Investigating Measurement Invariance under Different Missing Value Reduction Methods

Hüseyin Selvi[1] ✉
Devrim Alıcı[2]
Nezaket Bilge Uzun[3]

[1]*Medical Education Department, Mersin University, Mersin, Turkey.*
Email: hsyn_selvi@yahoo.com.tr
[2,3]*Measurement and Evaluation Department, Mersin University, Mersin, Turkey.*

(✉ Corresponding Author)

Abstract

This study aims to comparatively examine the resultant findings by testing the measurement invariance with structural equation modeling in cases where the missing data is handled using the expectation-maximization (EM), regression imputation, and mean substitution methods in the complete data matrix and the 5% missing data matrix that is randomly obtained from the same matrix. The data were collected from 2822 students. Of these students, who participated in the study, 1338 (49.2%) were females while 1434 (50.8%) were males. The data were collected using the "School Attitude Scale" developed by Alıcı (2013). In this study, the measurement invariance was tested with structural equation modeling in the complete data matrix and in cases of handling the missing data it was tested using EM, Regression-Based Imputation, and Mean Substitution methods. In the present study the measurement invariance decisions taken for the complete data matrix coincide with the mean substitution method in all sub-factors, with regression imputation from the other suggested methods in two sub-factors, and with expectation-maximization in one sub-factor. It was concluded that different data imputation methods change measurement invariance decisions, considering all the factors. The three techniques provided consistent results only in one sub-dimension. The ways of dealing with the missing data change the results; thereby, increasing fundamental studies in this regard is necessary. With reference to these results, it would be worthwhile to study different missing data structures and different proportions of missing data in terms of invariance decisions.

Contents

Contribution of this paper to the literature
This study aims to comparatively examine the resultant findings by testing the measurement invariance with structural equation modeling in cases where the missing data is handled using the expectation-maximization (EM), regression imputation, and mean substitution methods in the complete data matrix and the 5% missing data matrix that is randomly obtained from the same matrix.

1. Introduction

In the sciences that study human behaviors and characteristics, the data collection and the analysis of the collected data are the most important stages of the research procedure. Data collection requires the development of suitable measurement tools according to the principles and methods of measurement and evaluation science and/or using the existing tools, whereas the data analysis requires the use of statistical methods according to the principles and methods of statistics that suit the purpose of the study and the structure of data. The validity of a scientific research depends upon the quality of the measurement tools used in data collection and the suitability of statistical analyses applied in the analysis of the collected data.

The primary purpose of data collection is to collect valid and reliable information by means of specific measurement tools from a group of participants representing the purpose of study to answer the research questions. A variety of factors are encountered in academic research that may negatively affect this process. Two main factors that may affect the data collection process are the missing data and measurement invariance, which is also the subject of this study to investigate. Missing data is a critical problem that is confronted in data collection process, especially in academic studies conducted in behavioral sciences. Missing data can occur for many different reasons. The most common reasons attributed to are participants negligently skipping an item unanswered; leaving the item unfilled to think about the response a little more and then forget to return to that item or refrain from responding; leaving the item unfilled because of insufficiency of time; avoiding to respond the item thinking that information required by the item is too personal; the inability of the machine to detect the response provided to the relevant item when using an optical reader; missing data in the whole measurement tool due to the missing sample in longitudinal studies or simultaneous utilization of more than one measurement tool, boredom, exhaustion, time waste, and the like are the factors causing the measurement tools not to be responded (Banks, 2015; De Luca & Peracchi, 2007; Field, 2005; Goegebeur, De Boeck, & Molenberghs, 2010; Heerwegh, 2005; Schafer & Graham, 2002; Widaman, 2006).

Rubin (1976) discussed the missing data under the name of "missing data mechanisms", as a process of dealing with the probability that each set of data is observed or missing. The missing data mechanisms are defined in three categories: missing completely at random (MCAR), missing at random (MAR), and missing at non-random (MNAR). When the missing data is completely at random (MCAR), the missing data does not exhibit a particular pattern in relation to any variable. In this case, occurrence of missing data is attributed to randomness. Random missing data (MAR) indicates the probability of missing data that arises is systematically related to the observed data. Put differently, in addition to the presence of a pattern in missing data this pattern does not have any effect on the primary dependent variable; probability of leaving the item unfilled is dependent on the performance of individuals. In case that missing data is nonrandom (MNAR), it means that there is a pattern in the missing data and this pattern influences the primary dependent variable. In other words, the probability of missing data that may occur in the item is related to the situation caused by the item itself, such as inaccuracy of the item and the bias it contains (Peng & Zhu, 2008). The type of missing data mechanism illustrates the suitability of different analysis methods (Molenberghs, Fitzmaurice, Kenward, Tsiatis, & Verbeke, 2015). Allison (2002) states that the missing data with MCAR or MAR structures are ignorable while the missing data with a MNAR structure is non-ignorable and by mentioning 'ignorable' she emphasizes that there is no need for the missing data to be modeled for the analyses yet to be done.

