate statistics, which aims to find and explore conceptually meaningful fewer new variables (factor, component) by bringing a large number of inter-related variables together. The factor analysis can be considered under two headings, which are exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) discussed under the concept of structural equation modelling (SEM).

CFA, which is based on testing of theories about the latent variables, and used in advanced research, is a very sophisticated technique (Ullman, 2001). In this analysis, a predefined and constrained construct is tested whether it is confirmed as a model. It is also occasionally used to mean the confirmation of the theoretical structure (Maruyama, 1998). In this context, the determination of the construct validity for CFA is emphasized as a very powerful method (Floyd and Wideman, 1995; Kline, 2005; Stapleton, 1997).

Multi-group confirmatory factor analysis (multi-group CFA) is also a specific practice area in CFA. This analysis enables to test the equality of structural parameters for more than one group simultaneously. In this context, the assessment of equality between the groups in terms of factor structure is also termed as *measurement invariance*. Additionally, examining the fitness of structure brings about the concept of testing population heterogeneity. It is possible to encounter different terms for different tests of measurement invariance tests including equality test of factor structures, metric and factorial invariance in the literature (Brown, 2006).

Nowadays, interest of the researchers in social sciences towards measurement invariance is gradually increasing. In a plain definition, measurement invariance is the description of whether the structures of measurement tool are equal across individuals from different groups. This concept has a critical importance in comparing groups. When measurement invariance is not supported between the groups, it is not possible to interpret the findings that reveal differences concerning these groups. If the researcher does not have the evidence for measurement invariance, then the existence of different psychometric responses for scale items more than one group cannot be known. Measurement invariance analyses are used in intercultural comparison for groups speaking different languages in a culture, scale adaptation studies, the comparison of groups with different academic achievement, the comparison of employee groups in different areas of industry, comparisons based on gender and are also used to compare a control group and an experimental group in empirical research (Cheung and Rensvold, 2002). The frequently asked question on the use of psychological measurement tools is whether the factor design ensued as a result of factor analysis of the measurement tools valid for groups, which differentiated at such a level that may impact the measurement process concerning the ethnic characteristics, age or the way they respond to the items. In fact, the fundamental issue here is whether the measurement tool measures the same structural properties for different groups or not. When the factor structure is not equal across groups, naturally it is not possible to make meaningful comparisons between groups based on the factor scores. On the contrary, when measurement equivalence is supported

empirically, it is concluded that the group differences are completely reflected in terms of latent traits evaluated by factors. In this concept, the studies, which aim at determining equality of the measurement tools' factor structure, are becoming more and more substantial because cultural, developmental and contextual impacts related to the psychological structural traits have become well-known by the researchers recently (Floyd and Widaman, 1995). In addition to social science studies, using multi-group CFA becomes increasingly common in other majors such as psychology, education, management and organization, marketing, and communication, especially ones which carried out studies based on cross-cultural comparisons.

As Jöreskog, Sörbom and Toit (2000) claim, the factor structure of developed or adapted scale based on fundamental data set obtained from different groups or samples can be tested whether it is equal for more than one group or not concerning the national, territorial, regional, cultural or socio-economic status of the groups. It is highly functional to test the equality of factor structures for a scale or different numbers of items for more than a group. Thus, factors or structural relationships can be tested simultaneously whether they are equal across different samples (Baumgartner and Steenkamp, 1998).

According to Marcoulides and Schumacker (1996) in multi-group CFA, the question of "is each group measured under the same structure?" is investigated and this examination is carried out within the framework of the measurement model defined in advance. Similarly, Kline (2005) stated that, the focus of multi-group CFA is to test whether measurement invariance is supported for different groups within the same latent variables. This concept is defined as invariance of the psychometric properties of a scale across groups in the context of modelling in the literature of psychometry.

Determining whether the measurement invariance is supported for different groups or not has a critical role in the development of psychometric properties of psychological measurement tools. That implies whether the items of the same structure and all structures can be used for the sub-groups of a population. Likewise, the subject of testing measurement invariance plays a crucial role in terms of defining the generalizability of psychological structure across groups with different variables such as different cultures, age groups and genders. The equality tests of latent means, which are included in the analysis group, shows similarity with the comparison of observed group averages through t-test and ANOVA (Brown, 2006).

According to Byrne (2006), the researchers often seek answers to any of the following five questions for evidence related to the multi-group equality: (i) do certain structures of the items on the measurement tool work equally across different groups? In other words, does the measurement model have a group equality? (ii) is the factor structure of the tool or theoretical structure measured by multiple scales equal for each level of the group? (iii) are the paths of the experimental structures equal across the groups? (iv) does the latent means in the model for a particular structure vary between groups? and (v) is the factor structure of a measurement tool equal for independent samples of the population? The author particularly emphasizes that there could be a cross-validation study in his last question. The analysis results reach the conclusion that if the factor structures are not equal between the groups, the validity of interpretations based on a comparison of scores for these groups decreases.

According to Brown (2006), the process steps below should be followed in the evaluation of the multi-group CFA and measurement invariance: (i) performing CFA for each group included in the analysis separately, (ii) testing the equality of structures simultaneously (factor loadings, factor correlations and error variances constant), (iii) testing the equality of the factor structures (factor loadings free; factor correlations and error variances constant),

(iv) testing the equality of factor structures and the error variances indicators (factor loadings and error variances free; factor correlations constant), (v) testing the equality errors variances of indicator (error variances free, factor loadings and factor correlations constant), (vi) testing the equality of factor variances, (vii) testing the equality of factor covariances (if more than one factor), and (viii) testing the equality of latent means. Hereunder, the first step is one of the multi-group CFA's assumption. The processes between the second and fifth steps are about testing measurement invariance, and the processes between the sixth and eighth steps are about testing the population heterogeneity.

1.1. Measurement Invariance Test and Models

Before computing the multi-group CFA, first of all, correlation or covariance matrix of the groups in the same sample is evaluated by comparing each other. In other words, before setting up the configural invariance model (Model 1), the establishment of the test equality of covariance matrices (Model 0) must be made. If the equality of covariance matrices is provided for each group ($\Sigma^g = \Sigma^{g'}$) the configural invariance model can be developed and tested. The equality of covariance structures of the groups should be discussed only after the null hypothesis (H_0) has been rejected. Subsequently, the models for other hypothesizes should be tested separately. The configural model derived from different groups should be defined in the same sample. Thus, the defined model for each group of multi-group analysis would be simultaneously tested. In this case, it is expected to see high fitness between correlation or covariance matrices of different groups (Brown, 2006; Byrne, 2006; Dunn, Everitt and Pickles, 1993; Vandenberg and Lenca, 2000). In general, the measurement invariance is tested with four basic models. These models are summarized in Table 1 (adapted from Cheung and Rensvold, 2002).

Table 1. *Measurement invariance models*

Models	Hypothesis	Hypothesis Name	Symbolic Statement	Process
1	H_{form}	Configural invariance	$\Lambda_{form}^{(1)} \therefore \Lambda_{form}^{(2)}$	Invariance is supported for all groups regarding construct and items. Factor loadings, factor correlations, and error variances are equal for all groups.
2	H_{Λ}	Weak Factorial Invariance (Metric Invariance)	$\Lambda^{(1)} \therefore \Lambda^{(2)}$	Invariance is supported for all groups regarding factor correlations and error variances. The factor loadings have been freed for groups.
3	H_{λ}	Strong Factorial Invariance (Scalar Invariance)	$\lambda_{ij}^{(1)} \therefore \lambda_{ij}^{(2)}$	Invariance is supported for all groups regarding factor correlations. The factor loadings and error variances have been freed for groups.
4	$H_{\Lambda, \theta(\delta)}$	Strict Factorial Invariance (Residual Variance Invariance)	$\theta_{\delta}^{(1)} \therefore \theta_{\delta}^{(2)}$	Invariance is supported for all groups regarding factor loadings and correlations. The error variances have been freed for groups.

