ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
2016; 7(2);432-442

The Examination of Model Fit Indexes with Different Estimation Methods under Different Sample Sizes in Confirmatory Factor Analysis*

Doğrulayıcı Faktör Analizinde Farklı Örneklem Büyüklüklerinde Farklı Kestirim Yöntemleriyle Hesaplanan Uyum İndekslerinin İncelenmesi

Ayfer SAYIN**

Abstract
In adjustment studies of scales and in terms of cross validity at scale development, confirmatory factor analysis is conducted. Confirmatory factor analysis, multivariate statistics, is estimated via various parameter estimation methods and utilizes several fit indexes for evaluating the model fit. In this study, model fit indexes utilized in confirmatory factor analysis are examined with different parameter estimation methods under different sample sizes. For the purpose of this study, answers of 60, 100, 250, 500 and 1000 students who attended PISA 2012 program were pulled from the answers to two dimensional "thoughts on the importance of mathematics" dimension. Estimations were based on methods of maximum likelihood (ML), unweighted least squares (ULS) and generalized least squares (GLS). As a result of the study, it was found that model fit indexes were affected by the conditions, however some fit indexes were affected less than others and vice versa. In order to analyze these, some suggestions were made.

Keywords: comfirmatory factor analysis, sample size, fit index, scale development and adjustment

Öz
Ölçme araçlarının uyarlama çalışmalarında ve ölçme aracı geliştirirken çapraz geçerlik çalışmaları kapsamında doğrulayıcı faktör analizi gerceklestirilmektedir. Çok değişkenli bir istatistik olan doğrulayıcı faktör analizinin hesaplanmasında kullanılan birçok parameter kestirim yöntemi ve model uyumunun degerlendirilmesinde kullanılan farklı uyum indeksleri mevcuttur. Bu çalışmada, doğrulayıcı faktör analizinde kullanılan model-veri uyum indekslerinin farklı örneklem büyüklüklerinde farklı parametre kestirim yöntemleri ile hesaplama sonuçları incelenmiştir. Bu doğrultuda PISA 2012 çalışmasına katılan ve iki boyuttan oluşan "matematiğin önemine yönelik görüşler" maddelerine cevap veren öğrenciler içerisinden 60, 100, 250, 500 ve 1000 kişilik veri setleri belirlenmiştir. Hesaplamalarda en çok olabilirlik (EÇO), ağırlıklandırılmamış en küçük kareler (AEKK), genelleştirilmiş en küçük kareler (GEEK) hesaplama yöntemleri kullanılmıştır. Araştırmanın sonucunda oluşturulan koşullardan bazı uyum indekslerinin daha fazla, bazılarının ise daha az etkilendiği belirlenmiştir. Bu doğrultuda analizler için önerilerde bulunulmuştur.

Anahtar Kelimeler: doğrulayıcı faktör analizi, örneklem büyüklüğü, uyum indeksi, ölçek uyarlama ve geliştirme

INTRODUCTION

Observing humans, as they are one of the most basic subjects of measurement, can lead to information about humans, objects, actions, or processes (DeVellis, 2003). However, objects, actions, or behaviors cannot always be observed directly in the studies concerning social sciences. Psychological constructs that are subject to research via indirect methods bring about problems with

* . This study was presented at the V. Measurement and Evaluation in Education and Psychology Conference organized by Akdeniz University in Antalya, Turkey, on September 1–3, 2016.

** Dr., Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme, Ankara-Türkiye, ayfersayin@yahoo.com

Received: 26.09.2016
DOI: 10.21031/epod.280205
Acceptedi: 07.11.2016

the improvement of the measurement devices. For instance, psychological measurements are always based on limited behavior patterns and there is always a chance for faulty measurements. Units that are not defined precisely on the measurement tools and methods create a different problem for measurements. Psychological constructs should be demonstrated with their relative underlying constructs and observed phenomena because they cannot be defined as standalone, thus require observed responses (Crocker & Algina, 1986).

Confirmatory factor analysis (CFA), which in terms of structural equation modeling explores the relationships between latent and observable variables, analyses the measurement model where factor count and relations indicated beforehand. CFA focuses on one variable explaining another variable (or variables) and its error variances from all its sources. Error in measurement varies from each other and unbound to the factors. Relations between factors don’t have to be completely analyzed (factors are assumed to differ as a set) (Kline, 2011).

Since CFA is a multivariate statistics model; normality of the data set, single and multivariate outliers, linearity, multicollinearity and homogeneity of variances should be analyzed before the estimation (Tabachnick & Fidell, 2007). When multivariate normality assumption is not supported, unweighted least squares method can be utilized.

In the estimation process of CFA, firstly the description of the model is presented, then the degrees of freedom in the model is explored by identifying the model. After that, in accordance with the features of the data set, a parameter estimation method is settled upon and used to determine the path coefficients and error variances in the observed-latent variables. Lastly, to evaluate the model-data fit, fit indexes are calculated and necessary adjustments are made according to the modification suggestions.