Depending on the pattern of missing data (MAR, MCAR, or MNAR), two basic approaches are used to deal with the missing data. The first one is the "deletion" approach, which means removing all the data from the matrix in a row or column that contains missing data. In the second approach, the imputation of a representing value is in focus, where a representative value is added to each missing value preserving the multivariate structure of relationships between the variables in the dataset (Widaman, 2006). To apply deletion and simple imputation methods, the structure of missing data should exhibit an MCAR structure (Allison, 2002; Alpar, 2011). Since the deletion methods cause a significant decrease in the number of observations, the calculated statistics can be biased as well as the reliability and generalizability of the study can be compromised. However, simple imputation methods are reported to result in biased parameter estimation and bias, erroneous hypothesis test results, substitution of the missing data with a value that does not meet the score range, and so forth (Allison, 2002; Alpar, 2011; Banks. & Walker, 2006; Lord, 1974). In addition to the deletion and simple imputation methods in the literature, the missing data problem can be often compensated through maximum likelihood and expectation-maximization methods (Allison, 2002; Alpar, 2011; Demir, 2013; Little & Rubin, 1987). Compared to the deletion and simple imputation methods, the most important advantage of these methods is that they can be used even if the missing data has a MAR structure (Allison, 2002; Alpar, 2011). However, generally, the proportion of missing data should not exceed a certain value for the missing data reduction to be applied. Schafer (1999) stated that this proportion should be less than 5%, Bennett (2001) 10%, and Peng., Harwell, Liou, and Ehman (2006) less than 20%. If not, they stated, the resultant findings of the study are probably biased.

Buuren (2012) argues that we ignore the causes of missing data while applying the missing data reduction methods and criticizes this as a major problem. The primary measure to take, according to the researcher, is to identify factors that may cause the emergence of missing data and look for how to prevent them. Thus, measures that can be taken during the data collection phase are: to encourage responding, to determine the data collection method (face-to-face, online, etc.) according to the sample structure, to complete the missing data from other sources for the non-responders where possible, to minimize the intensity of the study in experimental studies, to prepare a good leaflet that explains the purpose and benefit of the study, to perform a pilot study for identifying possible problems in advance, to economize the total number of variables, to collect information that is only

necessary for the study, to use short measurement tools where possible, to avoid item blocks that force staying to a certain page for a longer time, to use computer-based tests where possible, to respect item skip, and to create unanswered forms that provides an understanding of why they do not respond or why do they skip; which could be considered as seeking expert opinions (Buuren, 2012). However, missing data still arises, especially in the studies conducted in social sciences, and the presence of missing data is a problem for statistical analyses. The main reason behind this is that most of the statistics software and, particularly, multivariate statistical analyses work over the complete data matrix. Besides, the presence of many missing values in the data matrix is a major problem that threatens the validity and reliability of the results and if this proportion increases, it entails to an increase in Type I Error in statistical methods/techniques used, decrease in power of the test, erroneous estimation of standard error, inability to well-estimate the latent characteristics, and the like (Bernhard et al., 1998; Molenberghs & Kenward, 2007; Woodward, Smith, & Tunstall-pedoe, 1991). Tsiatis (2006) states that "When some of the data are missing, it may be that, depending on how and why they are missing, our ability to make an accurate inference may be compromised". Meanwhile, the missing data arise as a significant factor affecting the psychometric properties of measurement tools. There is a large number of studies in the existing literature that indicate items and test parameters are affected by missing data and missing data reduction methods (Akbas & Tavşancıl, 2015; Akın & Soysal, 2018; Cokluk & Kayri, 2011; Cüm, Demir, Gelbal, & Kışla, 2018; Cüm & Gelbal, 2015; Demir, 2013; Enders, 2004; Sahin Kürşad & Nartgün, 2015; Sayın, Yandı, & Oyar, 2017; Zhang & Yuan, 2016). In case of the missing data, one of the issues that must be addressed regarding the measurement tool is the measurement invariance. Measurement invariance refers to a psychometric characteristic that is not affected when individuals are placed in different sub-groups according to whatsoever characteristics, that is, the measurement tool can measure the desired characteristic independently afar from the influence of measurement done on different sub-groups. When the missing data cannot be prevented, attempts are made to overcome this problem by means of different missing data reduction methods under different missing data mechanisms and proportions. Differences in the proportion of missing data and missing data reduction methods are expected to have an impact on the invariance of measurement outcomes.