1.1.1. Configural invariance (baseline model)

Developing a configural invariance (also known as baseline model) begins with identifying and testing the model, which was developed within the framework of a specific hypothesis for each group. In this context, the number of sub-scales in configural invariance model for each group (e.g. factors), the positions of the items (e.g. which factors the items are loaded) and correlations between sub-scales (e.g. setting such factors covariance) are determined. Secondly, the validity of the configural invariance model is tested separately for each group. Ideally, the model is expected to well fit and significant. However, the evidence, which shows a well fit, provides the information to the researcher that only the factor structure is similar but does not give any certain information about the equality of factors for each group. The evidence act as a design for invariance tests to be carried out subsequently. This model has two important functions: (i) the parameters are tested simultaneously for all groups, (ii) equal initial value is generated for the integration of configural invariance model

for testing (Byrne, 2008). Hence, the criterion, which will be obtained from further models to be tested, is occurred. In this model, invariance, regarding structure and items are supported for all groups (factor loadings, factor correlations and error variances are equal for all groups). When weak, strong or strict factorial invariance hypotheses are rejected, the "factor structure is equal across all groups" hypothesis, which is developed within the framework of configural invariance, is accepted.

1.1.2. Weak factorial invariance (metric invariance)

In this model, the equality of factor loadings (λ), ($\lambda^1 = \lambda^2 = \dots = \lambda^G$) is tested for all groups. (Spini, 2003; Vandenberg and Lence, 2000). If the fit, which is obtained from weak factorial invariance test, is better than the fit of configural invariance, configural invariance model is rejected. In other words, it indicates that the equivalence is not supported. According to Byrne and Stewart (2006) although measurement units are identical for groups in terms of underlying factors (e.g. factor loads), it constitutes one of the constraints of this model because scaling (e.g. intercepts) is not identical. Therefore, Meredith (1993) describes this invariance level as *weak factorial invariance*. This invariance is tested,

$$M_g \cong \hat{\tau}_g \hat{\tau}'_g + \hat{\Lambda}(\hat{\alpha}_g \hat{\alpha}'_g + \hat{\Psi}_g) \hat{\Lambda}' + \hat{\Theta}_{\varepsilon g} = \hat{M}_{g'}$$

with this equation (Widaman and Reise, 1997).

1.1.3. Strong factorial invariance (scalar invariance)

It is tested whether the regression constant (τ) of observed variables on the latent variables is equal across groups ($\tau^1 = \tau^2 = \dots = \tau^3$) (Schmitt and Kuljanin, 2008). In this model, there are a series of additional constraints described in weak factorial invariance. These additional constraints include the intercepts of the variables that are observed in the matrices $\hat{\tau}_g$. If estimations are problematic in terms of invariance on groups, subscript g on matrix τ is removed. In this case, invariance is tested,

$$M_g \cong \hat{\tau} \hat{\tau}' + \hat{\Lambda}(\hat{\alpha}_g \hat{\alpha}'_g + \hat{\Psi}_g) \hat{\Lambda}' + \hat{\Theta}_{\varepsilon g} = \hat{M}_{g'}$$

with this equation (Widaman and Reise, 1997).

1.1.4. Strict factorial invariance (residual variance invariance)

In this last model of the measurement invariance, about error terms across the groups $H_{\Lambda\phi}$ model limits ($H_{\Lambda\phi\theta}$) model equally ($\theta^1 = \theta^2 = \dots = \theta^G$). With the addition of this constraint, testing the hypothesis of equality of measurement errors becomes possible for independent samples of the population. If the error variances are equal, it means the items have equal reliability in terms of groups (Spini, 2003). Strict factorial invariance is also created through strong factorial invariance as it occurs in strong factorial invariance created through the weak factorial model constraints. These additional constraints are defined as strict factorial invariance, which contains unique factorial invariance in $\hat{\Theta}_{\varepsilon g}$ matrix or measurement errors. This invariance is tested,

$$M_g \cong \hat{\tau} \hat{\tau}' + \hat{\Lambda}(\hat{\alpha}_g \hat{\alpha}'_g + \hat{\Psi}_g) \hat{\Lambda}' + \hat{\Theta}_{\varepsilon} = \hat{M}_{g'}$$

with this equation (Widaman and Reise, 1997).

It should be noted that there are various classification in the related literature. Therefore, it is worth to consider following aspects suggested by Meredith (1993) and Dimitrov (2010), in the testing process of the equality of factor structures across groups, metric invariance is the general name of weak factorial invariance, strong factorial (scale invariance) and strict factorial invariance (invariance of error variance) models. However, there are some research

in literature that discuss the weak factorial invariance with the term of metric invariance. (Gregorich, 2006; Meade, Michels and Lautenschlager, 2007; Schmitt and Kuljanin, 2008; Spini, 2003; Vandenberg and Lance, 2000; Wu, Li and Zumbo, 2007). Besides, Cheung and Rensvold (2002) used the terms *metric invariance on construct-level* for weak factorial invariance, *item-level metric invariance* for strong factorial invariance and *error variance invariance* for strict factorial invariance.

Multi-group CFA for measurement invariance can be computed with such software statistical programs like LISREL, Amos, SAS/STAT, Mplus and EQS. The analysis starts with the creation of separate covariance matrices for the levels of the groups. It can be carried out by typing the syntax analysis in LISREL program or by following the instructions prescribed by the program (Toit and Toit, 2001). Measurement invariance is carried out in four models. The syntax samples of these models are named as EX10A.SPL, EX10B.SPL, EX10C.SPL, and EX10D.SPL in LISREL program. In the first model (Model 1), also known as configural invariance model, factor loads of structure(s), correlations and error variance are assumed to be equal and the analysis is run in this regard. The configural invariance model, which is a fundamental model for the equality of factor structure, is developed with the hypothesis that factor structures are equal ($H_0=There\ is\ no\ difference\ between\ factor\ structures$). In order to make comparisons with model defined in the analysis, a second alternative model named as weak factorial invariance model (Model 2) is analysed. In the weak factor invariance model, freeing the factor loads for each group, keeping the factor correlations and error variances constant are discussed. In the third alternative model strong factorial invariance (Model 3) factor loads and error variances for each level of the group are released, factor correlations are kept constant. The last and fourth model of measurement invariance is strict factor invariance model (Model 4). In this model while error variances are released, factor loads and factor correlations are kept constant (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001).

1.2. Model Comparisons in the Decision of Measurement Invariance

In multi-group CFA invariance test, constrained and unconstrained model are compared. In terms of availability of different values for each group in constrained model, model parameters (e.g. factor loads) are not constrained in this model. The parameters have the same value for all groups in constrained model. When the fit of unconstrained model is better than the constrained one, it implies that constrained model is incorrect. In other words, if the unconstrained model fits better when the constrained parameters are released, they are allowed to get different values for each group, and the constrained model developed within the invariance hypothesis framework is rejected (Cheung and Rensvold, 2000).

For comparisons of models with multi-group CFA in the studies in which measurement invariance is tested, it can be said that there are two widely used approaches in literature. The first one is the comparison between configural invariance model developed with the hypothesis that there is no difference in factor structure for each group and alternative models (e.g. weak factorial invariance, strong factorial invariance and strict factorial invariance models). Hereunder, the first comparison is made between configural invariance model and weak factorial invariance model (model 1 and 2), the second is between configural invariance model and strong factorial invariance model (model 1 and 3), and the last one is between configural invariance model and strict factorial invariance model (model 1 and 4). According to this approach, in the case of equality of fit between any alternative model and configural invariance or in the event of deterioration, the configural invariance model developed with the hypothesis that there is no difference in factor structure for each group is accepted. On the other hand, if the alternative model indexes differ from the configural invariance indexes

significantly (in favour of alternative models), H_0 hypothesis is rejected. In this case, the equality of factor structure and thus, the measurement invariance cannot be supported (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001). In the second approach, the comparisons are performed by following stepwise process. According to this, the analysis starts with less limited models and then the models are assessed by using nested χ^2 method (Brown, 2006). Accordingly, in comparison to nested models; $H_{form} > H_{\Lambda} > H_{\lambda} > H_{\Lambda, \theta(\delta)}$ is used as base. In other words, comparisons are made between configural invariance model and weak factorial model (model 1 and 2), weak factorial invariance model and strong factorial invariance model (model 2 and 3), strong factorial invariance and strict factorial invariance (model 3 and 4). According to Van de Vijyer and Leung (1997) if the fit of nested models is equal, more constrained model is frequently accepted. If this is not the case, the equality hypothesis is rejected (as cited in Spini, 2003). Cheung and Rensvold (2002) also suggest another comparison containing only one difference from the first approach. Although the first two comparison is the same, the authors suggest a comparison between weak factorial invariance and strict factorial invariance (model 2 and 4).

