ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2024; 15(2); 166-182

The Effects of Missing Data Handling Methods on Reliability Coefficients: A Monte Carlo Simulation Study

Tugay KAÇAK* Abdullah Faruk KILIÇ**

Abstract

This study holds significant implications as it examines the impact of different missing data handling methods on the internal consistency coefficients. Using Monte Carlo simulations, we manipulated the number of items, true reliability, sample size, missing data ratio, and mechanisms to compare the relative bias of reliability coefficients. The reliability coefficients under scrutiny in this study encompass Cronbach's Alpha, Heise & Bohrnsted's Omega, Hancock & Mueller's H, Gölbaşı-Şimşek & Noyan's Theta G, Armor's Theta, and Gilmer-Feldt coefficients. Our arsenal of techniques includes single imputation methods like zero, mean, median, and regression imputation, as well as multiple imputation approaches like expectation maximization and random forest. We also employ the classic deletion method known as listwise deletion. The findings suggest that, for missing completely at random (MCAR) or missing at random (MAR) data, single imputation approaches (excluding zero imputation) may still be preferable to expectation maximization and random forest imputation, thereby underscoring the importance of our research.

Keywords: missing data, reliability coefficients, missing data handling methods

Introduction

A common challenge in surveys and data collection processes is missing data, which can occur when respondents forget, skip, or choose not to answer one or more questions in a questionnaire. Especially for achievement tests or personality tests, sometimes test administrators suggest skipping the item if it is complex or confusing to respondents. Therefore, there are several reasons for missingness. Even the researchers do not care about the reason for missing data after the data collection process. However, the mechanism of missing data may affect the results of the analysis. There are three main mechanisms for missing data: i) Missing Completely at Random (MCAR), ii) Missing at Random (MAR), and iii) Missing Not at Random (MNAR) (Enders, 2010). So, mechanisms of missing data are the primary consideration when handling the missing data. In this study, we focused on MCAR and MAR because these mechanisms are more commonly encountered.

MCAR mainly points to randomness about missing data (Enders, 2010). As a formal definition, the probability of missing data about a variable is unrelated to other variables and itself. However, in MAR, the probability of missing data about a variable is related to other variables but unrelated to itself (Baraldi & Enders, 2010; Enders, 2010; Graham, 2012; Graham et al., 2013; Howell, 2007). The mechanisms of missing data may reduce the power of analysis or alter the distribution of the data set, which is another detail that can affect the results of statistical inferences. Even after adjusting for other factors, the likelihood of missing data on a variable is correlated with the values of that variable, which results in missing, not at random (MNAR). The degree of a patient's illness, for instance, may determine how likely they are to drop out of a clinical experiment, therefore affecting the missing data (Enders, 2010; Howell, 2007). MNAR presents substantial difficulties because it relies on unobserved values. Selection models and pattern mixture models provide methods for addressing missing not-at-random (MNAR)

*Research Assistant, Trakya University, Faculty of Education, Edirne-Türkiye, kacaktugay@gmail.com, ORCID ID: 0000-0002-5319-7148
**Associate Professor, Trakya University, Faculty of Education, Edirne-Türkiye, abdullahfarukkilic@gmail.com, ORCID ID: 0000-0003-3129-1763

To cite this article:
Kaçak, T., & Kılıç, A. F. (2024). The effects of missing data handling methods on reliability coefficients: A Monte Carlo simulation study. *Journal of Measurement and Evaluation in Education and Psychology, 15*(2), 166-182. https://doi.org/10.21031/epod.1485482

Received: 17.05.2024
Accepted: 21.06.2024

data, but they necessitate stringent assumptions. Conducting sensitivity analysis and actively collecting data are essential for dealing with complications arising from missing not-at-random (MNAR) data. So, we excluded the MNAR mechanism from the simulation study.

