Analyzing the Impacts of Alternated Number of Iterations in Multiple Imputation Method on Explanatory Factor Analysis[1]

Duygu KOÇAK[2] & Ömay ÇOKLUK-BÖKEOĞLU[3]

ABSTRACT

The study aims to identify the effects of iteration numbers used in multiple iteration method, one of the methods used to cope with missing values, on the results of factor analysis. With this aim, artificial datasets of different sample sizes were created. Missing values at random and missing values at complete random were created in various ratios by deleting data. For the data in random missing values, a second variable was iterated at ordinal scale level and datasets with different ratios of missing values were obtained based on the levels of this variable. The data were generated using "psych" program in R software, while "dplyr" program was used to create codes that would delete values according to predetermined conditions of missing value mechanism. Different datasets were generated by applying different iteration numbers. Explanatory factor analysis was conducted on the datasets completed and the factors and total explained variances are presented. These values were first evaluated based on the number of factors and total variance explained of the complete datasets. The results indicate that multiple iteration method yields a better performance in cases of missing values at random compared to datasets with missing values at complete random. Also, it was found that increasing the number of iterations in both missing value datasets decreases the difference in the results obtained from complete datasets.

Key Words: Construct validity, Explanatory factor analysis, Factor number, Missing value, Multiple iteration

---

[1] This study was presented as an oral presentation at the 26th International Conference on Educational Sciences (20-23 April 2017).

[2] Asst. Prof. Dr. - Alanya Alaaddin Keykubat University, Faculty of Education - duygu.kocak@alanya.edu.tr
[3] Assoc. Prof. Dr. - Ankara University, Faculty of Education Sciences - cokluk@education.ankara.edu.tr

INTRODUCTION

Subject matters explored in the fields of education and psychology are often abstract issues. Observing such abstract matters depends on being able to transform these patterns in a way that their concrete reflections on observable behavioral traits can be traced. Therefore, suitable instruments that can measure such patterns are developed and studies that can reveal to what extent the developed instruments are valid are conducted. In order to determine construct validity, analyzing response processes, clustering analysis, consulting expert opinions, analyzing internal consistency, correlation with other measures, or factor analysis could be conducted, and the results obtained can be the evidence for construct validity (Baykul, 2000). As one of these methods, Factor Analysis aims to reduce variables into fewer latent variables, called factors, based on the relationships among variables of a pattern, i.e., moving from covariance among variables. This method is particularly common in social science studies in order to develop scales, to confirm validity in replication studies, to determine the factors of a pattern, or to examine the validation of previously determined factors. Factor analysis can be performed for confirmatory or explanatory purposes. Explanatory factor analysis (EFA) aims to reduce the number of factors by identifying the relationships among them (Çokluk, Şekercioğlu, & Büyüköztürk, 2012). In conducting explanatory factor analysis, explained variance and the variance that factors explain are among the fundamental points while clustering and defining the factors are among the most important steps. At this point, the reactions or responses towards variables (the items or statements), also referred to as “answers,” as indicators of a pattern are important since explanatory factor analysis is conducted to explore a pattern and the factors that form that pattern moving from available data. Therefore, being unable to explain some variables can lead to have missing data and thus can affect the reliability of the conclusions (Koçak, 2016).

Hohensinn and Kubinger (2011) state that missing value is a problem experienced in almost every study, no matter how carefully designed. It is not always possible to conduct statistical analysis without solving missing value problems since statistical analysis is calculated using matrices composed of lines that represent individuals and columns that represent variables. Therefore, having all the matrices filled is the desired case, if not an obligation, as statistical analyses require datasets without missing values to in order to produce reliable results. In the 20th century, this need created an increased interest in the field to search for sound solutions for missing values. That is why most statistical analysis methods developed during the 20th century require datasets without missing values. In order to compensate for such problems, deleting listing columns, iteration regressions, iteration with mean values, iteration with expectation maximization, or multiple iterations are among the methods developed based on deleting or iteration used to deal with missing values (Graham, 2009). The power of these methods have been investigated under various conditions and found that performances varied depending on the size of the missing values and the degree of importance of these missing values (whether or not they could be ignored) (Allison, 2007; Little, & Rubin, 1987; Rubin, 1987; Schaffer, 1997).

Negligible missing value refers to missing values that are random; i.e., they do not have a pattern that could result in deviation or difference in data distribution (Rubin, 1976; Enders, 2010). On the other hand, when the missing values cannot be ignored, the conclusions drawn based on the findings would be biased, while analyses conducted using datasets with random missing values would result in similar findings to those from datasets without any missing values. Therefore, it is very important to determine whether missing values are completely random (CRM) or random (RM).

CRM mechanism refers to the case where missing values have no relationship with the variables they are in and with the variables of the individuals (Rubin, 1976). Donders, van der Heijden, Stijnen, and Moons (2006) state that when the dataset is in CRM mechanism, the dataset composed of individuals without missing values is a random sample of the research population. In other words, in the case of CRM, the dataset comprised of variables without missing values will be a random sample of the population, which includes the main data. Random missing value mechanism means that there is a possibility of having missing values in any variable which is related to the other variables in the model, but not with the observed values in the variable it is in. Enders (2010) explains that the case of missing values at random means that there is a systematic relationship between missing values and one or more variables. If the possibility of having missing values has a relationship with the variable they are in or the other variables, then we have "missing values not at random" mechanism. It is suggested that the methods used to cope with missing values have different performance powers in each missing value mechanism. In other words, the method to be used to deal with missing values for each mechanism has a varying performance in obtaining results close to the values without any missing values in the same dataset, thus, it is not possible to claim that one single method would perform well in all mechanisms under all conditions. For example, Satıcı (2009) claims that several methods perform well in CRM mechanism. Studies conducted by Donders et al. (2006) and Satıcı (2009), on the other hand, suggest that model-based methods and multiple iteration methods result in desired performances for RM mechanisms. Enders (2010) states that when the values missing are not at random, no method would produce a good performance.

