

Exploratory and Confirmatory Factor Analysis: Which One to Use First?*

Fatih ORÇAN**

Abstract

There exist differences between the use of Exploratory and Confirmatory Factor analysis at scale adaptation or development studies. The order of factor analysis used would cause the discrepancy in the results. Besides, multiple confirmatory factor analysis would fit well on a single data set. In this study simulated data sets were fitted to three different models. Based on the results 64% of the data sets fit well on all three models. Also, a different data set was fit both on a confirmatory and an exploratory factor analysis. The result showed that confirmatory factor analyses were not sufficient to detect the best fitting model.

Keywords: Scale adaptation, scale development, confirmatory factor analysis, exploratory factor analysis

INTRODUCTION

Confirmatory Factor Analysis (CFA) and Exploratory Factor Analysis (EFA) are two common techniques used in scale development and scale adaptation studies. If the relationship among the items is not known it is recommended to use EFA, but if the relationship is tested and the factors and related items are known, CFA is recommended to be used (Bandalos & Finney, 2010; Büyüköztürk, 2002; Kline, 2011). For scale adaptation studies the use of these methods and their orders of use showed diversity from one study to another. Güvendir and Özkan (2015) explored scale adaptation and development studies published in Turkey between 2006 and 2014. Based on their results, total of 25 studies used EFA out of 26 scale development studies and 16 of them used CFA. Moreover, 22 scale development studies started with EFA to analyze their data while 11 started with CFA.

Experts may guess how items will be structured beforehand; however, a statistical technique is required to decide about the structure of the items and number of latent factors. Thus, the items which works (explains variation) could be determined easily. Therefore, for a scale development study first an EFA should be used in order to discover underlying latent structure (Brown, 2006; Schumacker & Lomax, 2010). In fact, 96% of the studies in Turkey used EFA (Güvendir & Özkan, 2015). Besides, in a scale development process, CFA should be run using a data set different from the EFA data set (Schumacker & Lomax, 2010). Thus, the validity of the EFA structure found as a result of EFA will be shown by using CFA with a different data set. Two different ways can be discussed in the creation of the data set to be used for factor analysis. First, after a sufficient number of samples are collected in a single run to make both EFA and CFA, some of them (eg 50%) can be randomly selected for EFA and the rest for CFA. Another way is to collect two different data sets and analyze one for EFA and the other for CFA.

In adaptation studies, the use of EFA and CFA varies. For example, the process of translating the items from the original language to a new language is an important step for scale adaptation studies. Failure of transferring the original item meanings may cause a variation called scale error in scale scores. As a result of this meaning shift, it is possible to create a structure different from the original scale structure. Therefore, in an adaptation study, it is necessary to make sure that the translation of the item is done correctly before starting the analysis. A coherent translation process is very important for the elimination of structural differences. Sousa and Rojjanasrirat (2011) defined a step by step process for

* A part of this paper was presented at VII. International Congress of Research in Education.

** Dr. Öğr. Üyesi, Trabzon Üniversitesi, Fatih Eğitim Fakültesi, Trabzon-Türkiye, e-posta: fatihorcan@ktu.edu.tr, ORCID ID: <https://orcid.org/0000-0003-17270456>

To cite this article:

Orçan, F., (2018). Exploratory and confirmatory factor analysis: which one to use first?. *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 414-421. DOI: 10.21031/epod.394323

Received: 13.02.2018

Accepted: 28.09.2018

translating a scale into another language. According to Sousa and Rojjanasrirat (2011), at least two people should translate the article first (forward translation). Then the work of these two independent translators should be reviewed by a third expert and the translation should be finalized. In the third step, the translated materials should be translated back into the original language at least by two different experts (back translation) and the final version of the scale should be obtained after these translations are examined by a third independent expert. In subsequent steps, the pilot and actual implementation stages are presented in detail (Sousa & Rojjanasrirat, 2011). A similar process was made by Sperber (2004). Sperber (2004) also stressed that word to word translation may not be accurate and the items translation should be culturally adaptive in order to prevent meaning shifts.

It is obvious that the translation error will affect the validity and reliability of the adapted scale. Therefore, in this study psychometric tests, as mentioned by Sousa and Rojjanasrirat's (2011) the seventh step of the process, used to test the validity and reliability (EFA and CFA) were considered. In this study the reasons and the order of use of these techniques were examined via simulated data in order to explore the possible differences in the results.