Measurement invariance studies in the literature are often conducted by using multi-group factor analysis techniques in structural equation modeling and differential test functioning through differential item functioning methods (Raju, Laffitte, & Byrne, 2002; Schmitt & Kuljanin, 2008; Stark, Chernyshenko, & Drasgow, 2006; Vandenberg & Lance, 2000; Xu & Tracey, 2017). Measurement invariance studies conducted with multi-group confirmatory factor analysis comprise of four steps that need to be tested: (1) configural invariance, (2) metric invariance, (3) scalar invariance, and (4) strict invariance (Meredith, 1993). In configural invariance, the factor structure of measurement tool; in metric invariance, the factor loadings related to the items forming the measurement tool; in scalar invariance, the regression constants created for the items forming the measurement tool, and in strict invariance, the estimated parameters concerning the suggested model, are all tested for the invariance concerning the measurements obtained from the sub-groups (Başusta & Gelbal, 2015; Uzun & Öğretmen, 2010). The hierarchy in the tests performed means securing the condition to the preceding invariance for each invariance test. In other words, when the condition for the preceding invariance is not met the subsequent invariance test cannot be performed. However, a partial invariance is assumed when the configural, the metric, or the scalar invariance is satisfied (Byrne, Shavelson, & Muthén, 1989; Byrne., 2006; Kline, 2011; Meredith, 1993).

Only two studies were found in the literature that addressed how the measurement invariance is affected in case of applying different missing data structures and/or different missing data reduction methods. In the first study, Tsai and Yang (2012) proposed the learning vector quantization estimated stratum weight (LVQ-ESW) method for determining measurement invariance in stratified sampling to deal with the missing group membership and complete the stratum weights, drawing on the effect of sample studies upon the statistical analyses. The study employed a multi-group confirmatory factor analysis (MG-CFA) and simulations to examine the accuracy and consistency of measurement invariance detection. Results from computer simulations demonstrated that the proposed method outstripped the conventional methods, such as case deletion, which accurately and consistently defines the measurement invariance. However, in another study, Işıkoğlu (2017) handled the missing data by applying the case deletion, series mean substitution, regression imputation, expectation-maximization, and the multi-group methods by forming different sample sizes and different proportions of missing data based on the data from 5496 students who participated in PISA 2015 and then examined the inter-gender measurement invariance by applying multi-group confirmatory factor analysis (MGCFA) method on the collected data. The findings revealed that the expectation-maximization method produced more consistent results at different sample sizes and missing data proportions than did the other methods.

Since there are a limited number of studies investigating the effects of missing data reduction methods on measurement invariance, it bears the necessity to conduct new studies in this regard. This study aims to comparatively examine the resultant findings by testing the measurement invariance with structural equation modeling in cases where the missing data is handled using the expectation-maximization (EM), regression imputation, and mean substitution methods in the complete data matrix and the 5% missing data matrix that is randomly obtained from the same matrix. The present study is assumed to make a significant contribution to the field in determining methods of handling missing data that could be more effective in examining the measurement invariance in different missing data structures.

2. Method

In this study, the findings of measurement invariance in a complete dataset, obtained by employing a scale, were mutually examined with the findings of measurement invariance obtained when the 5% missing data generated from the same dataset were handled using different methods. As fundamental research, it includes a methodical comparison, as it aims to investigate the effect of different methods of missing data reduction on measurement invariance and to offer new methodical knowledge to the literature.

2.1. Participants

The data were collected from 2822 students studying at different high schools. Of these students, who participated in the study, 1338 (49.2%) were females while 1434 (50.8%) were males.

2.2. Data Collection Tool

The data were collected using the "School Attitude Scale" developed by Alıcı (2013). The scale consists of 20 items of a 5-point Likert type for determining the attitudes of high school students towards the school. Of scale items, 12 were positive and 8 were negative. The negative items were scored in reverse. An increase in the scores acquired from the scale is considered to be an indicator of a positive attitude towards school but negative in case of a decrease. Both the findings of exploratory factor analysis performed using Promax rotation and the findings of confirmatory factor analysis performed on a different sample revealed that the scale was of a one-factor structure comprising of three components. The components have been labeled as "School as a Barrier to Personal Growth", "School as a Supporter to Personal Growth", and "School as a Missed Entity" respectively. The Cronbach alpha reliability for the entire scale was reported as 0.907 and for the sub-components as 0.871, 0.813, and 0.789 respectively.