There are various methods used to cope with missing values. Among these are deleting methods such as deleting list or in pairs, as well as simple methods like iteration based on mean value or regression, and finally model-based methods such as iteration on expectation maximization algorithm or multiple iteration. Studies on missing values recommend methods with multiple iteration and high probability since they have stronger theoretical basis with fewer limiting assumptions and some advantages reducing subjectivity (Baraldi, & Enders, 2010; Young, Weckmen, & Holland, 2011; Demir, 2013; Enders, 2013; Kang, 2013).

Multiple imputation, as one of the model-based imputation methods, is commonly used since it yields very close datasets to the authentic sets and allows for the possibility of multiple iteration. In terms of decreasing the subjectivity in standard errors, multiple imputation method is more advantageous than single imputation models. While deleting values list-wise or adding mean values work only in CRM, the multiple imputation method yields strong performances in both CRM and RM mechanisms (Ginkel, van der Ark, Sijtma, & Vermunt, 2010). Rubin (1987) claims that multiple imputation models are better than single iteration methods as they increase predictability value when iterations are random and allow for multiple iterations. It is also suggested that in order to decrease subjectivity in variance assumption, value iterations in multiple imputation should be random and follow producing predictability values based on linear regression, and that error distribution should be identified for each regression equation. These errors should be added to each regression equation created to develop data (Allison, 2003, 2009).

In multiple imputation, m number of data iteration is performed for every missing value and m number datasets are produced as a result of number iterations. Afterwards, the mean value of these datasets are calculated and they are transformed into one single dataset. This single set is treated as the original dataset for statistical analyses (Rubin, 1987; Ginkel et al., 2010; Granberg-Rademacker, 2007). Multiple imputation provides a different solution for each iteration. If m number of results are close, the method should be chosen; however, if the results differ significantly, i.e. if the results of the iterations differ, then standard errors should be considered when evaluating these differences (Acock, 2005). Therefore, the number of iterations is an important aspect in multiple imputation method and is explained from different perspectives in the literature. While McKnight, McKnight, Sidani, and Figueredo (2007) claim that between three and ten iterations can be sufficient, Cheema (2012) and Rubin (1987) suggest between two and ten iterations; whereas Schafer (1997) and Graham (2009) state between three and five iterations. Schafer (1997) states that iteration intervals vary in multiple imputation and when there are more than 10 iterations, these differences decline whereas if between three and five iterations are used, the difference will increase but not necessarily resulting in unreliable conclusions since it includes an error factor. These discussions remind us that differences resulting from iteration numbers in multiple imputation methods reflect an error factor and that this could affect the psychometric qualities of the dataset. Thus, when multiple imputation method is used to cope with missing values, it is worth evaluating the effects of iteration number on Bartlett Sphericity Test results, on factor construction (KMO), on total variance explained, and on the number of factors and their loadings. In this respect, the current study aims to determine the effects of the number of iteration in multiple imputation method on the results of explanatory factor analysis.

METHOD

Design of the Study

The current study is a simulation-based research which aims to identify the effects of iteration number used in multiple iteration method, one of the methods used to cope with missing values, on the results of factor analysis. Thus, basic research design has been adopted in order to conduct the study.

Data Generation and Analysis

The analyses in the study were conducted on artificial (simulative) datasets. To create the datasets, "sim.poly.ideal.npn(…)" code in {psych} packet in R program was used. Since the study is based on explanatory factor analysis, factor analysis assumptions were considered in creating the datasets. One of these assumptions is the size of the samples. Although there are different views on the sufficient sample size for factor analysis, it is generally accepted that 200 is good and 500 is a very good number for factor analysis (Cattell, 1978; Gorsuch, 1983; Comrey, & Lee, 1992). Accordingly, the analyses for the current study were conducted on two samples, n=200 and n=500. When creating the number of items in the datasets, the common numbers in the fields of psychology and education were considered; and thus 30 items for both datasets with six factors were determined. As a result, two separate multi-categorized 200x30 and 500x30 datasets scored at 1-2-3-4-5 were created.

Generating Missing Values

Missing values were created in the developed datasets using CRM and RM mechanisms. The missing values were generated using {psych} program in R software, while "dplyr" program was used to create codes that would delete values according to the determined conditions of missing value mechanism. Not only how far the missing values are negligible but also how big their rates are is important since the existence of missing values will impact the power of interpretations from the results as it restricts the dataset (Allison, 2009; Little & Rubin, 1987; Rubin, 1987; Schafer, 1997). The rates of missing values in the data generated were determined to be 2%, 5% and 10% based on the suggestions provided in previous studies (Enders, & Bandalos, 2001; Chen, Wang, & Chen, 2005; Fiona et al., 2006).

When developing MCR mechanism, cells were deleted in 2%, 5% and 10% rates independent of any variables. For example, there are 6,000 cells (200x30) in the n=200 sample. Here, in order to obtain 2% missing value rate in CRM mechanism, 120 cells (6000x2/100) were deleted. As for RM mechanism, a new variable with three categories (1-2-3) was added to all individuals on ordinal scale since RM refers to missing values resulting from another variable. It was also important to have equal number of individuals at all levels. For example, in the n=200 sample, 1 for 66 individuals, 2 for 67, and 3 for 67 individuals were assigned randomly. Then, missing values at different rates were created in each. The missing values were 10% for 1, 30% for 2, and 60% for 3 since this variable is on ordinal scale and requires different rates of missing values based on the different levels of variables. For instance, for n=200 sample size, 120 cells were deleted to have 2% missing value rate, where 10% (12 cells) were in 1, 30% (36 cells) in 2 and 60% (72 cells) were in 3. Following the same procedure for CRM and RM mechanisms for the n=200 and n=500 samples, 12 different datasets with 2%, 5%, and 10% missing value rates were created.

Completion of missing Values

Schafer (1997) states that when the number of iterations is over 10, there is a significant decrease in differentiation. Based on this assumption, the number of iterations in multiple imputation method used was limited to between 1-10 by applying a different number of iterations for each dataset for missing values.

For example, in the CRM sample of 200 with 2% missing value, the number of iterations varied between one and ten for each dataset in 10 different sets. As a result, 120 datasets with different iteration numbers were created. SPSS program was used to complete the missing values by multiple imputation method. The completed datasets were then analyzed using explanatory factor analysis. The results from the completed sets were then compared to the results obtained from originally complete datasets. The datasets without missing values were used as reference values.