CFA, which is a special instance of structural equation modeling that all the equations involving multiple regression equations being calculated at the same time, is a multivariate statistics and it depends on certain assumptions. These assumptions are level of scale, outlier, missing value, normality, multiple connections, variable count, linearity, uncorrelated error term and sample size (Schumacker & Lomax, 2004; Tabachnick & Fidell, 2007; Byrne, 2010). Kaplan (2001) suggests that basic assumptions of structural equation modeling are sufficient sample size, mutlivariate normality, missing data and proper building of the model. Fan, Thompson & Wang (1999) focuced in their studies that effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Researchers state that although there are so many fit indexes which were developed under different theoretical rationales, it is not known which fit indexes are the ideal. Therefore, it is necessary to investigate the variables in which fit indixes are sensitive (Fan and others, 1999). Fan and Sivo (2007) examined the sensitivity of model fit indices to different types of models while controlling for the severity of model misspecification. At each level of misspecification, the severity of misspecification is comparable across the three different models. Ximénez (2009) first examined recovery for models correctly specified with the known number of factors, and then investigated recovery for models incorrectly specified by underfactoring. Savalei (2012) was studied with RMSEA under different types of misspecification.

Within the scope of this research, two models consisting of two dimensions were created. The difference between the first model and the second model is the items of the dimensions. Within the scope of research, three questions are answered:

- What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 1?
- What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 2?
- How do fit indexes differ within the Model 1 and Model 2, under different sample sizes and different parameter estimation methods?

Purpose of the study

CFA, which is used commonly to test a theoretical construct, used mainly in scale development, scale adjustment, cultural comparison, and comparing groups. CFA should be properly managed in areas like development and adaptation of scales used in the education field. When the literature in the field is examined, it can be seen that confirmatory factor analysis along with structural equation modeling is used frequently. However, in those studies sample size assumption is not properly utilized or not determined in accordance with assumptions of parameter estimation methods. This study aims to investigate the model fit indexes with different parameter estimation methods under different sample size conditions.

METHODOLOGY

Research model

In this research, the effect of varying parameter estimation methods and different sample sizes on fit indexes is meant to be explored. At the same time the effect of models set on the model-data fit indexes are meant to be determined. From this point of view, the research can be classified as a "basis research".

Research Group

The research group consists of students who attended the PISA 2012 in Turkey. In PISA, 4848 students answered the questions in the "views on the importance of mathematics" dimension. The missing or invalid and n/a data (n=1693) are removed from the data set and also analyses of the assumptions were done. From the remaining student answers (n=3155), different clusters of 60, 100, 250, 500 and 1000 answers were pulled randomly and processed within the research.

Data Collection Tool

In PISA 2012, six items in the mathematics dimension of the research were scored in a four-point likert scale. First three of these items ask about their friends' view on mathematics, and the remaining three asks about their families' views on mathematics. Exploratory factor analysis was used with the answers of 3155 students who responded to the survey items in Turkey. At the end of the calculation, it was concluded that the dimension regarding the views of the friends of the students accounted for %37.193 of the total variance, and the family dimension accounted for %32,740 of the total variance. The total variance of the six items was calculated as %69,933. Cronbach's alpha of the first dimension was calculated as 0,824; reliability coefficient of the second dimension was calculated as 0,705. The items can be seen in Table 1 below.

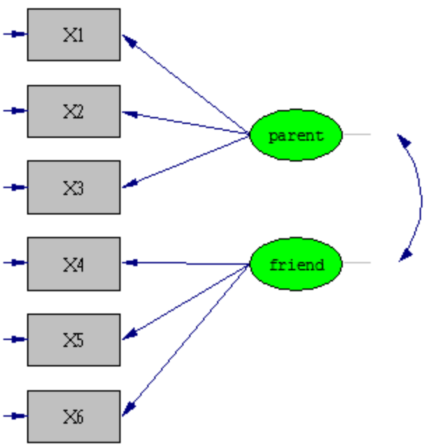
Table 1. Items and Dimension

Dimension	Items
Friends	X1. Most of my friends do well in mathematics.
	X2. Most of my friends work hard on mathematics.
	X3. My friends enjoy taking mathematics tests.
Parents	X4. My parents believe it's important for me to study mathematics.
	X5. My parents believe that mathematics is important for my career.
	X6. My parents like mathematics.