Many statistical programs offer limited methods to handle missing data. The most popular one is listwise deletion (LD). LD is an essential way to get a complete case by excluding the observations with empty cells. Complete case analysis is the primary choice. There are two common imputation approaches to obtain complete case: single imputation methods and multiple imputation methods. Mean, median, and regression imputation may be given as examples for single imputation (SI) methods. Random Forest imputation, classification and random trees, expectation-maximization may be given as examples for multiple imputation (MI) methods. It is known that imputation methods affect SEM fits (Fan & Wu, 2022; Li & Lomax, 2017), model parameters (T. Dai et al., 2024), growth curve parameters (D. Y. Lee et al., 2019), performances of factor retention methods (Goretzko, 2021; Goretzko et al., 2020), MANOVA results (Finch, 2016), the Rasch model statistics, Mokken's scalability coefficient H, Cronbach's Alpha (Roth et al., 1999; Sijtsma & Van Der Ark, 2003; Van Ginkel et al., 2007), and item parameters in IRT (S. Dai, 2021). However, there are limited studies focused on reliability coefficients except for Cronbach's Alpha. Therefore, in this study, we aim to compare the performance of reliability coefficients like Heise & Bohrnstedt's Omega, Hancock & Mueller's H, Gölbaşı-Şimşek & Noyan's Theta G and Armor's Theta, and also Cronbach's Alpha in terms of handling missing data methods. The present research was expected to elucidate the impact of handling missing data approaches on reliability coefficients.

How to Handle Missing Data?

Deletion methods can be performed either at the row or column level. Listwise deletion involves excluding individuals with missing data from the analysis, while column deletion entails not including variables containing missing data in the analyses (Enders, 2010; Little & Rubin, 2019; Schafer & Graham, 2002; Scheffer, 2002). While these approaches offer practical solutions to researchers, listwise deletion may introduce bias in analysis results due to sample reduction, and column-wise deletion may have a diminishing effect on content and construct validity. Therefore, contemporary approaches are progressing towards preserving both rows/lists (individuals) and columns (variables), utilizing techniques such as regression imputation between cells, expectation maximization (EM) for imputation, and considering variables' interrelationships, similarities among individuals in terms of measured characteristics, and various other statistical methods (Scheffer, 2002). All this literature raises this question: *"Which method for handling missing data provides more unbiased results for my survey?"*

In this study, we compared a deletion method (Listwise Deletion), single imputation methods ($\text{Mean}_{\text{Person}}$, $\text{Median}_{\text{Item}}$, Zero, Regression Imputation), and multiple imputation methods (Expectation Maximization and Random Forest Imputation). Deletion and single imputation methods were chosen because they are the default methods in most statistical software. Multiple imputation methods were also chosen because they are reported to give unbiased results in the current literature (Leite & Beretvas, 2010).

Listwise Deletion

The extent to which listwise deletion will cause problems in the data analysis process depends on the missing data mechanism. However, it is known to cause power loss of analysis in studies (Myers, 2011; Newman, 2014).

Mean Imputation

The mean imputation method (ME) involves replacing missing data in a row with the arithmetic mean of the non-missing values in that row or cell (Enders, 2010). The mean can be imputed by row or by column. In the case of categorical data, the row mean may be applicable in a continuous structure. In such cases, categorical observed variables should be treated as continuous, and analyses should be conducted accordingly. In this study, we used listwise mean imputation, which is actually named $\text{Mean}_{\text{Person}}$.

Median Imputation
Median imputation (MD) involves replacing missing data in a cell with the median of the values in the row or column where the missing data occurs (Zhang, 2016). In the case of categorical data, if the number of rows or columns is even, the imputed data may be treated as continuous. Therefore, observed variables should be considered continuous, and analyses should be conducted accordingly. In this study, we used column-wise median imputation which names as $\text{Median}_{\text{Item}}$.

Zero Imputation
Zero imputation involves imputing value of "0" to the cell with missing data (Wei et al., 2018). Clustering responses to variables to a single value can affect the variance, skewness, and kurtosis of the variable.

Regression Imputation
Regression imputation (RI) involves assuming linear relationships between variables and creating a regression equation. Using this equation, it predicts and imputes a value for missing cells (Enders, 2010). The predicted value is typically continuous.