EFA is a statistical technique used in the social sciences for determining underlying latent variables (factors). In other words, EFA stands out as a technique used in scale development. It is used where there is no knowledge among the items of the scale, that is, how many factors there are between the items and which factors are determined by which items. As the name suggests, EFA helps *explain* the structure that exists (Hayton, Allen, & Scarpello, 2004; Hurley, Scandura, Schriesheim, Brannick, Seers, et al., 1997). Some critical decisions need to be made during EFA, such as which method of estimation will be used, whether rotation will be made or by which criteria the number of factors will be determined. There are many studies in the literature about these (Costello & Osborne, 2005; Hanson & Roberts, 2006; Schmitt, 2011). Therefore, these factors were kept constant in the study. More detailed information about these concepts (transformation, sample size or number of factors) can be found in Büyüköztürk (2002), Costello and Osborne (2005).

Unlike EFA, CFA is used when there is a strong model assumption. With CFA, the existence of a previously proven structure is investigated with a new data set. In scale development studies, CFA should be used to test the validity of the structure obtained after EFA (Worthington & Whittaker, 2006). However, the use of CFA in scale adaptation studies differs in practice. In some adaptation studies, it is seen that both EFA and CFA are used, while in others only CFA is used. Use of CFA only in adaptation studies may cause some problems. For example, if a translation error occurred in an adaptation study, using the CFA only might result in a different situation than would actually occur, and the model could be misleading. In addition, a data set may fit with more than one CFA model, so it would be more appropriate to conduct an EFA first to introduce possible cultural differences in the adaptation. In such a case, if an EFA is not performed, a researcher will not test a second model since the first tested model fit to the data. Thus, it is important to run an EFA first to recognize the possible error.

Purpose of Study

The main purpose of this study was to determine how a data set can fit to more than one CFA model, and also how the use of EFA or CFA first may differ in the outcome model. In this study, the models were compared in two respects. Firstly, data were generated according to the model shown in Figure 1 in the R-cran program and these data were tested according to three different CFA models. Second, a simulated data set was evaluated on the basis of item and the possible differences that could occur as a result of using EFA or CFA first (as an example of scale adaptation or development procedures) were revealed. Thus, it is aimed to show what the different scale development procedures can produce.

METHOD

For the simulation part of the study, 100 data sets were simulated via R-cran for the sample size of 300 with model shown at Figure 1. Since the sample size is not a design factor in this study, as indicated

in Orçan and Yang (2016), 300 samples will be sufficient for this study. However, as the model complexity increases, the need for sample size will increase. The model consisted of two factors and eight observed variables. For four items loaded on each factor, factor loadings were set at .40, .50, .60 and .70 to provide diversity. In addition, the correlation between factors was determined as .70. This model was preferred in order not to increase the model complexity. The aim of the study is not to compare the behavior of CFA and EFA under different conditions. But it is aimed to show that the same data set can fit more than one model. Therefore, it would be sufficient to examine a single case to show that a data set might fit well more than one CFA model.

To produce data according to the model, first, the factor scores were randomly generated with the mean of 0 and the standard deviation of 1. Then, Cholesky method has been applied to ensure that the item data is multinomial in accordance with the specified factor loadings. Observed values were obtained as a linear combination of factor scores (Orçan & Yang, 2016). Finally, the simulated continuous variables converted into five categories in order to reflect the five-point Likert item properties. For the second aim of the study, a data set with sample size of 300 was tested with EFA and CFA models.

Analysis

Mplus 5.1 (Muthén & Muthén, 1998-2008) was used to analyze the data. 100 data sets were categorically generated based on the model (Model 1) as in Figure 1 that is correctly specified model. The data sets were analyzed according to model 1 and two misspecified models; Model 2 where item 5 is loaded on the first factor and Model 3 where item 4 is loaded on the second factor (See figure 2). These models are shown in figure 2. Factor loadings of items 4 and 5 are lower than others. In case of high factor loadings, the item-factor correlation will be higher and misspecification will be more prominent. However, for the purpose of the study lower factor loadings will be sufficient. In another study, factor loadings can be taken as research design.