FINDINGS

In this section, results are presented for the impacts of the number of iterations in multiple imputation method used for missing values on the explained variance rate in explanatory factor analysis, and on the number of factors obtained for CRM and RM conditions in the n=200 and n=500 sample size datasets.

The results of KMO and Bartlett Sphericity Test for the n=200 sample size with CRM are presented in Table 1.

Table 1. *KMO & Bartlett Sphericity Test in CRM (n=200)*

		Missing Value Ratios								
		2%			5%			10%		
		KMO	χ^2	p	KMO	χ^2	p	KMO	χ^2	p
Complete data		*.877*	*2,106.57*	*.00*	*.877*	*2,106.57*	*.00*	*.877*	*2,106.57*	*.00*
No. Iteration	1	.852	2,138.63	.00	.843	2,189.38	.00	.823	2,311.35	.00
	2	.866	4,379.97	.00	.847	4,424.80	.00	.833	4,675.00	.00
	3	.866	6,657.38	.00	.855	6,722.55	.00	.840	6,841.77	.00
	4	.863	8,923.14	.00	.857	8,957.90	.00	.846	9,369.43	.00
	5	.864	11,175.18	.00	.858	11,209.42	.00	.844	11,379.07	.00
	6	.864	13,374.09	.00	.859	13,531.56	.00	.847	13,714.61	.00
	7	.862	15,586.40	.00	.853	15,725.89	.00	.846	16,044.69	.00
	8	.864	17,916.26	.00	.856	17,979.70	.00	.849	18,233.82	.00
	9	.862	20,223.12	.00	.857	20,226.51	.00	.847	20,477.07	.00
	10	.863	22,430.72	.00	.859	22,405.41	.00	.849	22,979.70	.00

As can be seen in Table 1, KMO value for the complete dataset is .877. In the case of CRM, KMO values for all sets with different missing value rates and different number of iterations varied between .823 and .866. Having KMO values between .8 and .9 is considered "good" in terms of obtaining factors from the data (Şencan, 2005; Tavşancıl, 2005). When the missing value rate was 2%, it is observed that similar KMO values were obtained in different numbers of iterations. It is also observed that when the missing value rate is small (2%) in a small sample size (n=200), similar values are obtained in all number of iterations. When the missing value rate is 5%, similar values to other completed datasets and the complete sets are obtained when the number of iterations is three or higher. As the missing value rate increases to 10%, then the number of iterations need to be four or above in order to obtain close values. However, Bartlett Sphericity Test results indicate meaningful values for all rates of missing values ($p<.01$). These results indicate that simulated complete datasets and the datasets with completed missing values in different rates are suitable to obtain factors. It is found that when the missing value rates in CRM conditions are at low levels, the number of iterations can be two or over in order to obtain results similar to those from complete datasets. If the missing value rates are 5% or 10%, the iteration number should be three or over in order to obtain results similar to those from complete datasets. When the performances of multiple imputation methods used in datasets with different missing value rates are compared, it is found that the difference in the obtained values increases as the missing value rates increase; in other words, the method's performance decreases. Although multiple imputation method is accepted to have a better performance compared to other methods in obtaining values similar to those from complete datasets, its performance weakens like all the other methods as the rate of missing values grows (Enders, 2013; Kang, 2013; Young et al., 2011). The results obtained in this study seem to yield parallel results with previous studies.

KMO and Bartlett Sphericity Test results of for the n=500 sample size with CRM are presented in Table 2.

Table 2. *KMO & Bartlett Sphericity Test in CRM (n=500)*

		Missing Value Ratios								
		2%			5%			10%		
		KMO	χ^2	p	KMO	χ^2	p	KMO	χ^2	p
Complete data		*.922*	*4,473.58*	*.00*	*.922*	*4,473.58*	*.00*	*.922*	*4,473.58*	*.00*
No. Iteration	1	.919	4,477.09	.00	.915	4,560.60	.00	.901	4,617.79	.00
	2	.920	8,987.04	.00	.914	9,197.33	.00	.908	9,188.30	.00
	3	.921	13,610.58	.00	.916	13,870.69	.00	.911	14,011.45	.00
	4	.921	18,142.25	.00	.916	18,591.59	.00	.909	18,293.35	.00
	5	.922	22,713.46	.00	.917	23,202.10	.00	.910	23,046.18	.00
	6	.921	27,261.46	.00	.915	27,660.87	.00	.909	27,576.62	.00
	7	.922	31,925.85	.00	.917	32,284.53	.00	.911	32,361.23	.00
	8	.922	36,401.43	.00	.915	36,898.98	.00	.911	36,897.28	.00
	9	.921	41,075.94	.00	.916	41,504.54	.00	.910	41,502.83	.00
	10	.921	45,519.49	.00	.916	46,231.02	.00	.909	46,211.06	.00

Table 2 shows that KMO value for complete dataset is .922. In case of CRM, KMO values for all sets with different missing value rates and different number of iterations varied between .901 and .922.

When KMO values are above .9, it is considered to be "very good" in terms of obtaining factors from the data (Şencan, 2005; Tavşancıl, 2005). When the missing value rate was 2% in the n=500 sample size, similar KMO values were obtained in different numbers of iterations and in the complete dataset. It is also observed that when the missing value rate is small (2%) in small sample size (n=200), similar values are obtained for all number of iterations. When the missing value rate is 5%, similar performances were observed for all number of iterations. Although the results were not as close to each other as the ones in the 2% missing value rate, they are still close to the results obtained from the complete sets. As the missing value rate increases to 10%, two or above iterations could result in similar values to other completed datasets and the complete sets. Bartlett Sphericity Test results indicate meaningful values for all rates of missing values ($p<.01$). When the results are evaluated, it can be claimed that under CRM conditions and when the sample size is n=500, the performance of multiple imputation model decreases regardless of iteration number. However, this is similar for all other methods used to deal with missing values as the performance of these methods are associated with the rate of missing value, which result in lower performances as the rate increases (Allison, 2003; Baraldi & Enders, 2010; Graham, 2009; Rubin, 1987; Schafer, 1997). Similar performances are obtained from all numbers of iteration when the missing value rate is low (2%) or moderate (5%) while having high level (10%), the number of iteration needs to be two or above.