1.3. Decision Making of Measurement Invariance

While deciding whether the factor structures are equal for each group, the significance level of χ^2 matrix is required and the level should above .05 value, in other words, a non-significance value p is expected. This situation means that the covariance matrix of each of the defined groups do not differ significantly, thereby measurement invariance is supported. According to Jöreskog and Sörbom (1993), examples of acceptability of fit indices provided in Table 2 might be used for decision.

Table 2. *Acceptance of equality of factor structure in multi-group confirmatory factor analysis*

Problem	χ^2	df	p value	Decision
A	38.08	10	0.000	Reject
B	1.52	2	0.468	Accept
C	8.77	4	0.067	Accept
D	21.55	8	0.006	Reject
E	38.22	11	0.000	Reject

As seen in Table 2, models A, D, and E in which significance value p is a problem, are rejected whereas problem B and C are accepted. The criteria determined in the developing first years of multi-group CFA have been questioned over time. χ^2 has a possibility to increase its significance value if the number of samples increases, therefore, alternative models are investigated whether to accept the fit of factor structures within the model framework or not to assess the fit between covariance matrices. Among these, firstly, the value of χ^2 and degree of freedom should be compared. In this regard the χ^2 value obtained from the more constrained model, χ^2 value from less constrained model and the “delta” value (delta means the difference and its symbols is Δ) which is between the degree of freedom are calculated. $\Delta\chi^2$ and Δdf values are determined with this calculation. The significance level of χ^2 value obtained from this determination, is controlled in the level of $p < .01$ or $p < .05$ by comparing the critical values in the distribution table of χ^2 (Byrne, 2010; Jöreskog, 1971; Kline, 2005; Lee and Leung, 1982; Steiger, 2007; Van den Bergh and Van Ranst, 1998). In this case, H_0 and H_1 hypotheses can be developed in the following format:

H_0 : *There is no significant difference between the more constrained model and less constrained models in terms of fit.*

H₁: There is a significant difference between the more constrained model and less constrained models in terms of fit.

In this respect, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is less than critical table values, H_0 is accepted. In other words, there is no significant difference between two models in terms of fit, therefore, the researcher can make a decision about measurement invariance based on $\Delta\chi^2$. On the other hand, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is more than critical table values, H_0 is rejected. Thus, there is a significant difference between two models in terms of fit, and if this difference is in favour of the alternative hypothesis, the researcher can assume that measurement invariance is not provided on the basis of $\Delta\chi^2$.

In many studies in which analysis of SEM concept is applied, the distribution(s) may be remote from normal within certain tolerances. In the absence of normality in large samples, χ^2 value (S-B χ^2) obtained from Satorra-Bentler correction produces close values to the χ^2 that is produced when the number of people in the sample and the distribution of the produces is normal. S-B χ^2 is a rather reliable statistical test used to evaluate covariance structure models in various distributions and sample sizes (Byrne, 2006; Everitt and Howell, 2005). As in the other SEM analyses, such as multi-group CFA, which is carried out to obtain evidence of measurement invariance, S-B χ^2 can only be calculated if the distribution of each group is far from the normal distribution. In multi-group CFA, which is carried out with the maximum likelihood method, T_s value should be calculated for S-B χ^2 scaled difference in terms of evidence of measurement invariance between nested models. T_s is calculated

$$T_s = (T_0 - T_1) / c_d$$

with this equation. T_0 is the normal maximum likelihood χ^2 value for nested model, T_1 is the normal maximum likelihood χ^2 value for comparison (less constrained model) model, and c_d is the degree of difference test correction. c_d is calculated

$$c_d = [(d_0 * c_0) - (d_1 * c_1)] / (d_0 - d_1)$$

with this equation. d_0 is the degree of freedom of nested model, d_1 is the degree of freedom of comparison model, c_0 is the correction degree of nested model, and c_1 is the correction degree of comparison model. c_0 and c_1 are calculated

$$c_0 = T_0 / T_0^* \text{ and } c_1 = T_1 / T_1^*$$

with this equation. T_0^* is S-B χ^2 value of nested model, on the other hand T_1^* is S-B χ^2 value of comparison model. By comparing T_s , which is calculated for S-B χ^2 difference degrees, with the critical values in χ^2 distribution table, it can be determined whether measurement invariance is supported (Brown, 2006; Satorra and Bentler, 2011).

Recently, it is widely used as an alternative to utilize from fit indices as well as to evaluate the χ^2 differences among the models in many research due to a large number of n. According to Cheung and Rensvold (2002) it is inadvisable to reject null hypothesis in case of obtaining an insignificant χ^2 value. χ^2 is statistically sensitive test for large samples, however, it is not a practical test for model fit. In such case, alternative fit indices should be offered for χ^2 . The comparative fit indices (CFI, NNFI / TLI, RMSEA etc.) are among the most frequently recommended ones. Within this framework, it is observed that many goodness of fit indices are commonly used together to evaluate general fit of the model and to report it. ΔGFI 's are used as an alternative for χ^2 in multi-group CFA which is performed to determine whether the factor structures are equal or not. As in χ^2 , the configural model whose factor loads, factor correlation and error variance are released in covariance matrices of groups, in other

words, the model which is developed with the hypothesis that factor structures are equal, is the basic model like in alternative fit indices. For the evidence of measurement invariance, the differences between models can be evaluated with the comparison of indices such as Δ RMSEA, Δ CFI, Δ Gamma Hat, Δ Mc, Δ IFI, Δ AIC, Δ EVCI, Δ NFI, Δ TLI, and Δ SRMR.

The fit values are expected to become better for the equality of factor structures when the parameters like factor loads and error variance in covariance matrices of the group are released together or one by one. With this regard, the differences are evaluated by comparing the indices (e.g. Δ SRMR, Δ CFI and Δ RMSEA) between configural model and other alternative models or nested models. The configural model set up with the hypothesis that there is no significant difference between factor structures of each group is accepted if the fit indices of alternative models are lower than the ones in configural model. On the other hand, if the fit indices of other alternative model are higher than the ones in configural model or nested model, the fit across models is evaluated whether it differs significantly or not.

Cheung and Rensvold (2000; 2002) suggested cut-off points for Δ CFI significance level between modes in terms of measurement invariance after carrying out a study by using Monte Carlo method. Hereunder, when Δ CFI \geq -.01 is provided, then configural invariance model is accepted. In contrast to this situation, if Δ CFI is between -.01 and -.02, there will be increasing doubt about invariance. If it is more than -.02 it can be said that the difference between constrained and unconstrained model will increase. In this situation, configural model is rejected. In this context, it is decided that the factor structures are not equal and therefore an alternative model should be sought. In addition, the critical values of Δ Gamma hat and Δ McDonald NFI are -.001 and -.02.