Expectation Maximization Algorithm
Expectation Maximization (EM) algorithm was developed by Enders (2003) to mitigate information loss due to missing data. Essentially, it is an iterative method that performs imputation of the missing values. In the Expectation step, a series of regression equations are constructed to establish relationships between the variable with missing data and other variables. This process helps in developing estimated values for the empty cells. In the Maximization step, the covariance matrix is computed to minimize the residual variances after imputation. Iterations are repeated using the Maximum Likelihood method to minimize these residual variances each time. It relies on the assumption of multivariate normality of variables (Allison, 2002; Little & Rubin, 2002). This assumption is robust against some violations of categorical data. It is recognized that predictions may exhibit bias when dealing with Missing Not at Random (MNAR) as a missing data mechanism. Schafer and Graham's (2002) and Little and Rubin's (2002) study can be reviewed for further details.

Random Forest Imputation
Random Forest (RF) imputation relies on multiple regression and aims to impute missing data by creating a decision tree mechanism (Shah et al., 2014). It aims to predict missing cells by sampling randomly from the dataset. The Random Forest imputation method can be used for all types of variables, and unlike the EM method, it does not assume multivariate normality. It is known to perform well even for non-normally distributed datasets, and for more detailed information, studies by Doove et al. (2014) and T. Hayes & McArdle (2017) can be consulted.

Various studies have examined the effects of different approaches to handling missing data on various aspects of item response theory, such as item parameters (Finch, 2008), item difficulty and discrimination (Béland et al., 2018), methods for determining the number of factors (Goretzko, 2021), factor number and factor loadings (McNeish, 2017), results of confirmatory factor analysis (Lei & Shiverdecker, 2020), and reliability coefficients (Enders, 2004). This study aims to investigate the influence of different approaches to handling missing data on the reliability coefficients.

Reliability Coefficients
Reliability is one of the fundamental psychometric properties that need to be reported (Nunnally & Bernstein, 1994). While there are many different approaches to the concept of reliability, this study focuses on reliability coefficients in terms of internal consistency. Given the different assumptions they rely on, determining which reliability coefficient to report is essential for the reliability of research results. Among reliability coefficients, Cronbach's Alpha coefficient (Cronbach, 1951) is the most commonly reported in many social science research studies (Dunn et al., 2014; McNeish, 2018). However, Cronbach's Alpha coefficient is often reported without fulfilling its assumptions (Dunn et al., 2014). Despite many criticisms of using the Alpha coefficient, it continues to dominate among reported reliability coefficients (Edwards et al., 2021).

In order to properly utilize Cronbach's Alpha coefficient, many assumptions must be satisfied (Cronbach, 1951; A. Hayes & Coutts, 2020; McNeish, 2018). These are i) the presence of unidimensionality, meaning that the items being measured are all connected to the same underlying construct; ii) the equality of factor loadings, also known as tau-equivalence, which implies that all items have equal relationships with the underlying construct; iii) the usage of continuous variables that follow a normal distribution; and iv) the assumption that error terms are uncorrelated. In order to calculate Alpha, it is essential to provide evidence of construct validity for unidimensionality, considering that psychological attributes are often multidimensional (McNeish, 2018). Furthermore, the equivalence of taus, or the parallelism of items in the scale, is linked to the degree to which items equally account for the measured attribute. However, in practice, it is more typical to have congeneric measurements, which means that the factor loadings are not equal. Research has shown that Alpha tends to underestimate the reliability of congeneric measures, as demonstrated by Edwards et al. (2021) and McNeish (2018). Other coefficients based on different assumptions than Alpha coefficient have also been developed, such as McDonald's Omega (McDonald, 1999), Heise & Bohrnstedt's Omega (Heise & Bohrnstedt, 1970), Hancock & Mueller's H (Hancock & Mueller, 2001), Armor's Theta (Armor, 1974), Gölbaşı-Şimşek & Noyan's Theta G (Gölbaşı-Şimşek & Noyan, 2013), and Gilmer-Feldt reliability coefficient (Feldt & Charter, 2003). When calculating reliability coefficients, estimations such as covariance matrices, variances, or factor loadings can be used. We compared the performance of Heise & Bohrnstedt's Omega, Hancock & Mueller's H, Gilmer-Feldt, Armor's Theta, Gölbaşı-Şimşek & Noyan's Theta G, and Gilmer-Feld coefficients in the current study. Table 1 provides the equations used to calculate the reliability coefficients examined in this study.