When the performances of multiple imputation in the n=200 and n=500 sample sizes are compared, it can be stated that the n=500 sample size yields better results, which are closer to datasets without missing values. This is in line with the literature which explains that the larger the sample size, the closer the conclusions are to the population's parameter (Agresti & Finlay, 1997; Allison, 2002, 2009). In addition, Allison (2009) indicates that model-based methods create probability predictions and perform poor in small sample sizes.

KMO and Bartlett Sphericity Test results of for the n=200 sample size with RM are presented in Table 3.

Table 3. *KMO & Bartlett Sphericity Test in RM (n=200)*

		Missing Value Ratios								
		2%			5%			10%		
		KMO	χ^2	p	KMO	χ^2	p	KMO	χ^2	p
Complete data		*.877*	*2,106.57*	*.00*	*.877*	*2,106.57*	*.00*	*.877*	*2,106.57*	*.00*
No. Iteration	1	.861	2,143.83	.00	.836	2,223.70	.00	.819	2,266.36	.00
	2	.862	4,426.56	.00	.851	4,483.81	.00	.847	4,394.24	.00
	3	.865	6,703.33	.00	.852	6,775.72	.00	.837	6,785.00	.00
	4	.864	8,906.27	.00	.852	9,027.15	.00	.838	9,158.27	.00
	5	.864	11,174.11	.00	.856	11,365.67	.00	.844	11,233.44	.00
	6	.864	13,386.41	.00	.855	13,612.44	.00	.845	13,469.57	.00
	7	.865	15,634.17	.00	.857	15,852.14	.00	.839	15,720.04	.00
	8	.864	17,966.18	.00	.856	18,091.17	.00	.848	17,901.79	.00
	9	.864	20,251.90	.00	.859	20,457.37	.00	.848	20,076.97	.00
	10	.864	22,498.85	.00	.855	22,700.65	.00	.844	22,335.15	.00

Table 3 presents the results regarding the effects of various numbers of iteration in various rates of missing values on KMO and Bartlett Sphericity test in RM mechanism for the n=200 sample. When the size is 200 and the missing value rate is 2%, similar results in all numbers of iterations including the complete datasets were obtained. When the missing value rate is moderate at 5%, similar values to other completed datasets and the complete sets are obtained where the number of iterations is two or higher.

When the missing value rate is 10%, similar values to other completed datasets and the complete sets are obtained for iterations of two or above. However, it is found that when the missing value rate increases, the results in all numbers of iterations gets further from KMO results obtained from complete datasets. Also, the performance in RM mechanism weakens as the missing value rate rises regardless of iteration number applied. Bartlett Sphericity Test results indicate meaningful values for all rates of missing values ($p<.01$). The possibility of having factors was maintained for all conditions and numbers of iterations. Allison (2007), Little and Rubin (1987), Rubin (1987), and Schaffer (1997) state that multiple imputation method can be used especially for datasets with RM. When CRM and RM conditions are compared under similar conditions in similar sample sizes, it can be claimed that RM produces closer and more consistent results. This supports that multiple imputation method yields better performance in RM condition. As the missing value rate increases, the performance decreases.

KMO and Bartlett Sphericity Test results of for the n=500 sample size with RM are presented in Table 4.

Table 4. *KMO & Bartlett Sphericity Test in RM (n=500)*

		Missing Value Ratios								
		2%			5%			10%		
		KMO	χ^2	p	KMO	χ^2	p	KMO	χ^2	p
Complete data		*.922*	*4,473.58*	*.00*	*.922*	*4,473.58*	*.00*	*.922*	*4,473.58*	*.00*
No. Iterations	1	.914	4,470.338	.00	.910	4,501.86	.00	.907	4,443.65	.00
	2	.915	9,009.851	.00	.912	9,052.16	.00	.907	8,974.85	.00
	3	.916	13,599.698	.00	.914	13,611.72	.00	.908	13,466.23	.00
	4	.918	18,204.086	.00	.914	18,178.81	.00	.912	17,975.95	.00
	5	.917	22,740.191	.00	.914	22,724.87	.00	.910	22,397.08	.00
	6	.917	27,362.411	.00	.915	27,227.81	.00	.912	26,893.41	.00
	7	.917	31,886.458	.00	.915	31,792.22	.00	.911	31,477.08	.00
	8	.917	36,461.103	.00	.914	36,374.17	.00	.912	35,893.43	.00
	9	.917	41,196.83	.00	.915	40,887.70	.00	.912	40,189.80	.00
	10	.917	45,537.43	.00	.914	45,470.09	.00	.912	44,584.72	.00

When Table 4 is analyzed, it is seen that 2% missing value sets in RM condition display similar KMO values (.917). Having KMO values between above .9 is considered to be "very good" in terms of obtaining factors from the data (Şencan, 2005; Tavşancıl, 2005). When the missing value rate is 2% and the KMO is .922, it is observed that similar KMO values are obtained in both complete datasets and in sets completed with different numbers of iterations. It is also observed that when the missing value rate is small and the sample size is n=500, similar values are obtained in all number of iterations. The same case is observed in case of 5% and 10% missing value sets.

Similar KMO values were obtained in all numbers of iterations. Nevertheless, similar to the case in CRM conditions, all numbers of iterations show weaker performance and move away from the values obtained in complete datasets as the rate of missing value increases. Previous studies also indicate that methods applied to cope with missing values show weaker performance when the rate of missing value inclines (Enders, 2013, Rubin, 1987; Schaffer, 1997). The results from the current study are in line with this claim.

When the performances under CRM and RM conditions are compared, it can be stated that multiple imputation method performs better in small sample sizes under CRM conditions for all numbers of iterations. The same applies when the sample size is 500. Both missing value mechanisms show lower performances as the missing value rate increases.