Chen (2007) suggested cut-off points for decision of measurement invariance by considering situations like sample size of CFI, RMSEA and SRMR indices and sample sizes in groups after carrying out a study, which aimed at testing sensitivity of goodness of fit indices through Monte Carlo method. Accordingly, it can be concluded that measurement invariance cannot be supported (case of noninvariance) if sample size is small ($n < 300$), sample sizes of groups are not equal, pattern of variance is the same, there is a relationship like Δ CFI \leq -.005, Δ RMSEA \geq .010 or Δ SRMR \geq .025 between groups in terms of weak factorial invariance test, and there is a relationship like Δ CFI \geq -.005, Δ RMSEA \geq .010 or Δ SRMR \geq .005 between groups in terms of strong factorial invariance or strict factorial invariance. On the other hand, measurement invariance can be supported when sample size is sufficient ($n > 300$), numbers of groups compared are equal, there is a relationship like Δ CFI \geq -.010, Δ RMSEA \geq .015 or Δ SRMR \geq .030 between groups in terms of weak factorial invariance test, and there is a relationship like Δ CFI \geq -.010, Δ RMSEA \geq .015 between groups in terms of strong factorial invariance or strict factorial invariance.

An important point to be considered in assessing multi-group CFA comparison of the four basic models is type I and type II error possibilities. If the sample is small for a null hypothesis, type I error is likely occurred. However, if the sample is getting larger for alternative hypothesis, the difference of fit will be extended. In that case, type II error is likely occurred. For this reason, to minimize the type I and type II error possibility, the cut-off points should be determined efficiently (Hu and Bentler, 1998). In their maximum possibility $\Delta\chi^2$ studies which were performed with the indicators acting as continuous variables, French and Finch (2008) controlled type I error in the level of .01 and .05 between different models and sample numbers. The researchers revealed that the power of $\Delta\chi^2$ has a positive correlation with sample size, indicator number of each factors and factor number. Meade and Bauer (2007) also extrapolated the same results about $\Delta\chi^2$ (as cited in Sass,

Schmitt and Marsh, 2014). There is no doubt that this case is valid for other delta fit indices as well. However, this study didn't include detailed discussions on that subject because it was beyond the scope.

1.4. Objectives

Researchers of behavioural and educational sciences provide evidence through a sample on the validity of scores obtained from developed or adapted psychological measurement tools. After revealing the psychometric properties of the measurement tools, measurement process can be practiced on an independent group in the same sample and various decisions may be taken by means of obtained scores in the same or a different study. The fact that a measurement tool with confirmed factor structure for a sample may not be valid for the independent sub-groups in the relevant sample is a probability that researchers should pay attention. In such a case, the validity of decisions to be taken with scores obtained from groups will be suspicions. Within this scope, this research aims to discuss the conceptual basis of multi-group CFA in measurement invariance in terms of basic concepts and to introduce the subject through two hypothetic data set that one supports measurement invariance, but the other doesn't, for the researchers aiming to determine the psychometric properties of a measurement tool. Thereby, a new perspective will be introduced to the researchers aiming to determine the psychometric properties of measurement tools, suggestions about decisions to be taken for the tool without equalized factor structure will be asserted. Accordingly, the present study searches answers to the following research questions:

1. Is the five-factor structure of measurement tool 1 equal across the groups of science, health and social sciences?
2. Is the three-factor structure of measurement tool 2 equal across groups of males and females?

This research is limited to measurement invariance (measurement of configural invariance, weak factorial invariance, strong factorial invariance and strict factorial invariance). The heterogeneity of the population (factor variance invariance, factor covariance invariance and latent means invariance) is not included in the research.

2. Method

This study examines the method of multi-group CFA for the evidence concerning measurement invariance through two data set consisting of equal and unequal factor structure. Considering the findings of the study, the current study has the characteristics of correlational research concerning equality of factor structure for independent groups in two samples and due to the discussions on generation of construct validity evidences. The correlational studies analyse the relationship between two or more variables without intervening in these variables under any circumstances. These studies are the ones that are effective on revealing the relationships and determining the levels of relationships between variables and provides necessary cues for conducting high-level research on these relationships (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz and Demirel, 2012).

2.1. Research Data

The ready-made data was used in this study. They consist of two data set (equal and unequal factor structured) that the researchers collected them from his previous researches. The first hypothetic data set that measurement invariance is supported consist of 666 undergraduate students. When the distribution of the participants is examined based on scientific major, 32.28% (n=215) science, 31.83% (n=212) of health and 35.89% (n=239) of social science. The other hypothesis data set that measurement invariance is not supported

consist of 353 high school students. The distribution in terms of participants' gender is as follows, 62.32% (n=220) female, 37.68% (n=133) male.

2.2. Data Collection Tools

The study consists of two hypothetical data set, which are the subjects of measurement invariance analyses and the scores obtained from two different measurement tools. Some items were emitted from the tool in line with the results of EFA and CFA that were run on the data set collected from the participants. Moreover, the factor design differed for male participants in the second data collection tool whose factor design was not equal. The main purpose of this study is not to determine or discuss the psychometric properties of aforementioned tools. However the present study focuses on presenting the multi-group CFA in terms of measurement invariance through two hypothetical data set in which measurement invariance both was supported and was not, and creating a new view of validity for the researchers who aim to measure the psychometric properties of a measurement tool. Therefore, it is not appropriate to give the names of the tools and sub-scales in view of the probability that because they can form basis for further studies. For this reason, the data collection tools were mentioned as measurement tool 1 and measurement tool 2, and limited information about the psychometric properties of the tools was given because it was not wanted to reveal the tool.

Measurement tool 1 is a tool that consists of five sub-scales to measure an effective trait through using four point rating. In the original study, EFA and CFA were performed to determine psychometric properties of the tool in terms of gathering evidence about construct validity, concurrent validity was examined by comparing with a criterion score, to obtain reliability evidence for stability a test-retest method was run, and lastly to obtain reliability evidence for internal consistency, Cronbach alfa coefficients were calculated. In conclusion, it can be said that the scores obtained from measurement invariance tool 1 have a high level of validity.

This study starts with EFA to obtain construct validity evidence through hypothetical data set of measurement tool 1. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem across items. Also, there is no missing value in hypothetical data set. As a result of EFA, it is determined that items of measurement tool 1 are gathered under five factors, and they are also under their own factors in parallel with the results of original study. Since an item had high factor loading in more than a factor, it was emitted from the analysis. Factor loading values of the items are between .40-.80. The contributions of items to the total variance are as follows; for first factor 10.63%, for second factor 10.02%, for third factor 8.87%, for fourth factor 8.03%, for fifth factor 6.94% and the total variance explained is 44.49%. In CFA results, which was performed to produce additional evidence for construct validity, the standardized coefficients of items which had a significant t value may change between .32-.70, and the error variance values may change between .50-.90. As a result of the analysis, it is determined that fit indices are $S-B\chi^2(366)=699.22$, $p=.000$, $\chi^2/df=1.91$, $RMSEA=.037$, $GFI=.92$, $NNFI=.96$ and $SRMR=.049$. It is observed that the Cronbach Alfa coefficients which were calculated to determine internal consistency of factor are for the first factor .75, for the second .78, for the third .72, for the fourth .69, for fifth .57. The total Cronbach Alfa coefficient of the tool is .84

Measurement tool 2 is a tool that consists of three sub-scale to measure an affective trait through using four-rating scoring. In original study, EFA was performed to determine psychometric properties of tool and to obtain construct validity evidence, the discriminant validity was investigated in the direction of the scores collected from two different groups. Item-test correlations were calculated to determine item discrimination, test-retest method

was applied to obtain reliability for stability and Cronbach Alfa coefficients were calculated to obtain reliability for internal consistency.