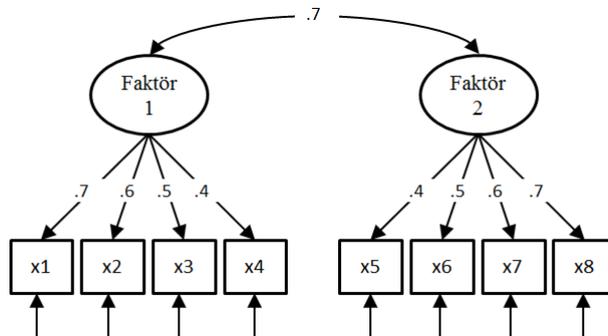


Figure 1. Data Generation Model (Model 1)

Although the data were generated according to normal distribution, during the categorization process the data were distanced from normality. Therefore Maximum Likelihood (MLR) estimation method was used for the models. For each of the models the p-value of the chi-square test, the comparative fit indices (CFI), the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) values were compared with Hu and Bentler's (1999) criteria. Besides, the correlation between the factors was examined via the descriptive statistics and the root mean square error (RMSE).

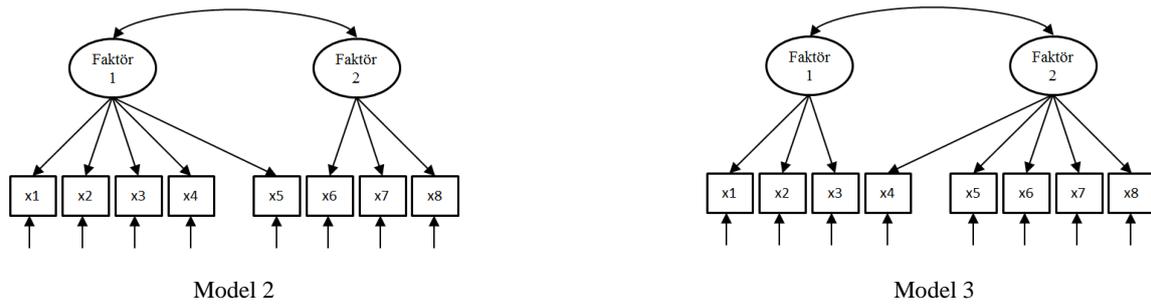


Figure 2. Misspecified Models

RESULTS

First, the models were evaluated based on chi-square, CFI, RMSEA and SRMR values. The number of data where model fit indices indicated good fit was shown at table 1. For example, 87% of the data fitted well for model 1 in terms of chi-square. However, 64% of the same data sets fit models 2 and 3. Similarly, for the RMSEA value, all the data fit to model 1 (100%), whereas for model 2 and model 3 these values were 91% and 89% respectively. Therefore, misspecified models model 2 and model 3 were considered to be true at 91% and 89%, respectively.

Table 1. Model-Data Fit Counts

	Model-Data Fit: Yes		
	Model 1	Model 2	Model 3
Ki-Kare	87	64	64
CFI	98	80	70
RMSEA	100	91	89
SRMR	100	100	100

The CFA models are generally evaluated based on the four fit indices. That is, a CFA model said to show a good model-data fit if the p-value of chi-square test is higher than .05, the CFI is higher than .95, the RMSEA and SRMR values are less than .06 and .08, respectively (Hu & Bentler, 1999). Table 2 shows the number of fit indices which indicates good model-data fit. For example, under model 1, 87% of the data-sets indicated good fit for four indices at the same time. This value is 64% for model 2 and 3. In detail, out of these 87 data-sets 63 of them also showed good model data fit for four indices under model 2 and 3. That is, even for misspecified model all four fit indices indicated good model data fit.

Table 2. Number of Fit Indices

	Number of Indices	Model 1				Total
		1	2	3	4	
Model 2	1	0	1	6	2	9
	2	0	1	2	8	11
	3	0	0	1	14	16
	4	0	0	1	63	64
Model 3	1	0	2	5	4	11
	2	0	0	5	14	19
	3	0	0	0	6	6
	4	0	0	1	63	64
Total		0	2	10	87	100

The correlation coefficient between the factors is an important research question for many studies. In this study, root mean square error (RMSE) was calculated for the correlation coefficients obtained from the models by using .70 for the true correlation coefficient. Table 3 shows the descriptive values

of the correlation coefficients obtained from each model. Considering the table, the correlation coefficients for models 2 and 3 seem to have been greater than the true value. As shown in Table 3, the mean correlation coefficient for these models was .74 and .75, respectively. Besides, the RMSE values of models 2 and 3 were also higher than for the value of Model 1.