2.3. Data Analysis

The inter-gender invariance of the model constructed under the explained construct of school attitude scale was investigated using a multi-group confirmatory factor analysis (MGCFA) method of structural equation modeling methods. Four different datasets, one of which was a complete data matrix, were utilized for this investigation. A dataset containing 5% missing data was created from the complete dataset in a completely random manner. When creating the missing data matrix, primarily all the items were added one after another to form a single column and as such a single column matrix consisting of 56440 rows in total was formed. After that, using the MEDCALC software, 2822 rows corresponding to 5% of 56440 rows were randomly deleted and a matrix containing missing data was created by moving the items back to their original locations.

A test of Little MCAR was performed to examine the resultant missing data structure and the missing data was observed to demonstrate an MCAR structure (Chi-Square = 7058.109, DF = 7006, Sig. = 0.328). The 5% missing data were handled by expectation-maximization (EM), regression assignment (RA), and mean substitution methods and measurement invariance studies were performed on both the complete data matrix and the matrices formed via three different methods.

Lisrel 8.7 software was utilized to analyze the data and so the multi-group confirmatory factor analysis (MGCFA) was performed. The MGCFA, a method which is commonly employed in group comparisons where there is more than one group, ensures that the means of latent factor are compared having the group parameters remain equal. The analysis performed on the means of latent factors in MGCFA is a sensitive technique compared to the traditional mean comparison and more accurately reveals the variation in different sub-groups (Thompson, 2004). Measurement invariance investigations are carried out in a hierarchical form with increased limitations. In this study, and as widely used and proposed, four different hypotheses were tested with configural invariance, metric invariance, scalar invariance, and strict invariance respectively. Instead of the chi-square statistics that produce more erroneous results due to statistical weaknesses arising from the sample size when comparing the goodness-of-fit in decision studies on measurement invariance, the difference values between the values of comparative fit indices (CFI) were used (Brown, 2006; Cheung & Rensvold, 2002; Wu, Zhen, & Zumbo, 2007). The differences between the CFI values for the hierarchically investigated invariance steps were examined in terms of "0.01>ΔCFI>-0.01"to ensure if the invariance conditions are met or not. In addition, CFI, $\chi2$, *df*, and Root Mean Square Error of Approximation (RMSEA) indexes were reported at the end of each test of invariance to learn about the fit indices. When the RMSEA values are less than 0.05 and the CFI values are higher than 0.90, they indicate a good fitting model (Cokluk, Sekercioğlu, & Büyüköztürk, 2010).

3. Results

The results of the measurement invariance for the complete data matrix are given below. The measurement invariance operations were separately examined for each component in the structural measurement model of the school attitude scale. The findings obtained for this table are elaborated in detail in terms of measurement invariance. However, since the main purpose of this study was to ensure if the decisions regarding the measurement invariance are accepted when different missing data imputation methods are applied, the findings related to other datasets were reported within the scope of the decision taken and the summary decision table were discussed in conjunction with other datasets in Table 1.

In the subgroups examined by gender in Table 1 the resultant fit indices of $\chi2$ /sd, RMSEA, NFI, CFI for the invariance conditions were within the acceptable ranges for the three subcomponents (Meredith, 1993). According to the results of measurement invariance for the subcomponents given in the table, the configural and metric invariance were achieved for the "School as a Missed Entity" component in the scale but the scalar invariance condition was "0.01<ΔCFI", that is, the scalar invariance condition was rejected due to the 'difference of regression constant for the groups'. Put differently, the measured factorial construct is similar in subgroups. The results indicate that the items making up the construct provide similar loadings to different subgroups. However, the relationship between the observed variables and the latent construct is not similar in different subgroups. The scores of the individuals with the same latent construct differ from the scores they obtain in relation to the observed constructs according to groups. Since the evidence for the scalar invariance was not secured, the strict invariance step, in which the error variances are also fixed for the groups, were not tested. Simply put, the studies of measurement invariance require conducting a hierarchic test of the most basic level of configural invariance and the comparison of the more restrictive model by developing hypotheses (Steenkamp & Baumgartner, 1998; Wu et al., 2007). Therefore, in these tests, which require hierarchy, it is not possible to pass to another step leaving the preceding step undone. It was concluded that, for the "School as a Supporter to Personal Growth" component, the

conceptual construct is invariant only for gender. In other words, for the "School as a Supporter to Personal Growth" component in the scale for gender, it can be said that conceptual construct is the same and the items under the components measure the same psychological construct (Horn & McArdle, 1992). Other invariance steps did not pass the invariance test except that; the metric, scalar, and strict invariance conditions did not satisfy this sub-factor. It could be concluded that there is a bias for this component in terms of the subgroups. Moreover, the configural and metric invariance were accepted for the "School a Barrier to Personal Growth" component as was for the "School as a Missed Entity" but not for the scalar and strict invariance, being rejected in invariance tests with respect to the constraints in these steps.