In this study, the analysis of measurement tool 2 through the hypothetical data set starts with EFA. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem between items. Also, there is no missing value in hypothetic data set. As a result of EFA, it is determined that the items are gathered under three factors. Some items are emitted from the analysis because they give low factor loading value ($\lambda^2 < .32$) or they are overlapped items. The factor loading of items ranges between .45-.75. The contributions of items to the total variance are as follows, for first factor 21.04%, for second 17.98%, for third 8.98%, and total 48%. In CFA results, which was performed to produce an additional evidence for the construct validity, the standardized coefficients of items which have significant t value may change between .45-.74 and their error variance may change between .45-.80. As a result of the analysis, it is determined that fit indices are $S-B\chi^2(227)=423.46$, $p=.000$, $\chi^2/df=1.87$, $RMSEA=.050$, $GFI=.89$, $NNFI=.97$ and $SRMR=.052$. It is seen that Cronbach Alfa coefficients, which were calculated to determine internal consistency of factor are for the first factor .89, for the second factor .84, and for the third factor .65. It is not necessary to calculate the total point within the frame of theoretical and logical view, so the whole scale was not calculated by Cronbach Alfa coefficient.

2.3. Data Analysis

To find answers to the research questions of study, EFA, CFA, Cronbach Alfa analysis, covariance matrices equality test and multi-group CFA were performed. The factor analysis aims to find a few but significant new (common) unrelated variable by combining the variables related with each other in p-variable situation (p-dimensional space). In other words, the factor analysis is a method in which common components are determined and construct dependence is dispelled (Diekhoff, 1992; Gorsuch, 1974; Tatlıdil, 1992; Thompson, 2004; Tucker and MacCallum, 1997). Factor analysis is a technique, which is used to confirm whether the items of a certain scale or sub-scale are gathered under a certain construct or factor (Gable and Wolf, 2001). Beyond reducing variable and naming the emerging factors, the EFA reveals whether the analysis results are similar to the structure of the theory (unobserved latent variables) that enables to figure out the behaviour. After the analysis, a query is made for determining whether the indicators, which are gathered under a certain factor, are indicators of theoretical construct. In CFA, it is firstly aimed to test and confirm the structural hypotheses regarding the relationships between variables. Within this frame, it is focused on examining the relationships between factors and variables, and the relationships between factors in this research through the hypothesis developed. Therefore, the researcher should have the information about the construct of variables that s/he defined in model before the analysis. By this way, the model can be based on a strong theoretical or empirical basis (Raykov and Marcoulides, 2008; Stevens, 1996). Multi-group CFA, which is a special application of CFA can test the measurement and equality of construct models for multi-groups (Brown, 2006). The factor loads of measurement tool consist of measurement properties related to the variables that include constants and error variances. The multi-group CFA makes comparison between two or more groups simultaneously possible by using covariance matrices that are calculated for each compared groups. Thus, measurement invariance or equivalence can be tested by putting equality constraints the parameters of groups (Harrington, 2009).

For the model comparisons in the studies in which the measurement invariance is tested through the multi-group CFA, the first approach of two common approaches is the

comparison between the structural model developed by the hypothesis that there is no meaningful difference between the factor structures for each compared group and the alternative models. In the second approach, the fit between the more constrained nested model and the least constrained comparison model is evaluated by following a stepwise process. Although researcher suggests that evaluation of difference between models should be made between nested models, the comparisons were made for each methods to increase sample numbers and Δ 's were evaluated in this study.

Additionally, cut of points for factor loading in EFA are accepted as $\lambda^2 \geq .32$; χ^2 level of acceptance in hypothesis test for significance as .05; since $n > 300$ in each data set the cut of points for multi-group CFA in measurement invariance run for three-model comparison as $\Delta CFI \leq -.01$; as $\Delta SRMR \leq .03$ for weak factorial invariance and as $\Delta SRMR \leq .01$ for strict factorial invariance.

LISREL sample syntax for covariance matrices is in appendix 1, LISREL sample syntax for four models, which are based on for measurement invariance is in appendix 2.

3. Findings

The five-factor structure of measurement tool 1 was tested to determine measurement invariance with multi-group DFA for the groups in the majors of science, health and social science. Before giving the findings of measurement invariance, test statistics, normality tests and reliability coefficients are given in Table 3 in terms of basis assumption of analysis.

Table 3. *Test statistics, normality tests and reliability coefficients of science, health and social science groups*

Major	Factor	n	Mean	Median	Mode	s	Range	Skewness	Kurtosis	α^1
Science	1	215	17.52	18	24	4.68	18	-.429	-.686	.74
	2	215	16.89	17	15	3.94	17	-.184	-.521	.82
	3	215	21.28	22	23	4.01	19	-.490	-.224	.75
	4	215	16.19	16	20	3.06	13	-.672	-.060	.72
	5	215	15.42	16	17	3.31	15	-.796	.388	.65
	Scale	215	87.29	88	90	12.43	66	-.521	.374	.85
Health	1	212	17.07	18	20	4.80	18	-.422	-.681	.76
	2	212	17.65	18	19	4.03	18	-.433	-.224	.79
	3	212	20.97	21	23	4.42	20	-.637	.135	.73
	4	212	16.09	17	20	3.07	13	-.605	-.129	.68
	5	212	15.14	16	17	3.41	15	-.544	-.202	.62
	Scale	215	86.91	89	97	13.26	66	-.536	.014	.85
Social	1	239	17.79	18	24	4.49	18	-.374	-.766	.75
	2	239	17.37	18	16	4.03	17	-.188	-.676	.76
	3	239	21.89	22	21	4.17	17	-.406	-.674	.74
	4	239	16.27	17	20	3.14	15	-.805	.507	.72
	5	239	14.57	14	13	3.36	15	-.313	-.225	.56
	Scale	239	87.90	88	86	13.05	63	-.303	.014	.85

¹ Cronbach Alfa internal consistency coefficient

As can be seen in Table 3, measures of central tendency are relatively close to each other for the groups in the majors of science, health and social science in the level of both sub-scale scores and total scale scores. The fact that skewness and kurtosis coefficients are in the range of ∓ 1 indicate that the distribution is close to normal (Rosenthal and Rosnow, 2008). Although the coefficients are between ∓ 1 , it can be said that all of the sub-scales and total scale score points are partly negatively skewed distribution. Accordingly, multi-group CFA, which was performed to determine whether measurement invariance was provided or not for all groups, was computed through asymptotic covariance matrix and S-B χ^2 statistics was

used as base for model fit. On the other hand, it is seen that internal consistency coefficients of science, health and social science groups, which were calculated based on sub-scale and scale scores, are generally in an acceptable level. According to Nunnally and Bernstein (1994), the reliability coefficient may be accepted for the research if the value is between .70-.80. In all groups, .70 condition is fulfilled with the factors 1-2-3, and 4 at a level of scale. However, this acceptance cannot be provided at the level of factor 5 in all groups. It can be thought that the internal consistency coefficient of sub-scale is low because the number of items is low. The equality of covariance matrices in science, health and social science groups was tested before multi-group CFA.

As a result of the analysis, index values of fit between covariance matrices related to the groups are shown in Table 4.

Table 4. *The equality of covariance matrices of science, health and social science groups*

Groups	S-B χ^2 (df)	p	χ^2 /df	RMSEA	GFI	CFI	SRMR
Science, Health and Social	1025.2(870)	.000	1.178	.028 (.020-.035)	.91	.98	.060

As can be seen in Table 4, S-B χ^2 and degree of freedom are below 2, RMSEA is below .05, GFI is above .090, CFI is above .95 and SRMR is below .08. In this situation it can be said that there is a fit between three covariance matrices.

Multi-group CFA findings for five-factor structure equality of the measurement tool 1 are given in Table 4 for science, health and social science groups.