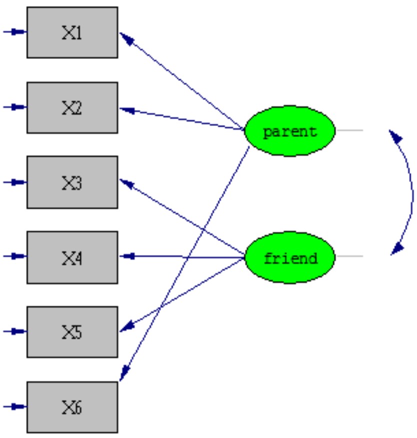
Analysis of The Data

In the research, analyses of assumptions were carried out before analyses of the data sets. Firstly, the students with missing answers to the items were removed from the data set. Following that, univariate (determined with z statistics) and multivariate (calculated with Mahalonobis coefficient) observations with extreme values were removed. In order to examine the normality assumption, coefficient of skewness and kurtosis were utilized. It was calculated that the answers to the items were between -1 and +1 in the given values. In addition, histogram graphics were examined and no extreme deviation was detected in the values. In order to check if there is a multicollinearity, correlation coefficients of the answers to the items were calculated and found to be between 0,074 and 0,709. Box's M test was carried out and calculated to check the homogeneity of variances (p>0,05). The aforementioned assumptions were analyzed in the groups of 60, 100, 250, 500, 1000 sample sizes that were pulled from the whole data set. In all data sets, it was found that there is no significant differences between items correlation coefficients (p>0,05).

In the analysis of the data, Model 1 which was reached via calculations with different sample sizes using exploratory factor analysis was calculated first. Then the factors with the least factor loads were determined and replaced to reach to Model 2, which the calculations were based upon. Estimations were carried out via LISREL packaged program.



Model 1



Model 2

Model 1: The model is specified in the first stage of DFA. Model 1 is specified in terms of AFA results and survey assumption.
Model 2: Factors with the least factor loads (X3-X6) were determined and replaced to reach to Model 2.

FINDINGS

1. What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 1?

Analyses for Model 1 which were based on maximum likelihood, unweighted least squares and generalized least squares estimation methods were carried out with sample sizes of 60, 100, 250, 500, and 1000. The fit indexes based on those estimations and their results are shown in Table 2.

When Table 2 is examined, it can be seen that X^2/df fit index tend to increase with sample size for all three estimation methods. RMSEA fit index tend to decrease as the sample size increases to 250 and to increase with a sample size of 500. CFI, which is a goodness of fit index, tend to increase as the sample size increases to 250 and it shows a stable value (similar in 500 and 1000 sample size) with increasing sample size. NFI and GFI show a similar pattern as CFI.

It is observed that overall model-data fit is achieved when the sample size is at least 250 in ML estimation method, 60 in ULS estimation method, and 250 in GLS estimation method.

Table 2. Estimation Results for the Model 1

Estimation methods	Sample Size	X^2/df	RMSEA	CFI	NFI	GFI
ML	60	2,78	0,174	0,88	0,83	0,89
	100	2,87	0,138	0,93	0,91	0,93
	250	3,27	0,095	0,97	0,96	0,97
	500	**6,76**	**0,107**	**0,97**	**0,96**	**0,97**
	1000	12,20	0,106	0,97	0,96	0,97
ULS	60	1,34	0,088	0,97	0,92	0,96
	100	1,46	0,058	0,99	0,96	0,99
	250	1,49	0,045	0,99	0,98	0,99
	500	3,01	0,063	0,99	0,98	0,99
	1000	7,06	0,075	0,98	0,98	0,98
GLS	60	1,53	0,111	0,87	0,77	0,81
	100	1,72	0,073	0,94	0,87	0,88
	250	1,60	0,049	0,97	0,94	0,95
	500	3,09	0,065	0,96	0,94	0,95
	1000	6,16	0,072	0,95	0,94	0,95

2. What are the fit indexes that were estimated via different parameter methods with different sample sizes in Model 2?

Analyses for Model 2 were carried out with the same three parameter estimation methods under five sample sizes. Results are shown in Table 3.

Table 3. Estimation Results for the Model 2

Estimation methods	Sample size	X^2/df	RMSEA	CFI	NFI	GFI	Statistically insignificant article count
ML	60			**Model does not converge**			
	100			**Model does not converge**			
	250	269,29	0,345	0,51	0,51	0,74	
	500	564,80	0,362	0,52	0,52	0,73	
	1000	959,99	0,373	0,54	0,54	0,76	
ULS	60	14,13	0,114	0,97	0,94	0,95	
	100	63,93	0,266	0,77	0,75	0,97	2
	250	183,71	0,297	0,72	0,71	0,88	
	500	435,92	0,327	0,70	0,70	0,87	
	1000	316,27	0,278	0,77	0,76	0,89	
GLS	60	15,39	0,125	0,85	0,77	0,66	
	100	164,78	0,441	0,00	-0,81	0,27	3
	250	301,43	0,445	0,00	-0,49	0,40	
	500	784,54	0,384	0,00	-0,82	0,33	
	1000	565,00	0,374	0,00	-0,42	0,41	

When Table 3 is examined, it can be seen that model did not converge with 60 and 100 sample sizes in ML method. In ULS method, with 100 sample size (2 items) and in GLS method, with 100 sample size (3 items), some items did not produce statistically significant results.

When the information on Table 3 is analyzed, it can be seen that generalized least squares and maximum likelihood method are affected the most from Model 2, and unweighted least squares method was the least affected among the three. At the same time, it is observed that ULS method overall manages to achieve data model fit in 60 sample size (relative to bigger fit indexes).