When the missing value rate is 2% in all samples, iterations of two or above yield closer performances to complete datasets. When the sample size is 500 and the missing value rate is 5%, all numbers of iterations show similar results. In small sample sizes with 10% missing value rates, the number of iterations should be three or above in order to obtain similar results to the complete datasets. In samples of 500 with 10% missing value rates, all numbers of iterations show similar results. However, when the missing rate is high under RM conditions, closer results to complete datasets are obtained as the number of iterations is increased. This would imply that increasing the number of iterations in multiple imputation method could decrease the subjectivity created by missing values.

When the sample size is 500, better results were obtained than the samples of 200 under both CRM and RM conditions. In multiple imputation method, closer results can be obtained as the sample size increases and the predictions get closer to the population's parameter (Agresti, & Finlay, 1997; Allison, 2002, 2009; Baykul, & Güzeller, 2013). Multiple imputation method performed better in all numbers of iterations in bigger sample sizes. Table 5 presents the results for the impacts of the number of iterations in multiple imputation method used for missing values on the explained total variance rate in explanatory factor analysis and on the number of factors obtained for CRM conditions in the n=200 sample size datasets.

Table 5. *Factors, Eigen Values, & Total Variance Explained for CRM (n=200)*

			No. Iteration									
	Factor	*Complete data*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
2% Missing	1	*7,116*	7,029	7,118	7,093	7,083	7,098	7,078	7,097	7,088	7,084	7,090
	2	*3,990*	3,958	3,919	3,980	3,987	3,971	3,949	3,929	3,968	3,947	3,965
	3	*2,070*	2,074	2,057	2,059	2,064	2,065	2,071	2,052	2,047	2,076	2,059
	4	*1,445*	1,406	1,438	1,433	1,413	1,417	1,438	1,430	1,437	1,433	1,421
	5	*1,114*	1,151	1,131	1,125	1,125	1,126	1,130	1,128	1,125	1,127	1,132
	6	*1,025*	1,068	1,059	1,069	1,060	1,051	1,054	1,055	1,047	1,052	1,053
	Ex var	*55,973*	55,623	55,741	55,864	55,775	55,767	55,734	55,639	55,711	55,977	55,735
5% Missing	1	*7,116*	6,979	7,006	7,121	7,050	7,084	7,093	7,053	7,053	7,033	7,020
	2	*3,990*	4,005	3,902	3,888	3,901	3,927	3,921	3,906	3,909	3,919	3,921
	3	*2,070*	2,074	2,114	2,074	2,087	2,083	2,122	2,089	2,100	2,121	2,095
	4	*1,445*	1,514	1,460	1,463	1,457	1,480	1,437	1,445	4,446	1,438	1,445
	5	*1,114*	1,153	1,129	1,107	1,145	1,121	1,107	1,124	1,133	1,130	1,149
	6	*1,025*	1,075	1,051	1,048	1,065	1,051	1,061	1,058	1,041	1,057	1,056
	Ex var	*55,973*	56,005	55,538	55,670	55,686	55,648	55,808	55,582	55,609	55,662	55,625
10% Missing	1	*7,116*	6,921	7,010	6,953	7,099	6,973	6,931	6,966	6,965	6,928	6,986
	2	*3,990*	3,998	3,870	3,854	3,916	3,820	3,922	3,899	3,881	3,871	3,864
	3	*2,070*	2,270	2,196	2,179	2,202	2,221	2,225	2,165	2,186	2,221	2,218
	4	*1,445*	1,504	1,523	1,517	1,449	1,474	1,450	1,512	1,494	1,474	1,462
	5	*1,114*	1,172	1,131	1,180	1,169	1,161	1,137	1,616	1,137	1,138	1,169
	6	*1,025*	1,041	1,080	1,051	1,010	1,048	1,022	1,031	1,022	1,026	1,043
	7		1,025	1,026	1,007	1,011	1,014	1,014	1,001	1,001		
	Ex var	*55,973*	59,771	59,455	59,139	59,510	59,150	59,000	59,120	58,951	55,528	55,809

*Ex var = Explained variance

As Table 5 indicates, when the missing value rates are 2% and 5%, all explained variance and the total number of factors complies with the results from the complete dataset. However, when the missing value rate is 10% and the number of iterations are nine or ten, the results are still in line with those from the complete dataset, while the performance created too many factors. Under CRM condition, similar results were obtained for all missing value rates and all numbers of iterations. Yet, when the missing value rate increases, the number of iterations should be nine or above to avoid subjectivity. This could indicate that the performance of multiple imputation method decreases as the rate of missing values increases. This result is also in line with previous studies in the literature that state that performance weakens as the missing value rate increases (Enders, 2013; Rubin, 1987; Schaffer, 1997). Furthermore, the current study's results indicate that the number of iterations should be higher as the missing value rate increases to avoid subjectivity.

Table 5 also shows that when the missing value rate is 10% and the number of iterations is between one and eight, there are too many factors which leads to higher explained variance values. However, it should be noted that when evaluating the performance of a method used to cope with missing values, the criteria should include the closeness of the results to complete datasets. Increasing the number of iterations as the missing value rate increases is an option to obtain closer results to those from complete datasets.

Table 6 presents results for the impact of the number of iterations in multiple imputation method used for missing values on the explained total variance rate in explanatory factor analysis, and on the number of factors obtained for CRM conditions in the n=500 sample size datasets.