In order to obtain reliability coefficients in the presence of missing data, it is necessary to handle missing values in the cells where they occur through deletion or imputation approaches. Each missing data handling method may affect the values and parameters used to calculate reliability coefficients. The examination of how missing data handling methods affect reliability coefficients in datasets with missing data is becoming more critical and significant for researchers.

Table 1.
Reliability Coefficients

Coefficients	Formula	Variables
Cronbach's Alpha (Cronbach, 1951)	$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_X^2}\right)$	k: number of variables, s_i^2: variance of each variable s_X^2 : sum of variances of variables
Heise & Bohrnstedt's Omega (Heise & Bohrnstedt, 1970)	$\Omega = 1 - \frac{\sum_{i=1}^{n} \sigma_i^2 - \sum_{i=1}^{n} \sigma_i^2 h_i^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} Cov(x_i, x_j)}$	h_i^2 : communalities x : variable
Hancock & Mueller's H (Hancock & Mueller, 2001)	$H = \frac{\sum_{i=1}^{k} \frac{l_i^2}{1-l_i^2}}{1+\sum_{i=1}^{k} \frac{l_i^2}{1-l_i^2}}$	k : number of variables, l_i : i[th] items' standardized factor loading
Gölbaşı-Şimşek & Noyan's Teta G (Gölbaşı-Şimşek & Noyan, 2013)	$\theta_G = \frac{k}{k-m}\left(\frac{\sum_{i=1}^{m} \gamma_i - m}{\sum_{i=1}^{m} \gamma_i}\right)$	m : number of factors k : number of variables γ_i : i[th] eigenvalue
Armor's Teta (Armor, 1974)	$\theta = \left[\frac{p}{p-1}\right]\left[1 - \left(\frac{1}{\lambda_i}\right)\right]$	λ_i : the largest eigenvalue obtained by principal component analysis, p : number of variables
Gilmer-Feldt's coefficient (Feldt & Charter, 2003)	$r = \left[\frac{Q}{Q-W}\right](T/S_{tot}^2)$	S_{tot}^2 : variance of total scores T : represents the sum of the last columns of the covariance matrix. D should be calculated for Q and W. For D, the largest of the row sums of the covariance matrix is determined. The elements in this row are subtracted from the current row (e.g., ∑1 - A) and the largest row sum (∑h-A). D values are obtained by performing the process for each row. Q is calculated as the square of the sum of D's. W is calculated as the sum of the squares of each D.

Effects of Missing Data on Reliability Coefficients

Most of the previous research on reliability coefficients was conducted with complete data. Enders' (2003) work considered incomplete data in 10 and 20 variables with unidimensional model. Factor loadings were λ = 0.55 and λ = 0.75. Data sets were generated with factor loadings, and the true reliability was calculated. Variables follow a normal distribution and are also categorical (3, 5, and 7). Missing mechanisms were used in three types (MCAR, MAR, and MNAR). Missing data ratios were at two levels (%15 and %30). EM, LD, PD, $Mean_{Column}$, and $Mean_{Person}$ were used to handle missing data. Only Cronbach's Alpha was analyzed in this study. Findings show that EM outperformed the other imputation methods, and LD and $Mean_{Column}$ may cause an underestimate of Cronbach's Alpha. Also, EM, LD, and Pairwise Deletion (PD) produced relatively high coverage rates.