Table 3. Estimated Correlation Coefficients Statistics

	Min	Max	Mean	S. Deviation	RMSE
Model 1	.56	.89	.71	.07	.069
Model 2	.58	.97	.74	.07	.086
Model 3	.60	.94	.75	.08	.092

An example for Application

How the use of EFA or CFA first may changes the results of scale adaptation or development study was investigated in this section. Table 4 shows the correlation coefficients between the items and the mean and standard deviation values of the items in a data set with sample size of 300.

Table 4. Item Correlation and Descriptive Statistics

Items	M1	M2	M3	M4	M5	M6	M7	M8
M2	.34							
M3	.38	.32						
M4	.21	.21	.22					
M5	.19	.22	.19	.02*				
M6	.26	.20	.18	.15	.17			
M7	.23	.23	.24	.18	.12	.37		
M8	.32	.24	.22	.13	.21	.34	.37	1.00
Mean	3.00	3.01	3.01	2.97	3.04	3.01	3.03	3.06
Standard Deviation	.93	.92	.89	.91	.84	.94	.94	.85

* p>.05

This data set was first tested with each CFA models in the Mplus 5.1 program. MLR was used for the estimation. The model results were shown in table 5. Based on the results each model indicated good model-data fit in terms of all the fit indices. Mplus modification indexes had also given no warning. In the light of these results, a researcher who had started to research with any of the model will not need to try a second model because he/she had a good model-data fit already. Due to the nature of CFA, there is also no need for such a search when the model was *confirmed*. Therefore, the model set as default will be presented as the result. However, as it was seen, a data set fit well with all three models at the same time.

Table 5. The Result of CFA Models

	Chi-Square	Sd	p-value	CFI	RMSEA	SRMR
Model 1	14.13	19	.78	1.00	.00	.03
Model 2	12.25	19	.87	1.00	.00	.03
Model 3	21.52	19	.31	.99	.02	.04

In the second case, an EFA was run on SPSS with the same data to answer the question of what kind of a situation would occur. Thus, how the results can be changed when a researcher runs the same data set with EFA or CFA can be pointed. Since the PCA was not a factor analysis (Brown, 2006; Schmitt, 2011), principal axis factoring (PAF) was used as the estimation method. Since it was expected to have correlation between possible factors *promax* rotation was used. According to the results of this factor analysis (EFA1), the KMO value was .80 and Bartlet's test was significant ($\chi^2 = 319.08$, $p < .01$). As a result, the data set was suitable for an EFA. Table 6 shows EFA1 results. According to the factor analysis, a two-factor structure was formed.

Factor loadings in a factor analysis are expected to be higher than .30 (Martin & Newell, 2004; Seçer, Halmatov & Gençdoğan, 2013). The fact that the factor loadings of item 5 were less than this value in both factors which indicated inadequacy of the item. In addition, the internal consistency (Cronbach alpha) of the five items loaded on the first factor increases slightly, in the case the item 5 was removed. Accordingly, it was decided to remove item 5 from the analysis.

The result of new factor analysis (EFA2) was given in table 6. KMO and Bartlett test indicate that the data was suitable for factor analysis. According to the result, there were four items in the first factor and three items in the second factor. Internal consistency of the factors were .61 and .63 respectively. Each item was only loaded on one factor and these loadings were higher than .30. To conclude, the model without item 5 (EFA2) gave better result for the given data set.

Table 6. Results of Exploratory Factor Analysis

	EFA 1		FFA 2	
	Factor 1	Factor 2	Factor 1	Factor 2
Item 1	.58		.59	
Item 2	.55		.53	
Item 3	.63		.63	
Item 4	.31		.34	
Item 5	.23	.12	-	-
Item 6		.61		.61
Item 7		.61		.63
Item 8		.54		.53
KMO	.80		.79	
Bartlett's test	319.08		289.70	
p-value	.00		.00	
Correlation between factors	.64		.64	
Eigenvalues	2.66	1.04	2.55	1.04
% Variation	24.91	4.71	27.01	5.36

CONCLUSION and DISCUSSION

There is not a common way of using EFA or CFA first for scale adaptation studies. For an adaptation study some studies started with EFA while others started with CFA (Güvendir and Özkan, 2015). EFA is used when it is not known how many factors there are between the items and which factors are determined by which items while CFA is used if there is a strong theory about the structure. In this study, a data set is examined to fit to more than one CFA model via a simulation study. In addition, a data set was investigated to show how the use of EFA or CFA first might affect the results a scale adaptation. The simulated data generated according to a model specified in the R program were analyzed in the Mplus and SPSS programs.