Table 5. *Findings of multi-group confirmatory factor analysis for science, health and social science groups (maximum possibility)*

	S-B χ^2 (df) ¹	MC ²	$\Delta\chi^2$ (Δ df)	χ^2 /df	$\Delta\chi^2$ / Δ df	CFI	Δ CFI	SRMR	Δ SRMR	Decision
Science	522.35(366)	–	–	1.427	–	.95	–	.073	–	–
Health	477.91(366)	–	–	1.306	–	.97	–	.064	–	–
Social	536.72(366)	–	–	1.466	–	.95	–	.068	–	–
Model 1 ^A	1735.47(1234)	–	–	1.406	–	.95	–	.079	–	–
Model 2 ^B	1671.52(1176)	M1–M2	63.95(58)	1.421	-.015	.95	0	.075	.004	H ₀ Accept
Model 3 ^C	1859.75(1205)	M1–M3	-124.28(29)	1.543	-.137	.94	.01	.081	-.002	H ₀ Accept
		M2–M3	-188.23(-29)	–	-.122	–	.01	–	-.006	H ₀ Accept
Model 4 ^D	1920.04(1265)	M1–M4	-184.57(-31)	1.518	-.112	.94	.01	.085	-.006	H ₀ Accept
		M3–M4	-60.29(-60)	–	.025	–	0	–	-.004	H ₀ Accept

¹ p<.05

² Model comparison (M=Model)

^A Configural Invariance (Factor loads, factor correlation and error variance are constant)

^B Weak Factorial Invariance (Factor loads, factor correlation and error variance are constant)

^C Strong Factor Invariance (Factor loads and error variance are free, factor correlation is constant)

^D Strong Factorial Invariance (Error variance is free, factor loads and factor correlation are constant)

Firstly, when the fit indices obtained as a result of CFA which was performed separately for science, health and social science groups are examined, it can be said that fit indices obtained from each of the three groups largely meet the acceptance levels. Accordingly, it can be seen that S-B χ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. After analysing the fit indices in general, it can be said that for five-factor structure of the measurement tool 1 was confirmed separately for science, health and social science groups.

The configural model, which was developed with the hypothesis about there is no significant difference between factor loads, factor correlation and error variance for science, health and social science groups was tested to evaluate measurement invariance. The analysis results show that S-B χ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. Also, after analysing the fit indices in general, it can be acceptable that fit indices of configural model meet the acceptance levels.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that fit gets worse in terms of the ratio of $S-B\Delta\chi^2$ and Δdf . In addition, it can be said that there is no change in ΔCFI value and the change is not significant ($n < .025$) in $\Delta SRMR$.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared in terms of the ratios of $S-B\Delta\chi^2$ and Δdf , it's seen that the fit gets worse. Besides it can be said that the fit between models gets worse regarding the ΔCFI and $\Delta SRMR$ values. On the other hand, when weak factorial invariance (Model 2) and strong factorial invariance (Model 3) models are compared in terms of the second approach, between the ratios of both $S-B\Delta\chi^2$ and Δdf , ΔCFI and $\Delta SRMR$ values, it can be stated that the fit gets worse.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it can be seen that the fit gets worse in terms of $S-B\Delta\chi^2$ and Δdf ratios for both models. Besides it can be stated that the fit between models also gets worse in terms of ΔCFI and $\Delta SRMR$ values. On the other hand, according to the second approach when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, the fit in terms of $S-B\Delta\chi^2$ and Δdf ratios, gets better. In this direction T_s value calculated for difference ratio of $S-B\chi^2$ is 57.3 and it is confirmed that this value is smaller than the critical value in χ^2 distribution table, $\chi^2_{diff}(60)=79.08$, $p > .05$. Therefore, it can be said that there is no significant difference between strong factorial invariance and strict factorial invariance models. In other respects, it can be stated that there is no change in ΔCFI value and $\Delta SRMR$ value in the direction of the fit gets worse.

In the light of findings outlined above, among the four models, the model that works best based upon covariance matrices in the majors of science, health and social science is configural invariance model developed in assumption of the equality of factor structures. In this context, it is accepted that the five-factor structure of the measurement tool 1 is equal for relevant groups, in other words, measurement invariance is supported.

Measurement invariance for the three-factor structure of measurement tool was tested through multi-group CFA for both female and male groups. Before giving findings about measurement invariance, first in line with the basic assumption of the analysis, test statistics related to relevant groups, test of normality and reliability coefficients were given in Table 6.

Table 6. *Test statistics, tests of normality and reliability coefficients for female and male groups*

Gender	Factor	n	Mean	Median	Mode	S	Range	Skewness	Kurtosis	α^1
Female	1	220	9.57	8	0	7.45	28	0.51	-0.84	.89
	2	220	14.58	15	12	6.31	26	-0.05	-0.78	.83
	3	220	4.30	4	3	3.14	12	0.55	-0.53	.68
Male	1	133	8.99	7	3	7.39	30	0.82	-0.23	.87
	2	133	13.84	12	9	7.13	27	0.29	-1.04	.84
	3	133	4.55	4	3	3.01	12	0.44	-0.44	.61

¹ Cronbach Alfa internal consistency coefficient

As seen in Table 6, it can be stated that measures of central tendency for female and male groups in the level of both sub-scale and scale total points is close. It is stated that except one distribution of coefficient of skewness and kurtosis, all other distributions even though it is between ∓ 1 , points, to some extent, are negative-skewed. Also it is seen in data set for male students that sub-scale points are out of ∓ 1 ; when kurtosis coefficient is calculated to kurtosis' standard error, the obtained value is still out of ∓ 1.96 . In this direction, multi-group CFA, which intends to find out whether the CFA and measurement invariance are confirmed

for each group, is performed over asymptotic covariance matrix and model fit was based on $S-B\chi^2$. On the other hand, it is seen that internal consistency coefficient calculated on female and male groups' sub-scale and scale points is generally in an acceptable level. In all groups, .70 condition meets with 1st and 2nd factors in the scale level but in both groups, this acceptance cannot be met in 3rd factor level. It can be concluded that the internal consistency coefficient of the relevant scale is low because the number of items (4 items) in the scale is low.

The equality of covariance matrices for female and male groups was tested before the multi-group CFA. In the result of analysis made, index values regarding the fit between covariance matrices for these groups are presented in Table 7.

Table 7. *Equality of covariance for female and male groups*

Groups	S-B χ^2 (df)	p	χ^2 /df	RMSEA	GFI	CFI	SRMR
Female & Male	374.04(276)	.000	1.355	.042(.029-.053)	.89	.99	.083

As can be seen in Table 7, the ratio of $S-B\chi^2$ to the degree of freedom is below 2, RMSEA is below .05, GFI is below .90, CFI is above .95 and SRMR is above .08. It can be stated that when fit indices are assessed in general and GFI and SRMR indices are taken into account, the fit between two variances is moderate.

The multi-group CFA findings related to the equality of the three-factor structure of the measurement tool 2 for female and male groups are shown in Table 8.

Table 8. *Multiple-group confirmatory factor analysis findings of female and male groups (maximum possibility)*

	S-B χ^2 (df) ¹	MC ²	$\Delta\chi^2$ (Δ df)	χ^2 /df	$\Delta\chi^2$ / Δ df	CFI	Δ CFI	SRMR	Δ SRMR	Decision
Female	349.31(227)	–	–	1.539	–	.98	–	.061	–	
Male	298.51(227)	–	–	1.315	–	.98	–	.068	–	
Model 1	732.94(503)	–	–	1.457	–	.97	–	.100	–	
Model 2	694.38(480)	M1–M2	38.56(23)	1.447	.010	.97	0	.083	.017	H ₀ Accept
Model 3	651.04(457)	M1–M3	81.90(46)	1.425	.032	.99	-.02	.076	.024	H₀Reject
		M2–M3	43.34(23)	–	.022	–	-.02	–	.007	H₀Reject
Model 4	691.77(480)	M1–M4	41.17(23)	1.441	.016	.98	-.01	.098	.002	H ₀ Accept
		M3–M4	-40.73(-23)	–	-.016	–	.01	–	-.022	H ₀ Accept

¹ p<.05

² Model Comparison (M=Model)

Firstly, it can be said when fit indices obtained as a result of CFA's performed separately for female and male groups are analysed, fit indices of both groups generally meet the level of acceptance. According to this, it is seen that $S-B\chi^2$ and degree of freedom ratios are below 2, CFI is above .95 and SRMR is below .08. It can be stated that when fit indices are assessed in general, the three-factor structure of the measurement tool is confirmed separately for female and male groups.