In another study, Zhang & Yuan's (2016) work, Cronbach's Alpha and McDonald's Omega, were compared in terms of missing and outlying data. Listwise deletion and ML methods are used to handle missing data when tau equivalence is violated/is not violated. Findings show that listwise deletion causes the underestimate of Cronbach's Alpha and McDonald's Omega.

Also, Enders' (2004) study shows that EM is the best method to handle missing data for both missing data mechanisms (MCAR and MAR). The $Mean_{Person}$ method was the most negatively biased. The missing data ratio was 20%, and only Cronbach's Alpha was investigated.

In Sijtsma & Van Der Ark's (2003) study, predictive mean matching (PMM), two-way imputation (TW - $Mean_{Colum}$ and $Mean_{Person}$), response function (RF), and model-based response function (MRF) were used. Overall, PMM, TW, RF, and MRF resulted in slight biases on Cronbach's Alpha. The impact of imputation methods was not substantial.

Also, in the literature, there are several studies about the effects of missing data on reliability coefficients in real datasets (Parent, 2013; Şahin Kürşad & Nartgün, 2015). In real datasets, true reliability cannot be known; it can just be estimated. Therefore, this study adopts a simulation approach. Furthermore, this study is deemed significant for the following reasons: i) examining the impact of standard missing data handling methods found in statistical software on reliability coefficients, ii) investigating the effects of more contemporary and robust methods such as random forest, EM, as indicated by previous studies (T.

Hayes & McArdle, 2017; McNeish, 2017), alongside more straightforward methods on reliability coefficients, and iii) considering that the field of education primarily deals with categorical data, exploring the use of categorical variables and providing recommendations to researchers on which reliability coefficient to prefer in the presence of missing data.

The problem statement within the scope of the research is as follows: "In the examined simulation conditions, to what extent are the relative bias values of reliability coefficients:

- differ in complete datasets?
- differ in datasets with missing data?"

Method

This study, conducted to determine how reliability coefficients are affected by different missing data handling methods, is a Monte Carlo simulation. Monte Carlo simulation studies involve generating data according to a specific distribution, analyzing the generated data using different methods, and comparing the results (Sigal & Chalmers, 2016).

Simulation Conditions and Data Generation

This study investigates how reliability coefficients change according to missing data handling methods based on the number of items per factor, sample size, reliability level, missing data ratio, and missing data mechanism. The simulation conditions are presented in Table 2.

Table 2.
Examined Conditions

	Simulation Factors	Simulation Conditions	Number of Conditions
Datasets with missing data	Sample size	200, 1000	2
	Missing data ratio	5%, 10%, 20%	3
	Missing data mechanism	MCAR, MAR	2
	Test length	8, 16	2
	True reliability	0.70, 0.90	2
			2x3x2x2x2 = 48
			1000 replications
Datasets without missing data	Sample size	200, 1000	2
	Test length	8, 16	2
	True reliability	0.70, 0.90	2
			2x2x2=8
			Total 56 conditions
			1000 replications

In addition to Table 2, eight conditions have been included for datasets without missing data, resulting in a total of 56 conditions. For each condition, 1000 replications were performed. Thus, 56,000 datasets were generated, and estimates were obtained for six different reliability coefficients from each dataset. Consequently, we conducted 336,000 different analyses. The fixed conditions of the study include a unidimensional structure and variables with five categories. Continuous datasets generated to exhibit multivariate normal distribution were transformed into categorical form using cutoff points as utilized by Uysal & Kılıç (2022)

Sample sizes of 200 and 1000 were added to the simulation conditions to represent small and large samples, respectively. These sample sizes are commonly preferred in many Monte Carlo simulation studies (Beauducel & Herzberg, 2006; Enders, 2004), facilitating the comparability of analysis results.

In this study, missing data rates of 5%, 10%, and 20% were considered. These rates have been examined in various studies on missing data (Cheema, 2014; H. J. Lee & Huber, 2021). It is noted that rates of 10% and above may lead to bias in the analyses (Bennett, 2001). Accordingly, the 5% rate represents a small proportion of missing data, the 10% rate represents the threshold where analyses may exhibit bias, and the 20% rate represents the expected level of missing data that may introduce bias in the analysis results.