According to the information on Table 3, fit indexes of X^2/df showed increase with the sample size and a model data fit was not achieved in any sample size. RMSEA fit index similarly increased with the sample size in ML overall. In ULS and GLS estimation methods,, RMSEA showed a decline in 250 and 500 sample sizes.

In ML parameter estimation method, CFI and NFI fit indexes produced similar results in every sample size, whereas GFI fit index showed a relatively bigger estimation among others. In ULS method, apart from 60 sample size, a similar change is observed. In GLS parameter estimation method, GFI fit index produced higher estimations relative to other fit indexes.

3. How does fit indexes differ within Model 1 and Model 2, different sample sizes and different parameter estimation methods?

Fit indexes that reduced in value after the estimations of Model 1 and Model 2 and differences between them are analyzed. The differences between the reduced fit indexes are shown in Figure 1.

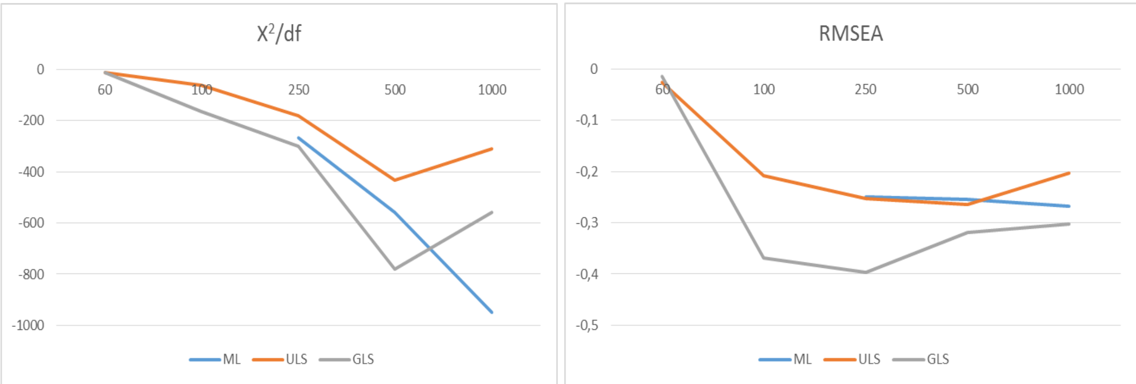


Figure 1. The Difference Among the Decremental Fit Indexes

As seen in Figure 1, X^2/df fit index showed the most increase in GLS parameter estimation method, and the least in ULS parameter estimation method. The biggest difference between the Model 1 and Model 2 model was estimated with ML parameter estimation method in 1000 sample size. RMSEA fit index is observed to not change in ULS and GLS in 60 sample size. ULS parameter estimation method is found to be the least affected from the Model 2 setting of the model.

As seen in Figure 2, the least affected parameter estimation method from the Model 2 setting of the model is ULS; whereas the most affected is GLS. In 60 sample size, the incremental fit indexes of CFI, NFI and GFI show little or no sign of being affected by Model 1 and Model 2 setting of the model. GFI fit index is found to be the least affected by the setting of the model among other incremental fit indexes.

The fit indexes showing increase with the Model 1 and Model 2 model are shown in Figure 2 below.