Table 6. *Factors, Eigen Values, & Total Variance Explained for CRM (n=500)*

			No. Iteration									
	Factor	*Complete data*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
2% Missing	1	*7,112*	7,079	7,058	7,097	7,088	7,103	7,097	7,104	7,099	7,095	7,095
	2	*3,684*	3,684	3,665	3,689	3,685	3,671	3,664	3,689	3,669	3,694	3,677
	3	*1,916*	1,926	1,912	1,895	1,897	1,914	1,912	1,901	1,906	1,908	1,900
	4	*1,125*	1,123	1,135	1,119	1,123	1,115	1,123	1,120	1,124	1,114	1,126
	Ex var	*46,120*	46,040	45,897	45,997	45,974	46,008	45,988	46,043	45,993	46,036	45,992
5% Missing	1	*7,112*	7,160	7,084	7,126	7,120	7,157	7,095	7,109	7,089	7,096	7,093
	2	*3,684*	3,667	3,699	3,663	3,701	3,660	3,669	3,673	3,682	3,660	3,695
	3	*1,916*	1,890	1,926	1,942	1,915	1,919	1,916	1,916	1,916	1,920	1,922
	4	*1,125*	1,133	1,107	1,132	1,130	1,110	1,128	1,125	1,130	1,128	1,120
	Ex var	*46,120*	46,164	46,050	46,211	46,453	46,157	46,026	46,074	46,056	46,018	46,101
10% Missing	1	*7,112*	7,061	7,006	7,108	6,997	7,037	7,030	7,078	7,057	7,047	7,033
	2	*3,684*	3,636	3,716	3,657	3,685	3,664	3,660	3,654	3,670	3,646	3,674
	3	*1,916*	1,894	1,897	1,936	1,890	1,929	1,916	1,908	1,886	1,928	1,905
	4	*1,125*	1,159	1,149	1,153	1,140	1,143	1,129	1,128	1,134	1,134	1,139
	5		1,023	1,004								
	Ex var	*46,120*	49,239	49,241	46,183	45,711	45,909	45,781	45,893	45,823	45,850	45,837

*Ex var = Explained variance

Table 6 shows that when the missing value rates are 2% and 5%, all explained variance and total number of factors comply with the results from the complete dataset. However, when the missing value rate is 10% and the sample size is 200, the results move further away from those obtained for a complete dataset. When the number of iterations are one or two, there are too many factors which leads to higher explained variance values. However, when the number of iterations is three or above, the results get closer to those from the complete dataset. This finding suggests that when the sample size is 500 with 2% and 5% missing value rates, closer results to those from the complete dataset are obtained, while less closer results are received when the missing value rate rises to 10% in case the number of iterations is three or above. The performance of multiple imputation method weakens as the missing value rate increases. When the number of iterations rises to 10%, the results are still in line with those from the complete dataset if the missing value rate is 2%, 5%, and 10%, respectively. Accordingly, it should be noted that the performance of multiple imputation method decreases as the rate of missing values increases.

Table 7 presents the results for the impacts of the number of iterations in multiple imputation method used for missing values on the explained total variance rate in explanatory factor analysis, and on the number of factors obtained for RM conditions in the n=500 sample size datasets.

Table 7. *Factors, Eigen Values, & Total Variance Explained for RM (n=200)*

			No. Iteration									
	Factor	*Complete data*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
2% Missing	1	*7,116*	7,078	7,074	7,093	7,073	7,084	7,046	7,041	7,072	7,065	7,071
	2	*3,990*	3,932	3,988	3,994	3,980	3,949	3,968	3,982	3,975	3,989	3,966
	3	*2,207*	2,123	2,100	2,089	2,071	2,095	2,113	2,096	2,099	2,088	2,095
	4	*1,445*	1,463	1,443	1,444	1,443	1,440	1,434	1,444	1,448	1,438	1,436
	5	*1,114*	1,157	1,163	1,169	1,151	1,066	1,171	1,165	1,156	1,164	1,175
	6	*1,025*	1,020	1,022	1,028	1,028	1,021	1,012	1,018	1,023	1,023	1,024
	Ex var	*55,973*	55,912	55,963	56,057	55,832	55,849	55,810	55,819	55,910	55,477	55,885
5% Missing	1	*7,116*	7,068	7,015	6,998	7,017	7,039	7,065	7,033	7,065	7,037	7,038
	2	*3,990*	4,063	3,987	4,058	3,979	4,046	3,975	3,979	3,975	4,040	4,010
	3	*2,207*	2,083	2,124	2,078	2,100	2,124	2,096	2,097	2,096	2,093	2,088
	4	*1,445*	1,428	1,456	1,460	1,442	1,426	1,426	1,437	1,426	1,423	1,419
	5	*1,114*	1,153	1,165	1,160	1,166	1,138	1,159	1,168	1,159	1,167	1,149
	6	*1,025*	1,025	1,018	1,006	1,006	1,003				1,017	1,011
	Ex var	*55,973*	56,065	55,880	55,864	55,701	55,918	52,399	52,382	52,399	55,924	55,718
10% Missing	1	*7,116*	7,050	6,898	6,897	6,958	6,915	6,895	6,911	6,903	6,918	6,906
	2	*3,990*	3,880	3,966	3,908	4,001	3,922	3,981	3,948	3,961	3,936	3,960
	3	*2,207*	2,059	2,108	2,076	2,066	2,095	2,042	2,084	2,067	2,074	2,067
	4	*1,445*	1,490	1,461	1,515	1,461	1,476	1,457	1,441	1,458	1,441	1,442
	5	*1,114*	1,201	1,193	1,215	1,206	1,197	1,217	1,194	1,205	1,182	1,206
	6	*1,025*	1,120	1,046	1,010	1,013	1,010	1,049		1,014	1,038	1,001
	Ex var	*55,973*	55,999	55,571	55,401	55,686	55,384	55,473	51,925	55,363	55,297	55,272

*Ex var = Explained variance

As Table 7 shows, when the missing value rate is 2%, all explained variance and total number of factors are in compliance with the results from the complete dataset for all number of iterations. Especially when the number of iterations are one, two or eight, the value of total variance explained is closest to the values of the complete dataset. This finding may suggest that there is no linear relation between the number of iterations and predictions of the complete dataset. In other words, it negates the claim that as the number of iterations rises under CRM and RM conditions, closer results to those from the complete dataset can be obtained, or vice versa. Similar results apply in cases of 5% and 10% missing values. Additionally, multiple imputation method results in fewer numbers of factors compared to those obtained in complete dataset in some missing value rates and number of iterations. When the sample size is 200 under RM conditions, the performance of multiple imputation method decreases as the rate of missing values increases, regardless of the number of iterations applied.