Firstly, three different CFA models were evaluated for the same data set. The results clearly showed that more than one CFA model can fit well to a data set. For example, 63 of the 87 data sets that fit to model 1 also fit to model 2. This situation creates an ambiguity. *Which model shows the actual factor structure? Should all possible factor combinations be tried to determine the actual factor structure?* Using CFA for exploratory purposes may be limiting and even misleading the results (Schmitt, 2011). For this reason, as the result of the study showed, having a good fitting CFA model for a data set does not indicate that this model is actually the best model. In a scale adaptation studies, there may be changes in the structure resulting from cultural differences, as well as changes that may result from the item translation. Translating a scale into a new language requires not only translating language, but also language, culture and psychology as a whole (van de Vijver & Tanzer, 2004).

The possible models that may occur can be clearly defined in EFA. Structures not recognizable in CFA can easily be discovered through EFA (Bandalos & Finney, 2010). In other words, the possible changes in the structure in adaptation studies can be easily understood with the help of EFA. It is normal to have a change in the structure when a scale is translated into another language. It may even be possible to remove an item from the scale in some cases. Based on the results of this study, in order to achieve a consistent result and to establish a standard in scale adaptation studies, it is suggested to start with

an EFA to notice possible differences across cultures and languages. Then, a CFA will be a good step to verify the structure of adapted scale by using a different data set.

The use of different approaches in scale adaptation studies results in quite different conclusions. In this study, points to be considered in scale adaptation or development studies were indicated. How the results may change was examined via a simulation study. Also, differences were highlighted on a data set. As a result, in adaptation studies as well as in scale development studies, it is recommended to run an EFA and then a CFA to show the validity of the structure. Let us assume that the structure of the adapted scale is the same for both the adaptation and the original language. In this case, it will not be a problem to start with FFA and the same result will be achieved in every way. Otherwise, if there is a change in the structure, as it is seen in this study, we may not be able to detect it only with CFA. Therefore, it would be more beneficial to run EFA first in adaptation studies. Changing the design factors can change the results.

In this study, simulation design factors were limited because the aim was to show that a data set can fit different models. In other words, the sample size was 300, the correlation was fixed to .70 between the factors and the factor loadings had fixed values. It can be said that changing the design factors of the simulation may change the results. But this will not affect the conclusion that a data set can fit more than one model. Therefore, this was sufficient, even for one case (constant correlation and sample size). However, by changing the design factors, the simulation studies can be repeated and be examined in terms of the fit indices.

REFERENCES

- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York, NY: Routledge.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, W. S., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation, 18*(6), 1-13.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yöntemleri, 32*, 470-483.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-8.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of explanatory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.
- Güvendir, M. A. & Özkan, Y. Ö. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi, 14*(52), 23-33.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205. doi: 10.1177/1094428104263675
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. doi: 10.1177/0013164405282485
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi:10.1080/10705519909540118.
- Hurley A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vanderberg, R. J., & Williams L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18*, 667-683.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Martin, C. R., & Newell, R. J. (2004). Factor structure of the hospital anxiety and depression scale in individuals with facial disfigurement. *Psychology, Health & Medicine, 9*(3), 327-336. doi:10.1080/13548500410001721891.
- Muthén, B., & Muthén, L. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

- Orçan, F., & Yang, Y. (2016). A note on the use of item parceling in structural equation modeling with missing data. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 59-72. doi:10.21031/epod.88204.
- Schmitt R. S. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304-321. doi:10.1177/0734282911406653.
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York, NY: Routledge.
- Seçer, İ., Halmatov, S., & Gençdoğan (2013). Duygusal tepkisellik ölçeğinin Türkçeye uyarlanması: Güvenirlilik ve geçerlilik çalışması. *Sakarya University Journal of Education*, 3(1), 77-89.
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17, 268-274. doi:10.1111/j.1365-2753.2010.01434.x
- Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, 126, 124-128. doi:10.1053/j.gastro.2003.10.016.
- van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée*, 54, 119-135. doi:10.1016/j.erap.2003.12.004