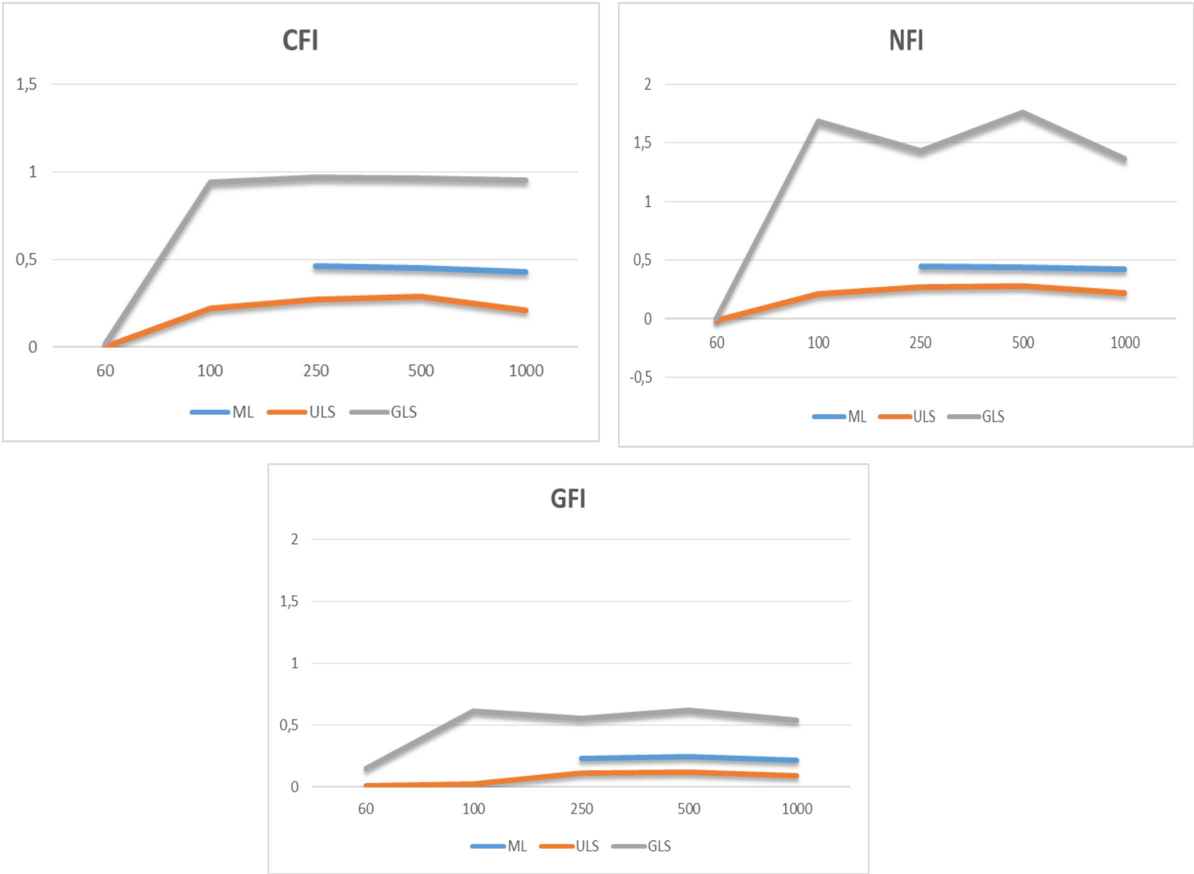


Figure 2. The Difference Among the Incremental Fit Indexes

RESULTS and DISCUSSION

In this study, the aim was to study the effect of different parameter estimation methods and sample sizes on the fit indexes for two models in confirmatory factor analysis. Towards this, sample data sets of 60, 100, 250, 500 and 1000 were pulled from answers to the "thoughts on the importance of mathematics" section in PISA 2012 data. Six items in the mathematics section of the research were scored in a four-point likert scale . Data sets were firstly analyzed in exploratory factor analysis and it was found that scale factors were two dimensionally gathered in friends (%37,193) and family (%32,740). Cronbach's alpha reliability coefficients were estimated to be 0,824 and 0,705 in dimension 1 and dimension 2 of the scale, respectively.

Firstly, Model 1 was formed with six items and two dimensions, and ML, ULS and GLS parameter estimation methods were carried out with five different sample sizes. Results of the estimations showed that X^2/df fit index shows increase with the sample size. RMSEA fit index, on the other hand, shows increase with sample size up to a point, then decreases with some parameter estimation methods. Those which were referred to as decremental fit indexes showed a tendency to rise with the sample size, meaning that they may be misleading with bigger sample sizes. Research shows that X^2/df fit index can be statistically underpowered in small sample sizes (Kenny & McCoach, 2003). On the other hand, Kline (2011) suggests X^2 fit index shows increase along with the sample size. Sayın & Gelbal (2016) also states in their study that fit index shows increase with increasing sample size, that model data fit cannot be achieved in especially greater sample sizes. On the other hand X^2 fit index may be quite affected from the Model 2 model. Kenny, Kaniskan & McCoach (2014) states that RMSEA, which is the most used to evaluate model data fits, fit index is not fit to be used in small df models. They state that in especially small sample sizes, model parameters are estimated lower than they should be with RMSEA.

The fit indexes that showed an increase, CFI, GFI and NFI fit indexes produced and increase until 250 sample size, and remained stable after that point, overall. Accordingly, it can be said that 250 sample size is enough to measure the model data fit. Boomsma (1982) states that at least a sample size of 200 is needed for structural equation modeling. Molwus, Erdogan, & Ogunlana (2013) suggests, when the effect of sample size on fit indexes are explored, small sample sizes produce low fit indexes, and as the sample sizes get bigger, fit indexes also do. The also suggest that is vital to consider this aspect when studying with small and big sample sizes at the same time.

When parameter estimation methods are observed, ML estimation method achieved model data fit at 250 sample size, ULS at 60 sample size, and GLS at 250 sample size. Fan and others (1999), state that ML and GLS should not produce different statistics from each other in theory, but there can be variations in practice against prediction. Sayın & Gelbal (2016) states that ML estimation method produces higher model fit index estimations than GLS. In this case if the prerequisite assumptions are met, It is suggested that ML is used.

In the research, in order to determine the effect of modeling on fit indexes, the items with the least factor loads were replaced and a Model 2 model was created. Same analysis procedures were carried out for Model 2. In the Model 2, one of the decremental fit indexes, X^2/df fit index showed an increase with the sample size and it showed no model data fit in any sample size. RMSEA fit index showed an overall increase with ML estimation method, along with the sample size. In ULS and GLS estimation methods, at 250 and 500 sample sizes, RMSEA showed a decline. In light of those informations, it can be said that X^2/df and RMSEA decremental fit indexes can be utilized if the model is set as Model 2 but they should not be evaluated for their results in different sample sizes. Fan and others (1999) states in their studies that RMSEA is more sensitive than GFI to model misspecification.