Table-1. The results of measurement invariance studied for the complete data matrix.

			χ2	df	RMSEA	NFI	CFI	ΔCFI
Complete Data	School as a Missed Entity	Model A (Configural Invariance)	11.96	4	0.075	0.99	0.99	–
		Model B (Metric Invariance)	17.72	8	0.058	0.99	0.99	0.00
		Model C (Scalar Invariance)	28.47	12	0.059	0.98	0.98	0.01
	School as a Supporter to	Model A (Configural Invariance)	371.68	40	0.12	0.95	0.96	–
		Model B (Metric Invariance)	390.21	48	0.11	0.95	0.95	0.01
	School as a Barrier to Personal Growth	Model A (Configural Invariance)	180.76	40	0.077	0.98	0.98	–
		Model B (Metric Invariance)	200.14	48	0.073	0.98	0.98	0.00
		Model C (Scalar Invariance)	267.55	56	0.078	0.97	0.97	0.01

After generating the missing data, the invariance test steps investigated for the datasets obtained by different methods of missing data imputation are presented together in Table 2.

Table-2. Measurement invariance results for the datasets obtained by different methods of missing data imputation.

			χ2	df	RMSEA	NFI	CFI	ΔCFI	
	School as a Missed Entity	Model A (Configural Invariance)	20.2	4	0.091	0.99	0.99	–	
		Model B (Metric Invariance)	26.64	8	0.07	0.98	0.99	0	
		Model C (Scalar Invariance)	37.28	12	0.067	0.98	0.98	0.01	
	School as a Supporter to Personal Growth	Model A (Configural Invariance)	401.04	40	0.13	0.95	0.95	–	
		Model B (Metric Invariance)	420.46	48	0.12	0.95	0.95	0	
		Model C (Scalar Invariance)	476.22	56	0.11	0.94	0.95	0	
		Model D (Strict Invariance)	540.74	71	0.11	0.93	0.94	0.01	
		Model A (Configural Invariance)	220.65	40	0.083	0.98	0.98	–	
		Model B (Metric Invariance)	271.31	48	0.086	0.97	0.97	0.01	
REGRESSION	School as a Missed Entity								
		Model A (Configural Invariance)	16.97	4	0.086	0.99	0.99	–	
		Model B (Metric Invariance)	23.49	8	0.067	0.98	0.99	0	
		Model C (Scalar Invariance)	31.89	12	0.063	0.98	0.98	0.01	
	School as a Supporter to Personal Growth	Model A (Configural Invariance)	418.26	40	0.12	0.96	0.96	–	
		Model B (Metric Invariance)	440.8	48	0.11	0.95	0.96	0	
		Model C (Scalar Invariance)	509.98	56	0.11	0.95	0.95	0.01	

		Model A (Configural Invariance)	147.6	40	0.074	0.98	0.98	-
	School as a Barrier to Personal Growth	Model B (Metric Invariance)	166.16	48	0.07	0.98	0.98	0
		Model C (Scalar Invariance)	227.13	56	0.074	0.97	0.97	0.01
	School as a Missed Entity							
		Model A (Configural Invariance)	26.44	4	0.084	0.99	0.99	-
		Model B (Metric Invariance)	33.1	8	0.066	0.98	0.99	0,00
		Model C (Scalar Invariance)	46.93	12	0.064	0.98	0.98	0.01
MEAN	School as a Supporter to Personal Growth	Model A (Configural Invariance)	378.41	40	0.11	0.96	0.96	-
		Model B (Metric Invariance)	402.55	48	0.1	0.96	0.95	0.01
	School as a Barrier to Personal Growth	Model A (Configural Invariance)	131.06	40	0.068	0.98	0.98	-
		Model B (Metric Invariance)	152.24	48	0.066	0.98	0.98	0
		Model C (Scalar Invariance)	219.56	56	0.072	0.97	0.97	0.01

As in Table 2 the measurement invariance hypotheses were tested for the data created by means of expectation-maximization (EM), Regression-based imputation, and mean substitution methods respectively.