The missing data mechanisms considered in the study are Missing Completely at Random (MCAR) and Missing at Random (MAR). In MCAR, missing data for a variable are unrelated to other variables (Allison, 2002; Enders, 2010; Graham, 2012; Little & Rubin, 2019). In MAR, there exists a relationship between the variable with missing data and other variables (Allison, 2002; Little & Rubin, 2019; Schafer & Graham, 2002). To ensure comparability with other studies in the literature, the MAR mechanism is included in this study.

Another condition included in the study is the number of items per factor. Brown (2006) suggests that there should be a minimum of 3 items per factor. However, there are studies suggesting that there should be at least 5 (Gorsuch, 2015) or at least 10 (Nunnally & Bernstein, 1994) items per factor. Since unidimensional structures are considered, the number of items per factor is set to 8 and 16 in this study. Thus, most of the recommendations regarding the number of items per factor in the literature are met. Additionally, this enables the examination of the impact of missing data handling methods on reliability coefficients in relatively short tests.

The reliability level represents the reliability coefficient in terms of internal consistency. In the literature, reliability coefficients of 0.70 and above are considered acceptable (McAllister & Bigley, 2002; Nunnally & Bernstein, 1994). Therefore, in this study, the true reliability conditions include the acceptable lower limit of 0.70 and the condition of 0.90, which can be interpreted as high reliability. The specified reliability levels, as also included in the study by Edwards et al. (2021), were obtained using McDonald's Omega coefficient (McDonald, 1970, 1999) with Formula 1:

$$\omega = r_{xx'} = \frac{(\sum_{i=1}^{k}\lambda_i)^2}{(\sum_{i=1}^{k}\lambda_i)^2 + \sum_{i=1}^{k}(1-\lambda_i^{\,2})} \qquad (1)$$

In Formula 1, λ represents the factor loading of each item.

Evaluation Criteria

The relative bias (RB) values, which are based on the difference between the calculated reliability coefficients for each condition and the true reliability level, were calculated using Formula 2:

$$RB = \frac{\bar{r}_{xx'} - r_{xx'}}{r_{xx'}} \qquad (2)$$

Here, $\bar{r}_{xx'}$ represents the average reliability coefficient obtained from 1000 replications, while $r_{xx'}$ represents the true reliability level. As recommended by Flora & Curran (2004), we employed a criterion of $|RB| < 0.10$ to determine acceptable bias.

Generating Datasets and Imputing Process

The process of creating datasets with missing data was carried out using the "mice" package (van Buuren & Groothuis-Oudshoorn, 2011) based on the specified missing data ratios (5%, 10%, 20%) and missing data mechanisms (MCAR and MAR). After generating datasets with missing data, we employed the following imputation methods using relevant R packages. We employed the imputeTS package (Moritz & Bartz-Beielstein, 2017) for listwise deletion and zero imputation and the missMethods package (Rockel, 2022) for median imputation, mean imputation, and EM imputation. We employed the mice package (van Buuren & Groothuis-Oudshoorn, 2011) for random forest (RF) imputation with the default iteration setting as m = 5 and regression imputation.

The analysis of the data involved calculating Cronbach's Alpha coefficient for the complete datasets and datasets obtained through various missing data imputation methods using the psych package (Revelle, 2024). Additionally, we utilized the reliacoef package (Cho, 2023) to calculate Heise & Bohrnstedt's Omega, Gilmer-Field's reliability coefficient, and Hancock & Mueller's H coefficient. We wrote the custom R script to calculate Gölbaşı-Şimşek & Noyan's Teta G and Armor's Teta coefficients.

We utilized Pearson correlation matrices to calculate reliability coefficients from factor analysis for datasets imputed using mean, median and regression imputation methods, which resulted in continuous data.