In ML parameter estimation method, CFI and NFI fit indexes produced similar results in all sample sizes, whereas GFI fit index was estimated to be higher than other incremental fit indexes. In ULS method, a similar pattern was observed over 60 sample size. In this case, it can be said that GFI fit index is the least affected from the Model 2 setting of the model, and whenever the model is inconsistent, it should not be used. Fan (1996) states in his study that a completely wrong model setting affects GFI fit index more than CFI and NFI. Schermelleh-Engel, Moosbrugger, & Müller (2003) also suggest in their study that GFI fit index estimates are higher than other incremental fit indexes, and it is not affected by a Model 1 or Model 2 setting of the model. Sharma, Mukherjee, Kumar & Dillon (2005) states in their studies that despite the percent of times misspecified model is accepted is quite low for higher degrees of misspecification, RMSEA index performs better than GFI. Also they noticed that the use of GFI should be discouraged.

GLS and ML methods are found to be the most affected from the Model 2 setting of the model, whereas ULS method was affected the least. At the same time ULS method achieves model data fit at 60 sample size (relative to the incremental fit indexes). In light of this information, it is suggested that when the model data fit is achieved preliminarily, ULS should not be used. Fan and others (1999) state that parameter estimation methods affect all fit indexes strongly. And they have found similar results to these research results that fit index values based on GLS invariably appeared to indicate a better fit than those based on ML. And also the research can be repeated with the simulation study.

REFERENCES

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. (Eds. K. G. Jöreskog & H. Wold). Systems under indirect observations: Causality, structure, prediction, Amsterdam: North-Holland.

Byrne, B. M. (2010). Structural equation modeling with AMOS basic concepts, applications, and programming. New York: Taylor & Francis Group.

Crocker, L. M. & Algina, L. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

DeVellis, R. F. (2003). Scale development theory and applications, applied social research methods series. California: SAGE Publications.
Fan, X. (1996). Structural equation modeling and canonical correlation analysis: What do they have in common?. Structural Equation Modeling: A Multidisciplinary Journal, 4, 64-78.
Fan, X., Thompson B. & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Structural Equation Modeling, 6(1), 56-83.
Fan, X. and Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. Multivariate Behavioral Research, 42(3), 509-529.
Kaplan (2001), Structural equation modeling. International Encyclopedia of the Social & Behavioral Sciences, 15215-15222.
Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. Structural Equation Modeling, 10, 333–351.
Kenny, D. A., Kaniskan, B. & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. Sociological Methods & Research, 1-22.
Kline, R. B. (2011). Principals and practice of structural equation modeling. New York. The Guilford Press.
Molwus, J. J., Erdogan, B. & Ogunlana, S. O. (2013). Sample Size and Model Fit Indices for Structural Equation Modelling (SEM): The Case of Construction Management Research. ICCREM, 338-347.
Savalie, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. Educational and Psychological Measurement, 72(6), 910-932.
Sayın, A. & Gelbal, S. (2016). Yapısal eşitlik modellemesinde parametrelerin klasik test kuramı ve madde tepki kuramına göre sınırlandırılmasının uyum indekslerine etkisi. International Journal of Education Science and Technology, 2(2), 57-71.
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.
Schumacker, R.E. & Lomax, R.G. (2004). A beginner's guide to structural equation modeling. Lawrence Erlbaum Associates, New Jersey.
Sharma, S., Mukherjee, S., Kumar, A. & Dillon, W.R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. Journal of Business Research, 58, 935–943.
Tabachnick, B. G. & Fidell, L. S. (2007). Using multivariate statistics. Boston: Allyn and Bacon.

GENİŞ ÖZET

Giriş

Ölçmenin temel konularından birini oluşturan insanlar gözlemlenerek; nesneler, olaylar ve süreçler hakkında bilgi edinilebilir (DeVellis, 2003). Ölçülmesi gereken psikolojik yapılar gözlenen değişkenler aracılığı ile belirlenebilir. Ancak daha çok sosyal bilimlerde doğrudan gözlemler yoluyla gerekli bilgiler sağlanamayabilir ve bir hata ortaya çıkabilir. (Crocker, & Algina, 2006).