To evaluate the measurement invariance, factor loads, factor correlations, and error variances for female and male groups of the three-factor structure are initially tested with the configural invariance model based on the hypothesis asserting that there is no significant difference among the variables stated. In the result of the analysis made the ratio of $S-B\chi^2$ to the degree of freedom is below 2, CFI is above .95 but SRMR is above .08. When fit indices are assessed in general, fit indices for the configural invariance model meet the level of acceptance in general.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that the fit gets better in terms of $S-B\Delta\chi^2$ and Δ df ratio for both models. In this direction T_s value calculated for difference ratio of $S-B\chi^2$ is 41.02 and this value is bigger than the critical value in χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, p<.05. In other words,

there is a significant difference between configural invariance and weak factorial invariance models. On the other hand, it is seen that there is no difference in ΔCFI value, and the change in $\Delta SRMR$ is not significant ($<.025$). When the findings are evaluated in general, based on the results asserting that the difference is not significant in two out of three fit indices, it is decided that fit indices of the configural invariance and the weak factorial invariance do not differ from each other.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared, it is seen that the fit gets better in terms of $S-B\Delta\chi^2$ and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of $S-B\chi^2$ is 78.14 and it has been determined that this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(46)=62.83$, $p<.05$. Hence, there is a significant difference between configural invariance and strong factorial invariance models. Also, when Model 1 and 3 are compared, it can be said that ΔCFI value ($<-.01$) and $\Delta SRMR$ value ($>.01$) considerably change. On the other hand, according to the second approach, when weak factorial invariance (Model 2) and strong factorial invariance models are compared, the similar results are obtained with the first approach comparison. According to this, the fit between $S-B\Delta\chi^2$ and Δdf percentages gets better. In this direction T_s value calculated for difference ratio of $S-B\chi^2$ is 38.37 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, $p<.05$. In other words, there is a significant difference between weak factorial invariance and strong factorial invariance models. Also in the context of Model 2 and 3, ΔCFI value ($<-.01$) and $\Delta SRMR$ value ($>.01$) significantly differ.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it is observed that the fit gets better in terms of $S-B\Delta\chi^2$ and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of $S-B\chi^2$ is 37.5 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, $p<.05$. According to that, there is a significant difference between weak factorial invariance and strict factorial invariance models. But on the other hand it is seen that the change is not significant in ΔCFI ($=-.01$) and $\Delta SRMR$ values ($<.01$). When the findings are evaluated, since there is no significant difference in two of three fit indices, it was decided that fit indices of the strict factorial invariance model and the configural invariance model don't differ from one another. According to the second approach, when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, it can be said that the fit between two models gets worse in terms of $S-B\Delta\chi^2$ and Δdf ratio and ΔCFI and $\Delta SRMR$ values.

The findings above reveal that the best working model is the strong factorial invariance model among these four models. Accordingly, it has been accepted that the three-factor structure of the measurement tool 2 is not equal for female and male groups, and measurement invariance cannot be supported.

In this step, different exploratory factor analyses have been computed for the data set obtained from female and male groups. As a result of analysis carried out for the female group, it has been observed that the three factors structure of measurement tool 2 is valid for this group, there is no considerable difference in factor loads (.38 and .75) and the contribution (%49.56) of these factors to the total variance explained. On the other hand, the analysis results, which were performed for male group have revealed that the items have been gathered under four factors. As a result of analysis repeated for these four factors, it is revealed that three items belonging to the second factor is showed up as another factor. In the data set for male group, it has been seen that factor loads (.47 and .83) and the contribution of

the factors to the total variance (57.55%) goes up, however the increase in the total variance explained has occurred because of the rise in the number of factors. When the results of analysis are examined in general, the factor numbers for measurement tool 2 are three for the female group and four for the male group.

4. Discussion and Conclusion

Validity is a concept that is referred to inferences from trait(s) that it measures, but not behalf of the measurement tool's name. Not only the ones who develop and adapt measurement tool but also the researchers who use the tool with a different purpose or different sample from its initial purpose have certain responsibilities to reveal scientific evidences about the validity. However, it may not be sufficient for determining the psychometric properties of the measurement tool through conventional methods in all circumstances. For a particular measurement tool and group, it is a problematic issue for construct validity whether the factor structures that were confirmed empirically by the factor analysis, have the same meaning for independent sub-groups in the sample. Therefore, it might be required for the researchers to test validity for different groups in the sample. It is thereby a crucial psychometric problem to test the equality for specific groups obtained factor structure design by explanatory and/or confirmatory factor analyses in scale development or adaptation studies.

The researchers usually make comparisons across groups to create theoretical information or contribute to existent theoretical knowledge one and naturally want their decisions as correct as possible about the population that they wish to generalize according to their findings. The researchers should examine whether the factor structure of the tool is equal across groups because these comparisons are usually made by scores of the measurement tool. In contrast, in case of inequality of the factor structures defined in the measurement tool across groups, the group scores of these structures do not mean the same. When the factor structures are equal among groups, factor design will be the same for groups and thus it can be evaluated that the collected group scores from scale or sub-scale are valid. That leads to develop a new additional empirical evidence for construct validity of the measurement tool. On the other hand, if psychometric properties of the measurement tools are not equal for groups, the factor structures will vary from one sub-group to another, the comparisons made with scores of measurement tool and the decisions about group will be faulty. Also if there is no empirical evidence, the contribution of the research results to the theory will be doubtful. In this respect, it is beneficial to develop a different perspective on the construct validity, in the framework of this basic problem, in scale development or adaptation studies in the majors of behavioural and educational sciences. Hence, the researcher has the responsibility to test the equality of test scores from factor structure of the tool and sub-scale for the compared groups. Therefore, the evaluations related to empirical evidence about the validity of measurement tool are never considered as "last word". Validating a tool requires everlasting effort.

This study aimed at developing a sample for what sort of decisions could be made for two different factor designs that either of them supports measurement invariance. In this respect, it was determined that the five-factor structure of the measurement tool 1, regarded as a first sample, supported measurement invariance for the undergraduate students in the majors of science, health, and social sciences. Based on EFA and CFA results, it can be claimed that the five-factor structure of the tool has high validity for the entire of the sample. With the multi-group CFA, the relevant factor design was approved to be separately valid for students of science, medical and social sciences. This result recommends that this factor design is valid for the whole sample and all of the science majors (any of these groups) like science,

health, and social sciences separately, so that the measurement tool 1 has high construct validity for the groups that maintain their academic lives in different majors.

As a second sample, the three-factor structure of the measurement tool 2 supports measurement invariance for high school students within the context of gender. In such cases, the researchers should consider the relevant factor structure as unequal for groups, therefore, they need to take into account the different factor designs or possibility of bias for each comparison. Therefore, the researchers are suggested to run the EFA for each group. In fact, it has been revealed that three factor structure is valid for females but not in males, in which the items are gathered under four factor in this study. It is required to give cues to make the findings more meaningful (reason of hiding names of scales are explained in method section). In the EFA that is run for males, the three items of the second factor named as stress subscale have been observed to be gathered under a new factor. When these items are examined, three of them are about “negative reaction showed in blocking situation”. These items, which are symptoms of stress for females, are loaded under a new factor called “intolerantness to blocking” for males. This finding reveals that the structure for female and male groups are different, in other words, doesn’t have the same psychological meaning for these groups. The researchers can produce different forms of measurement tool in this situation (e.g. female form-male form) and suggest these forms for the ones who study in the sub-groups of the sample. Although this situation causes a problem for practicality, it plays a crucial role in the construct validity.

The researchers can test whether the measurement tool, based on the theoretical basis of the trait which it intends to measure is equal for more than a group or a sample like age, gender, socio-economic level, class, education level, academic major, subcultures in a society, international comparisons, experimental researchers, and different occupational groups. It is surely beyond doubt that the evidences about whether the measurement invariance is supported for different groups will strengthen psychometric properties of the tool, and will therefore increase the validity of the results presented in the direction of the research findings. As a result, the contribution to the production of theoretical knowledge or existing theoretical knowledge accumulation will enhance as well.