The configural and metric invariance condition were met for the "School as a Missed Entity" component in the dataset created through the EM, but the scalar and strict invariance constraints were not and, thus, for these steps the hypotheses that the regression constant and error variances are the same in gender subgroups were rejected respectively. Moreover, the configural, metric, and scalar invariance were satisfied for the "School as a Supporter to Personal Growth" sub-factor but not for the strict invariance. With this finding, it was concluded that factor structure of the *support* factor in school attitude scale, the factor loadings, and regression constants are invariant and the error variances are different in gender subgroups. Further, for the "School as a Barrier to Personal Growth" component the hypothesis was only accepted in the most basic level of configural invariance step. However, since the metric invariance conditions were not satisfied and since the measurement invariance operations are tested hierarchically, the other invariance steps were rejected.

Table-3. The comparative decisions of invariance steps.

Factor	Invariance Hypotheses	Complete Data	Expectation Maximization	Regression	Mean
Feeling of Missing	Configural Invariance	Accepted	Accepted	Accepted	Accepted
	Metric Invariance	Accpeted	Accpeted	Accpeted	Accpeted
	Scalar Invariance	Rejected	Rejected	Rejected	Rejected
	Strict Invariance	Rejected	Rejected	Rejected	Rejected
Support	Configural Invariance	Accpeted	Accpeted	Accpeted	Accpeted
	Metric Invariance	Rejected	Accpeted	Accpeted	Rejected
	Scalar Invariance	Rejected	Accpeted	Rejected	Rejected
	Strict Invariance	Rejected	Rejected	Rejected	Rejected
Barrier	Configural Invariance	Accpeted	Accpeted	Accpeted	Accpeted
	Metric Invariance	Accpeted	Rejected	Accpeted	Accpeted
	Scalar Invariance	Rejected	Rejected	Rejected	Rejected
	Strict Invariance	Rejected	Rejected	Rejected	Rejected

When the fit indices of difference values between the configural and other invariance steps (ΔCFI), which is known as the basic level for all three components, were examined, it was concluded that the configural and metric invariance conditions were satisfied by regression-based imputation according to the missing data imputation

methods. Therefore, the hypotheses were accepted in that both the conceptual construct and the items had similar meanings to these groups in gender subgroups for all three components.

For the mean substitution, however, the hypothesis related to the configural and metric invariance conditions were accepted for the "School as a Missed Entity" and "School as a Barrier to Personal Growth" components. Besides, the configural invariance conditions were accepted for the "School as a Supporter to Personal Growth" component in the gender subgroup, where only the conceptual substructure of the factor was the same.

The decisions concerning the tested invariance steps for the complete data and missing data imputation methods are summarized in Table 3.

As indicated in Table 3 the same results were obtained for all three components in the complete dataset only by "imputing missing data with mean substitution" method. Based on the components, the invariance decisions concerning the whole datasets for the "School as a Missed Entity" component match up with the results produced by the complete data matrix, while there is variability in decisions on different steps of invariance for the *support* and *barrier* factors. The missing data imputation method, which indicated similarity with the decisions achieved for the complete data in "School as a Supporter to Personal Growth" component, is the mean substitution method; however, for the "School as a Barrier to Personal Growth" component the methods are "the mean substitution and regression-based imputation".

4. Discussion

In this study, the measurement invariance was tested with structural equation modeling in the complete data matrix and in cases of handling the missing data it was tested using EM, Regression-Based Imputation, and Mean Substitution methods.

To conclude, only in the first subscale it was determined that utilization of both the complete data matrix and the three different missing data handling methods yielded similar results and only the configural and metric invariance were satisfied. In case of handling the missing data with the expectation-maximization method in the "School as a Supporter to Personal Growth" component, it was observed that the scalar invariance was also satisfied; however, in cases of handling missing data with the mean substitution methods in the complete data matrix, only the configural invariance was satisfied. As such, the configural and metric invariance were satisfied when the missing data were handled with the regression method. Further, in the "School as a Barrier to Personal Growth" component, it was suggested that handling the missing data both with the regression-based imputation and the mean substitution methods may satisfy the configural and metric invariance but dealing with the missing data with the expectation-maximization method may only satisfy the configural invariance.

The findings show some similarities and differences when compared to Işıkoğlu (2017) study. Using different sample sizes, different missing data proportions that exhibit MCAR structure, and different methods of handling missing data in his study, Işıkoğlu (2017) revealed that determining the sample size as 2000 and the missing data as 5% ensures the inter-gender measurement invariance in all the methods of regression imputation, expectation-maximization, multiple imputation, case deletion, and series mean and, besides, the resultant fit indices from the regression imputation, expectation-maximization, and multiple imputation methods provide closer results to reference values. The researchers have suggested that datasets containing around 5% of missing data in large sample sizes can be completed by regression imputation, expectation-maximization, and multiple imputation methods. In the present study, however, the measurement invariance decisions taken for the complete data matrix coincide with the mean substitution method in all sub-factors, with regression imputation from the other suggested methods in two sub-factors, and with expectation-maximization in one sub-factor. It was concluded that different data imputation methods change measurement invariance decisions, considering all the factors. The three techniques provided consistent results only in one sub-dimension. The ways of dealing with the missing data change the results; thereby, increasing fundamental studies in this regard is necessary. With reference to these results, it would be worthwhile to study different missing data structures and different proportions of missing data in terms of invariance decisions.