Gözlenen cevaplar doğrultusunda gizil yapıların belirlendiği ve daha çok teorik bir yapının test edildiği Doğrulayıcı Faktör Analizi (DFA); ölçek uyarlama, geliştirme, kültürel karşılaştırma, grupları karşılaştırma gibi alanlarda sıklıkla kullanılmaktadır. Eğitimde kullanılan ölçeklerin geliştirme ve uyarlama sürecinde DFA'nın doğru bir şekilde kullanılması gerekmektedir. Alan yazında gerçekleştirilen araştırmalar incelendiğinde, yapısal eşitlik modellemesi ile doğrulayıcı faktör analizi çalışmalarının sıklıkla kullanıldığı ancak bu çalışmalarda örneklem büyüklüğü varsayımının incelenmediği, parametre kestirim yöntemlerinin varsayımları doğrultusunda belirlenmediği görülmektedir. Bu çalışma ile örneklem büyüklüğü ile parametre kestirim yöntemlerinin uyum iyiliği indeksleri üzerindeki ne düzeyde bir etkisi olduğu belirlenmesi amaçlanmaktadır. Bunun yanı sıra DFA'nın kullanıldığı çalışmalarda, özellikle geliştirme çalışmalarında, modelden tam olarak emin olunamadığı durumlarla karşılaşılmaktadır. Bu doğrultuda araştırmada aynı zamanda oluşturulan iki farklı modelin hesaplama sonuçları karşılaştırılmıştır. Sosyal bilimler ve davranış bilimlerinde doğrulayıcı faktör analizi sıklıkla kullanılmasına rağmen böyle bir çalışmanın yaygınlaşmaması araştırmayı ayrıca önemli kılmaktadır.

Yöntem

Bu araştırmada, farklı parametre kestirim yöntemleri ve farklı örneklem büyüklüklerinin doğrulayıcı faktör analizindeki uyum indeksleri üzerindeki etkisinin araştırılması amaçlanmaktadır. Aynı zamanda model-veri uyum indekslerinin oluşturulan farklı modellerde nasıl hesaplandığının belirlenmesi amaçlanmaktadır. Bu doğrultuda araştırmanın "temel araştırma" niteliğinde olduğu söylenebilir.

Araştırmanın çalışma grubunu Türkiye'de PISA 2012 çalışmasına katılan öğrenciler oluşturmaktadır. PISA çalışmasında "matematiğin önemine yönelik görüşler" maddelerine 4848 öğrenci cevap vermiştir. Eksik ve geçersiz veriler, veri setinden çıkarılmış ve varsayımların incelemesi gerçekleştirilmiştir. Kalan 3155 öğrencinin cevapları içerisinden rastgele seçilen 60, 100, 250, 500 ve 1000 öğrencinin görüşleri araştırma kapsamında incelenmiştir.

PISA 2012 çalışmasında öğrencilerin cevaplandırdığı matematik anketinde kişilerin matematiğin önemine yönelik 4'lü likert tipinde derecelendirilmiş altı madde yer almaktadır. Bu maddelerden ilk üçü arkadaşlarının, diğer üç madde de ailesinin matematiğe yönelik algılarını ifade etmektedir. Türkiye'de çalışmaya katılan öğrencilerin cevapları (n=3155) doğrultusunda öncelikle açımlayıcı faktör analizi hesaplanmıştır. Hesaplama sonucunda arkadaş boyutundaki maddelerin toplam varyansın %37,193'üne; aile boyutundaki maddelerin de varyansın %32,740'ına açıklık getirdiği tespit edilmiştir. Toplam altı maddenin varyansın %69,933'ünü açıkladığı belirlenmiştir. Birinci boyuttaki maddelerin Cronbach alfa katsayısı 0,824; ikinci boyuttaki maddelerin güvenirlik katsayısı da 0,705 olarak hesaplanmıştır.

Bu doğrultuda PISA 2012 çalışmasına katılan ve iki boyuttan oluşan "matematiğin önemine yönelik görüşler" maddelerine cevap veren öğrenciler içerisinden 60, 100, 250, 500 ve 1000 kişilik veri setleri belirlenmiştir. Belirtilen şekilde iki boyutlu oluşturulan Model 1 ile en çok olabilirlik (EÇO), ağırlıklandırılmamış en küçük kareler (AEKK), genelleştirilmiş en küçük kareler (GEKK) hesaplama yöntemleri kullanılarak hesaplamalar gerçekleştirilmiştir. Ardından boyutlar içerisinde en düşük faktör yük değerine sahip maddelerin yerleri değiştirilerek Model 2 oluşturulmuştur. Model 2 ile de hesaplamalar gerçekleştirilerek sonuçlar karşılaştırılmıştır.

Sonuç ve Tartışma

Araştırma kapsamında öncelikle altı madde ve iki boyuttan oluşan Model 1 kurulmuş ve EÇO, AEKK ve GEKK parametre kestirim yöntemleri ile beş farklı örneklem büyüklüğünde hesaplamalar gerçekleştirilmiştir. Hesaplama sonucunda X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği belirlenmiştir. RMSEA uyum indeksinin ise parametre kestirim yöntemlerine göre belirli örneklem büyüklüğüne göre azalma, daha sonra ise artma eğiliminde olduğu tespit edilmiştir. Söz konusu uyum indekslerinin örneklem büyüklüğüne bağlı olarak artış göstermesi, büyük örneklem büyüklüklerinde yanıltıcı olabileceklerini göstermektedir.