References

- Baumgartner, H. & Steenkamp, J. E. M. (1998). Multi-group latent variable models for varying numbers of items and factors with cross-national and longitudinal applications. *Marketing Letters*, 9(1), 21-35.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. (First Edition). New York: Guilford Publications, Inc.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32, 470-483.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2012). *Bilimsel araştırma yöntemleri* (11. Baskı). Ankara: Pegem Akademi.
- Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (On Dokuzuncu Baskı). Ankara: Pegem Yayıncılık.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Second Edition). New York: Taylor & Francis Group.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872-882.
- Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. (Second Edition). California: Sage Publications, Inc.
- Byrne, B. M. & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), 287-321.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504.
- Cheung, G., & Rensvold, R. (2000). *Testing measurement invariance using critical values of fit indices: A Monte Carlo study*. Retrieved May 20, 2004, from http://www.aom.pace.edu/rmd/cheung_files/cheung.htm
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. (First Edition). Orlando: Holt, Rinehart and Winston, Inc.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Diekhoff, G. (1992). *Statistics for the social and behavioral sciences: Univariate, bivariate, and multivariate* (First Edition). Dubuque, IA: William C. Brown Publishers.
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43, 121-149.
- Dunn, G., Everitt, B. & Pickles, A. (1993). *Modeling covariances and latent variables using EQS* (First Edition). London: Chapman & Hall.
- Everitt, B. S. & Howell, D. C. (2005). *Encyclopedia of statistics in behavioral sciences*. Chichester: John Wiley & Sons, Ltd.

- Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286-299.
- Gable, R. K. & Wolf, M. B. (2001). *Instrument development in the affective domain: Measuring attitudes, and values in corporate and scholl settings* (Second Edition). London: Kluwer Academic Publishers.
- Green, S. B., Salkind, N. J. & Akey, T. M. (1997). *Using SPSS for windows: Analyzing and understanding data*. New Jersey: Prentice Hall, Inc.
- Gregorich, S. (2006). Do self report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44(11-3), 78-94.
- Gorsuch, R. L. (1974). *Factor analysis*. (First Edition). Philadelphia: W. B. Saunders Company.
- Harrington, D. (2009). *Confirmatory factor analysis* (First Edition). New York: Oxford University Press, Inc.
- Jonson, J. L. & Plake, B. S. (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*. 58(5), 736-754.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood: Scientific Software International, Inc.
- Jöreskog, K. G. & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood: Scientific Software International, Inc.
- Jöreskog, K. G., Sörbom, D., Toit, M. & Toit, S. (2000). *LISREL 8: New statistical features*. Lincolnwood: Scientific Software International, Inc.
- Kline, P. (2000). *The handbook of psychological testing* (Second Edition). London: Taylor & Francis Group.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second Edition). New York: Guilford Publications, Inc.
- Lee, S. Y. & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, 47, 297-308.
- Marcoulides, G. A. & Schumacker, R. E. (1996). *Advanced structural equation modeling: issues and techniques* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.
- Maruyama, G. M. (1998). *Basics of structural equation modeling* (First Edition). Thousand Oaks, CA: SAGE Publications, Inc.
- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods*, 10, 322-345.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (Third Edition). New York: McGraw-Hill, Inc.

- Raykov, T. & Marcoulides, G. A. (2008). *An introduction to applied multivariate analysis* (First Edition). New York: Taylor & Francis Group.
- Rosenthal, R & Rosnow, R. L. (2008). *Essential of behavioral research* (Third Edition). New York: McGraw-Hill, Inc.
- Sass, D. A., Schmitt, T. A. & Marsh H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21, 167-180.
- Satorra, A. & Bentler, P. M. (2011). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika: A Journal of Quantitative Psychology*, 66(4), 507-514.
- Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210-222.
- Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.
- Spini, D. (2003). Measurement equivalence of 10 value types from the Schwartz value survey across 21 countries. *Journal of Cross-Cultural Psychology*, 34(1), 3-23.
- Stapleton, C. D. (1997). Basic concepts and procedures of confirmatory factor analysis. Austin: The Annual Meeting of the Southwest Educational Research Association.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42, 893-898.
- Stevens, J. (1996). *Applied multivariate statistics for social sciences* (Third Edition). New Jersey: Lawrence Erlbaum Associates, Inc.
- Tatlıdil, H. (1992). *Uygulamalı çok değişkenli istatistiksel analiz* (Birinci Baskı) Ankara: Engin Yayınları.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (First Edition). Washington: American Psychological Association.
- Toit, M. & Toit, S. (2001). *Interactive LISREL: User's guide*. Lincolnwood: Scientific Software International, Inc.
- Tucker, L. R. & MacCallum, R. C. (1997). Exploratory factor analysis. (Online Edition) Web: <http://quantrm2.psy.ohio-state.edu/maccallum/factornew.htm> adresinden 19 Şubat 2007'de alınmıştır.
- Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 653-771). Needham Heights, MA: Allyn & Bacon.
- Urbina, S. (2004). *Essentials of psychological testing* (First Edition). New Jersey: Wiley & Sons, Inc.
- 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.
- Westen, D. & Rosenthal, R. (2005). Improving construct validity: Cronbach, Meehl, and neurath's ship. *Psychological Assessment*, 17(4), 409-41.

- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.
- Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation, 12*(3), 1-26.

Appendix 1: LISREL Syntax Example for Test of Covariance Matrices Equality

SCALE I EQUALITY TEST OF COVARIANCE MATRIX (ACADEMIC MAJOR)

Group SCIENCE:

Observed Variables: V1-V30

Covariance Matrix from File SCIENCE.COV

Method of Estimation: Maximum Likelihood

Iterations: 20

Sample Size: 215

Latent Variables: f1-f30

Relationships:

V1 = 1*f1

V2 = 1*f2

V3 = 1*f3

V4 = 1*f4

V5 = 1*f5

V6 = 1*f6

V7 = 1*f7

V8 = 1*f8

V9 = 1*f9

V10 = 1*f10

V11 = 1*f11

V12 = 1*f12

V13 = 1*f13

V14 = 1*f14

V15 = 1*f15

V16 = 1*f16

V17 = 1*f17

V18 = 1*f18

V19 = 1*f19

V20 = 1*f20

V21 = 1*f21

V22 = 1*f22

V23 = 1*f23

V24 = 1*f24

V25 = 1*f25

V26 = 1*f26

V27 = 1*f27

V28 = 1*f28

V29 = 1*f29

V30 = 1*f30

Set the Error Variances of V1-V30 to zero

Group HEALTHCARE:

Observed Variables: V1-V30

Covariance Matrix from File HEALTHCARE.COV

Method of Estimation: Maximum Likelihood

Iterations: 20

Sample Size: 212

Latent Variables: f1-f30

Group HUMANITIES:

Observed Variables: V1-V30

Covariance Matrix from File HUMANITIES.COV

Method of Estimation: Maximum Likelihood

Iterations: 20

Sample Size: 239

Latent Variables: f1-f30

End of Problem

Appendix 2: LISREL Syntaxes Example for Test of Measurement Invariance

Model A

Group 1: Testing Equality Of Factor Structures

Model A: Factor Loadings, Factor Correlation, Error Variances Invariant

Observed Variables: V1-V30

Covariance Matrix from File SCIENCE.COV

Asymptotic Covariance Matrix from File SCIENCE.ACM

Method of Estimation: Maximum Likelihood

Iterations: 20

Sample Size: 215

Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5

Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
End of Problem
Model B
Group 1: Testing Equality Of Factor Structures
Model B: Factor Correlation and Error Variances Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
End of Problem
Model C
Group 1: Testing Equality Of Factor Structures
Model C: Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215

```

Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
End of Problem
Model D
Group 1: Testing Equality Of Factor Structures
Model D: Factor Loadings and Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
Set the Error Variances of V1-V30 free
End of Problem

```