This study was carried out only by randomly generating 5% missing data based on a real dataset and by using three different methods of handling missing data in the "School as a Supporter to Personal Growth" component. Even the dissimilarities of methods used in resolving the missing data problem may lead to different findings from the analyses (Harrington, 2009). Therefore, a similar study could be conducted using different proportions of missing data and/or different methods of handling missing data at different sample sizes.

The structural equation modeling and item response theory are the two approaches that are frequently used in measurement invariance studies. In this study, the measurement invariance was performed based on structural equation modeling by means of multi-group confirmatory factor analysis. A study, in which both the item response theory and the structural equation modeling are used together, could make significant contributions to the comparison of results.

References

Akbas, U., & Tavşancıl, E. (2015). Investigation of psychometeric properties of scales with techniques of handling missing data for different sample sizes and missing data Patterns in different sample sizes and lost data patterns. *Journal of Measurement and Evaluation in Education and Psychology, 6*(1), 38-57.

Akın, A. C., & Soysal, S. (2018). Investigation of reliability coefficients according to missing data imputation methods. *Hacettepe University Faculty of Education Journal, 33*(2), 316-336.

Alıcı, D. (2013). Developing an attitude scale towards school: Reliability and validity study [Development of an Attitude Scale towards School: A Study on Reliability and Validity]. *Education and Science, 38*(168), 319-331.

Allison, P. D. (2002). *Missing data.* California: Sage Publication, Inc.

Alpar, R. (2011). *Applied multivariate statistical methods [Practical Multivariate Statistics Methods]*: Detay Publishing.

Banks, K. (2015). An introduction to missing data in the context of differential item functioning. *Practical Assessment, Research, and Evaluation, 20*(1), 1-10.Available at: https://doi.org/10.7275/fpg0-5079.

Banks., K., & Walker, C. (2006). *Performance of SIBTEST when focal group examinees have missing data.* Paper presented at the Paper Presented at the Annual Meeting of the National Council of Measurement in Education, San Francisco, CA.

Başusta, N. B., & Gelbal, S. (2015). Examination of measurement invariance at groups' comparisons: A study on PISA student questionnaire. *Hacettepe University Journal of Education, 30*(4), 80–90.

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand journal of public health, 25*(5), 464–469.Available at: https://doi.org/10.1111/j.1467-842x.2001.tb00294.x.

Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., . . . Hürny, C. (1998). Missing quality of life data in cancer clinical trials: Serious problems and challenges. *Statistics in Medicine, 17*(5-7), 517-532.Available at: https://doi.org/10.1002/(sici)1097-0258(19980315/15)17:5/7<517::aid-sim799>3.0.co;2-s.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Buuren, v. S. (2012). *Flexible imputation of missing data*. Boca Raton: Chapman & Hall/CRC.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.Available at: https://doi.org/10.1037/0033-2909.105.3.456.

Byrne., B. M. (2006). *Structural equation with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associations.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.Available at: https://doi.org/10.1207/s15328007sem0902_5.

Cokluk, O., & Kayri, M. (2011). The effect of approximate value assignment methods on the validity and reliability of measurement tools. *Educational Sciences in Theory and Practice., 11*(1), 289 − 309.

Cokluk, O., Sekercioğlu, G., & Büyüköztürk, S. (2010). *Multivariate statistics in social sciences for social sciences*. Ankara: Pegem Publishing.

Cüm, S., Demir, E. K., Gelbal, S., & Kışla, T. (2018). Comparison of advanced methods used to assign approximate values instead of lost data under different conditions. *Mehmet Akif Ersoy University Journal of Education Faculty*(45), 230-249.Available at: https://doi.org/10.21764/maeuefd.332605.

Cüm, S., & Gelbal, S. (2015). The effect of different methods used in assigning approximate values instead of lost data on model data fit. *Mehmet Akif Ersoy University Journal of Education Faculty, 1*(35), 87-111.

De Luca, G., & Peracchi, F. (2007). A sample selection model for unit and item nonresponse in cross-sectional surveys. *CEIS Tor Vergata - Research Paper Series, 33*(99), 1-44.Available at: http://dx.doi.org/10.2139/ssrn.967391.