Artan uyum indeksleri olan CFI, GFI ve NFI uyum indekslerinin tüm parametre kestirim yöntemlerinde genel olarak 250 örneklem büyüklüğüne kadar artış gösterdiği, daha sonra ise genel olarak sabitlendiği belirlenmiştir. Bu doğrultuda 250 örneklem büyüklüğünün model-veri uyum tespitinde yeterli olduğu söylenebilir.

Parametre kestirim yöntemleri bazında incelemeler gerçekleştirildiğinde ise EÇO hesaplama yöntemi ile en az 250 örneklem büyüklüğünde, AEKK ile 60 örneklem büyüklüğünde ve GEKK hesaplama yöntemi ile de 250 örneklem büyüklüğünde model-veri uyumunun genel olarak sağlandığı tespit edilmiştir.

Model içerisinde maddelerin boyutlarındaki yerleri değiştirilerek Model 2 oluşturulmuştur. Model 2 ile gerçekleştirilen hesaplamalar sonucunda X^2/sd ve RMSEA azalan uyum indekslerinin değişim gösterdiği belirlenmiştir. Model 2 hesaplanmasında en çok GEKK yöntemi ile EÇO yönteminin, en az ise AEKK'nin Model 1'den farklı sonuçlar verdiği belirlenmiştir.

Araştırma kapsamında gerçekleştirilen hesaplamalar sonucunda X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği belirlenmiştir. RMSEA uyum indeksinin ise parametre kestirim yöntemlerine göre belirli örneklem büyüklüğüne göre azalma, daha sonra ise artma eğiliminde olduğu saptanmıştır. Araştırmalar X^2/sd uyum indeksinin küçük örneklem büyüklüklerinde istatistiksel gücünün düşük olacağını ifade etmektedir (Kenny & McCoach, 2003). Buna karşın Kline (2011) X^2 uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiğini ifade etmektedir. Sayın (2014) da çalışmasında X^2/sd uyum indeksinin örneklem büyüklüğüne bağlı olarak artış gösterdiği, özellikle de geniş örneklem büyüklüklerinde iyi uyum gösteren modelde model-veri uyumunun sağlanamadığına yönelik bilgiler verdiğini belirlemiştir. Kenny, Kaniskan & McCoach (2014) model-veri uyum değerlendirmesinde en çok kullanılan RMSEA uyum indeksinin serbestlik derecesinin küçük olduğu modellerinde kullanılmasının uygun olmadığını ifade etmektedir.

Artan uyum indeksleri olan CFI, GFI ve NFI uyum indekslerinin tüm parametre kestirim yöntemlerinde genel olarak 250 örneklem büyüklüğüne kadar artış gösterdiği, daha sonra ise genel olarak sabitlendiği belirlenmiştir. Bu doğrultuda 250 örneklem büyüklüğünün model-veri uyum tespitinde yeterli olduğu söylenebilir. Bu çalışma kapsamında 250 örneklem büyüklüğü aynı zamanda serbestlik derecesinin katını da ifade etmektedir. Bu doğrultuda hesaplamalarda serbestlik derecesine bağlı bir örneklem büyüklüğünün de belirlenebileceği söylenebilir. Boomsma (1982) yapısal eşitlik modellemesinde en az 200 örneklem büyüklüğüne ihtiyaç duyulduğunu belirtmektedir. Molwus, Erdogan, & Ogunlana (2013) örneklem büyüklüğünün uyum indekslerine etkisini incelediği çalışmasında küçük örneklem büyüklüklerinde uyum indekslerinin düşük, geniş örneklem büyüklerinde ise yüksek çıktığını belirlemiştir.

EÇO parametre kestirim yönteminde CFI ve NFI uyum indeksinin tüm örneklem büyüklüklerinde benzer sonuçlar ürettiği, GFI uyum indeksinin ise diğer artan uyum indekslerinden daha yüksek kestirildiği belirlenmiştir. Bu durumda modelin farklı tanımlanmasından GFI uyum indeksinin diğer indekslere göre daha az etkilendiği; modelden emin olunmadığı durumlarda söz konusu uyum indeksinin raporlanmaması gerektiği önerilmektedir. Fan, Thompson, & Wang (1996) ise yaptığı çalışmada ise modelin tam olarak hatalı kurulmasında GFI uyum indeksinin CFI ve NFI uyum indeksinden daha duyarlı olduğunu belirtmiştir. Schermelleh-Engel, Moosbrugger, & Müller (2003) çalışmalarında bu çalışma sonuçlarına benzer şekilde GFI uyum indeksinin diğer artan uyum indekslerine göre daha yüksek kestirimler gerçekleştirdiği; farklı tanımlanan modellerdeki hesaplama sonucunun çok fazla değişiklik göstermediği belirlenmiştir. Benzer şekilde Sharma, & diğerleri (2005) de GFI uyum indeksinin kullanılmasının uygun olmadığını belirtmektedir.