When the performances under CRM and RM conditions are compared for the same sample sizes, missing value rates and number of iterations, it can be stated that multiple imputation method performs better under CRM conditions. In the literature, it is stated that RM would perform better under these conditions (Allison, 2003; Baraldi, & Enders, 2010). In the process of prediction for the missing values, multiple imputation works mostly on probability aspect (Cheung, 2007). However, when the required conditions are met, it produces consistent, asymptotic normal and asymptotic prediction values (Agresti, & Finlay, 1997; Allison, 2002, 2009). RM mechanism performed weaker than CRM mechanism since the sample size was small and thus did not meet the assumptions.

Table 8 presents the results for the impacts of the number of iterations in multiple imputation method used for missing values on the explained total variance rate in explanatory factor analysis, and on the number of factors obtained for RM conditions in the n=500 sample size datasets.

Table 8. *Factors, Eigen Values, & Total Variance Explained for RM (n=500)*

			No. Iteration									
	Factor	*Complete data*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
2% Missing	1	*7,112*	7,029	7,037	7,060	7,083	7,069	7,068	7,070	7,076	7,074	7,067
	2	*3,684*	3,678	3,663	3,661	3,663	3,667	3,676	3,666	3,669	3,674	3,665
	3	*1,916*	1,928	1,920	1,917	1,918	1,905	1,914	1,906	1,902	1,913	1,905
	4	*1,125*	1,141	1,142	1,139	1,142	1,138	1,139	1,145	1,140	1,149	1,139
	Ex var	*46,120*	45,923	45,897	45,922	46,019	45,930	45,990	45,960	45,959	46,036	45,922
5% Missing	1	*7,112*	7,035	7,082	7,100	7,032	7,061	7,058	7,072	7,036	7,068	7,056
	2	*3,684*	3,636	3,600	3,162	3,640	3,628	3,635	3,618	3,637	3,617	3,636
	3	*1,916*	1,914	1,908	1,896	1,912	1,897	1,883	1,893	1,909	1,900	1,888
	4	*1,125*	1,164	1,139	1,126	1,157	1,139	1,156	1,143	1,152	1,138	1,143
	Ex var	*46,120*	45,827	45,765	45,024	45,806	45,750	45,775	45,754	45,784	45,741	45,740
10% Missing	1	*7,112*	6,999	6,930	6,955	6,934	6,968	6,956	6,940	6,988	6,935	6,948
	2	*3,684*	3,611	3,647	3,602	3,676	3,612	3,631	3,642	3,590	3,621	3,600
	3	*1,916*	1,876	1,916	1,920	1,924	1,897	1,906	1,932	1,911	1,913	1,915
	4	*1,125*	1,126	1,172	1,175	1,152	1,164	1,141	1,157	1,157	1,147	1,140
	Ex var	*46,120*	45,372	45,551	45,510	45,620	45,469	45,444	45,568	45,484	45,383	45,344

*Ex var = Explained variance

When Table 8 is analyzed, it is observed that total number of factors complies with the results from the complete dataset in all missing value rates and number of iterations. Also, when the missing value rates are 2%, all explained variance values are in line with the results from the complete dataset in all numbers of iterations. The case is the same for 5% and 10% missing value rates. When the iteration number is the same, it is found that the performance of multiple imputation model decreases as the rate of missing value increases.

CONCLUSION AND IMPLICATIONS

The current study aims to identify the effects of iteration number used in multiple iteration method, one of the methods used to cope with missing values, on the results of factor analysis. With this aim, artificial datasets in 200 and 500 sample sizes were created and missing values at random and missing values at completely random were created in 2%, 5% and 10% ratios by deleting data. The results were evaluated in terms of the number of factors, the total variance explained and Eigen values obtained.

The performance of multiple imputation method decreases when the rate of missing value increases under all conditions. In other words, as the rate of missing value rises, both CRM and RM conditions provide weaker results for both sample sizes regardless of the number of iterations. This finding is in line with the literature which states that the performance of all methods used to cope with missing values decreases as the rate of missing value increases gets bigger (Rubin, 1987; Schaffer, 1997; Collins, Schafer, & Kam, 2001; Baraldi, & Enders, 2010; Enders, 2013).

When the missing value rate is 10% and the sample size is 500 under CRM conditions, the number of iterations is few but still multiple imputation method is observed to perform strong and creates too many factors. However, with the same conditions, RM mechanism seems to create a similar number of factors to the original complete dataset. This would imply that RM mechanism performs better than CRM mechanism under the same conditions for multiple imputation method. This finding concurs with the literature in stating that multiple imputation method performs better in RM conditions (Rubin, 1987; Graham, Hofer, & Piccinin, 1994; Schaffer, 1997; Pigott, 2001; Arnold, & Kronmal, 2002; Streiner, 2002; Newman, 2003; Wayman, 2003; Barzi, & Woodward, 2004; Durrant, 2005; Baraldi, & Enders, 2010; Enders, 2013). However, the results differ for the n=200 sample size, where CRM mechanism performs better than RM in multiple imputation method.

In the process of predicting the missing values, multiple imputation method relies on probability assumptions (Cheung, 2007). As long as all the required assumptions are met, probability method produces consistent, asymptotic normal and asymptotic effective prediction values (Agresti, & Finlay, 1997; Allison, 2002, 2009). It is suggested that RM mechanism (as it does not meet the assumptions required) performs weaker than CRM mechanism. The reason for this could be that the missing values in RM mechanism might hinder obtaining asymptotic normality. The larger the sample size, the less likely will be the prediction of the unknown population's (or the complete dataset) parameter.

The performance of multiple imputation method decreases when the rate of missing value increases. At the same time, as the sample size increases, its performance gets stronger for all numbers of iteration. This would imply that multiple imputation method performs better for larger sample sizes and provides closer results to the population's parameter (Little, & Rubin, 1987; Rubin, 1987; Schaffer, 1997; Allison, 2007, 2009; Baykul, & Güzeller, 2013).