Demir, E. (2013). Estimation of item and test parameters in multiple choice tests in the presence of lost data: SBS example. *Journal of Educational Sciences Research, 3*(2), 47-68.Available at: https://doi.org/10.12973/jesr.2013.324a.

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement, 64*(3), 419-436.Available at: https://doi.org/10.1177/0013164403261050.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: SAGE Publications Inc.

Goegebeur, Y., De Boeck, P., & Molenberghs, G. (2010). Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical bayes estimates. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6*(1), 3-16.Available at: https://doi.org/10.1027/1614-2241/a000002.

Harrington, D. (2009). *Confirmatory factor analysis*. New York: Oxford University Press.

Heerwegh, D. (2005). *Web surveys explaining and reducing unit nonresponse, item nonresponse and partial nonresponse*. Doctoral thesis, Catholic University of Leuven, Faculty of Social Sciences, Leuven, Belçika.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*(3), 117-144.Available at: https://doi.org/10.1080/03610739208253916.

Işıkoğlu, M. A. (2017). *Comparison of influence of the missing data handling methods on measurement invariance*. Unpublished Master's Thesis. Hacettepe University, Department of Educational Science, Ankara.

Kline, R. B. (2011). *Principles and practices of structural equation modelling*. New York: The Guilford Press.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley & Sons, Inc.

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39*(2), 247-264.Available at: https://doi.org/10.1007/bf02291471.

Meredith, W. (1993). Measurement equivalence, factor analysis and factorial equivalence. *Psychometrika, 58*(4), 525-543.

Molenberghs, G., Fitzmaurice, G., Kenward, M., Tsiatis, A., & Verbeke, G. (2015). *Handbook of missing data methodology*. Boca Raton: CRC Press.

Molenberghs, G., & Kenward, M. G. (2007). *Missing data in clinical studies* (1st ed.). England: John Wiley & Sons.

Peng, C.-Y. J., & Zhu, J. (2008). Comparison of two approaches for handling missing covariates in logistic regression. *Educational and Psychological Measurement, 68*(1), 58-77.Available at: https://doi.org/10.1177/0013164407305582.

Peng., C. Y. J., Harwell, M., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), Real data analysis (pp. 31–78). Greenwich.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology, 87*(3), 517-529.Available at: https://doi.org/10.1037/0021-9010.87.3.517.

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

Sahin Kürşad, M., & Nartgün, Z. (2015). Comparison of various methods used in the context of validity and reliability of the scales [Comparison of Various Methods Used in the Context of Validity and Reliability of the Scales]. *Journal of Measurement and Evaluation in Education and Psychology, 6*(2), 254-267.

Sayın, A., Yandı, A., & Oyar, E. (2017). Investigation of the effects of coping with lost data on item parameters. *Journal of Measurement and Evaluation in Education and Psychology, 8*(4), 490 -510.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.Available at: https://doi.org/10.1037/1082-989x.7.2.147.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222.Available at: https://doi.org/10.1016/j.hrmr.2008.03.003.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292-1306.Available at: https://doi.org/10.1037/0021-9010.91.6.1292.

Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.Available at: https://doi.org/10.1086/209528.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Tsai, L. T., & Yang, C.-C. (2012). Improving measurement invariance assessments in survey research with missing data by novel artificial neural networks. *Expert Systems with Applications, 39*(12), 10456-10464.Available at: https://doi.org/10.1016/j.eswa.2012.02.048.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. New York: Springer.

Uzun, B., & Öğretmen, T. (2010). Evaluation of the measurement invariance of some variables related science achievement by gender in Turkey sample TIMSS-R. *Education and Science, 35*(155), 26-35.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.Available at: https://doi.org/10.1177/109442810031002.

Widaman, K. F. (2006). Missing data: What to do with orwithout them. *Monographs of the Society for Research in Child Development, 71*(3), 42–64.

Woodward, M., Smith, W., & Tunstall-pedoe, H. (1991). Bias from missing values: sex differences in implication of failed venepuncture for the Scottish Heart Health Study. *International Journal of Epidemiology, 20*(2), 379-383.Available at: https://doi.org/10.1093/ije/20.2.379.

Wu, A. D., Zhen, L., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research, and Evaluation, 12*(1), 1-26.Available at: https://doi.org/10.7275/mhqa-cd89.

Xu, H., & Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurenet invariance in counseling psychology research. *The European Journal of Counselling Psychology, 6*(1), 75-82.

Zhang, Z., & Yuan, K.-H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement, 76*(3), 387-411.Available at: https://doi.org/10.1177/0013164415594658.