Under CRM condition with the n=200 sample size and 10% missing value rate, there are deviations from the results of the complete dataset. This deviation is eliminated when the number of iterations is nine or above. Furthermore, in the case of smaller sample sizes with high missing value rates in CRM conditions, iterations should be nine or above for multiple imputation method in order to provide the desired results. McKnight et al. (2007) suggest that between three and ten iterations would suffice, while for Cheema (2012) and Rubin (1987) it is between two and ten, and for Schafer (1997) and Graham (2009) it is between three and five. The findings of the current study suggest that it is necessary to have two or more iterations for a small sample size (n=200) with low rates of missing values under CRM conditions. Increasing the number of iterations did not produce any difference in the performance. It is therefore suggested that the number of iterations should be increased as the missing value rate rises. For instance, when the missing value rate is 5%, there should be at least three iterations. When the rate is 10%, the number of iterations should be increased to at least nine or ten. With larger sample sizes (e.g., n=500), the minimum number of iterations decreases. In this larger sample size, one iteration is sufficient with 2% missing value rate, but a minimum of two iterations are required for 5% missing value rate. When the rate is 10%, then the number of iterations should be at least three.

Under RM conditions, on the other hand, the number of iterations need to be at least two for the n=200 sample size with 2% and 5% missing value rates. However, when the rate is 10%, the method does not perform well under RM conditions. When the sample size is n=500, the number of iterations should be increased to four or more. Even though no specific number can be suggested for iterations to obtain strong performance, iterations of three or above are generally accepted to be sufficient. However, it should also be noted that there are no significant changes in the performance in this mechanism even when the number of iterations are in excess of the minimum numbers.

REFERENCES

Acock, A. C. (2005). Working with missing values. *Journal of Marriage & Family, 67*(4), 1012-1028.

Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences (4th ed.)*. USA: Pearson Prentice Hall.

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research, 28*(3), 301-309.

Allison, P. D. (2002). *Missing Data*. Sage Publications.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545-557.

Allison, P. D. (2009). *Missing data.* In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 72-89). Thousand Oaks, CA: Sage Publications.

Arnold, A. M., & Kronmal, R. A. (2002). Multiple imputation of baseline data in the cardiovascular health study. *American Journal of Epidemiology, 157*(1), 74-84. DOI: 10.1093/aje/kwf156.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology, 48*(1), 5-37.

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology, 160*(1), 34-45. DOI: 10.1093/aje/kwh175.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme*. Ankara: ÖSYM Yayınları.

Baykul, Y., & Güzeller, C. O. (2013). *Sosyal bilimler için istatistik: SPSS uygulamalı*. Ankara: Pegem Akademi.

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum Press.

Cheema, J. (2012). *Handling missing data in educational research using SPSS* (Unpublished Doctoral dissertation). George Mason University, Fairfax, VA.

Chen, S. F., Wang, S., & Chen, Y. C. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications, 39*(4), 4026-4031.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences.* Hoboken, New Jersey: John Wiley & Sons.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve Lisrel uygulamaları.* Ankara: Pegem Akademi Yayıncılık.

Demir, E. (2013). *Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi* (Unpublished Doctoral dissertation). Ankara University, Ankara.

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087-1091. DOI: 10.1016/j.jclinepi.2006.01.014.

Durrant, G. B. (2005). *Imputation methods for handling item-nonresponse in the social sciences: A methodological review*. ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute, NCRM Methods Review Papers NCRM/002. Southampton, UK: University of Southampton.

Enders, C. K. (2010). *Applied missing data analysis* (1st ed.). New York: Guilford Press.

Enders, C. K. (2013). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological Methods, 8*(3), 322-337.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430-457.

Ginkel, J. R. V., van der Ark, L. A., Sijtma, K., & Vermunt, J. K. (2007). Two-way imputation: A bayesian method for estimating missing scores in tests and questionnaires and an accurate approximation. *Computational Statistics & Data Analysis, 51*, 4013-4027. DOI:10.1016/j.csda.2006.12.022.

Gorsuch, R. L. (1983). *Factor analysis.* Philadelphia: Saunders.

Graham, J. W. (2009). *Missing data: Analysis and design*. New York: Springer.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins & L. A. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (pp. 13-64). Rockville, MD: National Institutes of Health.

Granberg-Rademacker, J. S. (2007). A comparison of three approaches to handling incomplete state level data. *State Politics and Policy Quarterly, 7*(3), 325-338.

Hohensinn, C., & Kubinger, K. D. (2011). On the impact of missing values on the item fit and the model validness of the Rasch model. *Psychological Test and Assessment Modeling, 53*(3), 380-393.

Kang, H. (2013). The prevention and handling of missing data. *Korean J Anesthesiology, 64*(5), 402-406.

Koçak, D. (2016). *Kayıp veriyle baş etme yöntemlerinin madde tepki kuramı bir parametreli lojistik modelinde model veri uyumuna ve standart hataya etkisi* (Unpublished Doctoral dissertation). Ankara University, Ankara.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York: Guilford Press.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood and multiple imputation techniques. *Organizational Research Methods, 6*(3), 328-362. DOI: 10.1177/1094428103254673.

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation: An International Journal of Theory and Practice, 7*(4), 353-383.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Satıcı, E. (2009). *Kayıp gözlem olması durumunda kitle ortalaması tahmini* (Unpublished Doctoral dissertation). Hacettepe University, Ankara.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. USA: Chapman & Hall.

Streiner, D. V. (2002). The case of the missing data: methods of dealing with dropouts and other research vagaries. *Research methods in Psychiatry, 47*(1), 68-75.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik.* Ankara: Seçkin Yayıncılık.

Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi (2*nd ed.*).* Ankara: Nobel Yayınları.

Wayman, J. C. (2003, April). *Multiple imputation for missing data: What is it and how can I use it?* Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL.

Young, W., Weckmen, G., & Holland, W. (2011). A survey of methodologies for the treatment of missing values within datasets: Limitations and benefits. *Theoretical Issues in Ergonomics Science, 12*(1), 15-43.

Please cite as:

Koçak, D., & Çokluk-Bökeoğlu, Ö. (2017). Analyzing the impacts of alternated number of iterations in multiple imputation method on explanatory factor analysis. *Eğitim Bilimleri Araştırmaları Dergisi - Journal of Educational Sciences Research, 7*(2), 95-111. http://ebad-jesr.com/