Findings

The one-way ANOVA results indicate that each simulation factor has a statistically significant effect on the RB values. The findings are summarized in Table 3.

Table 3.
One-way ANOVA Results

	Factor	df	*F*	η^2
Datasets with missing data	True reliability	1	97.39*	0.05
	Test length	1	284.42**	0.12
	Missing data mechanism and ratio	5	23.84*	0.06
	Deletion/imputation methods	6	76.06**	0.19
	Sample size	1	89.91*	0.04
	Reliability coefficient	5	455.70***	0.53
Datasets without missing data	True reliability	1	0.519	0.01
	Test length	1	4.457**	0.10
	Sample size	1	2.195	0.05
	Reliability coefficient	5	12.606***	0.62

*Note. *p<0.05, **p<0.01,***p<0.001*

Upon examining Table 3, it is evident that the reliability coefficients have the most significant impact on RB values for datasets containing missing data, while the sample size has the most minor effect. For datasets without missing data, the most significant impact on RB values is associated with reliability estimation methods, with no significant effect observed for reliability level and sample size.

Figures 1-4 present the RB values obtained from simulation conditions. The parallel lines on the x-axis represent the acceptable range of |RB| as ±0.10.

Figure 1.
Relative Bias Values for Datasets Without Missing Data

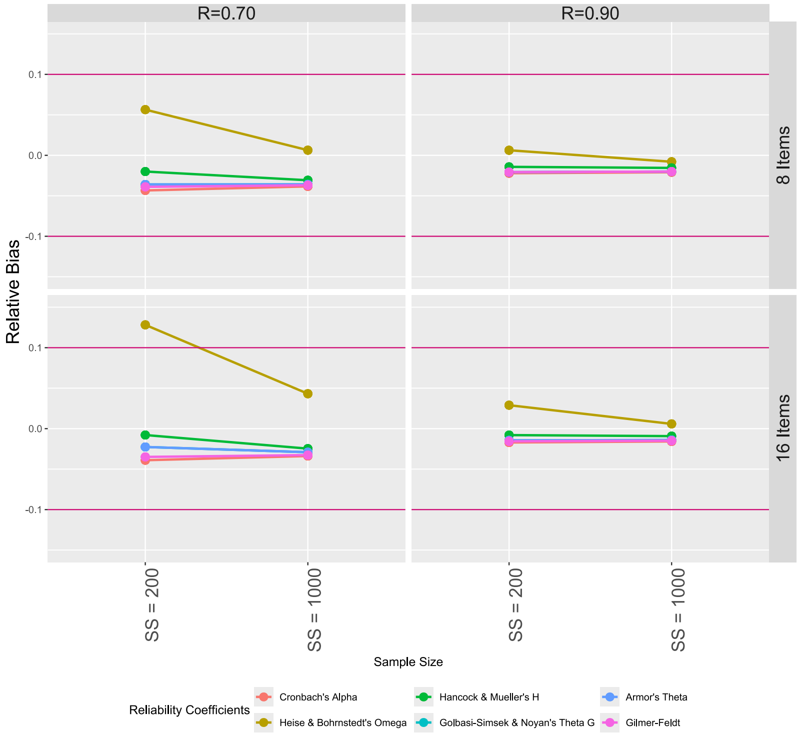


Figure 1 displays the relative bias (RB) values of reliability coefficients for datasets without missing data. It is observed that in the condition where the sample size is 200, the number of items is 16, and the reliability (R) is 0.70, Heise & Bohrnstedt's Omega coefficient overestimated reliability. However, all other reliability coefficients provided acceptable estimations across all conditions.

In Figure 2, the RB values of reliability coefficients are presented for datasets with a 5% missing data rate. Similar to datasets without missing data, in conditions where the sample size is 200, the number of items is 16, and reliability (R) is 0.70, Heise & Bohrnstedt's Omega coefficient overestimated reliability independent of the missing data mechanism and imputation method. However, all other reliability coefficients, except Heise & Bohrnstedt's Omega, provided reliability estimates within an acceptable RB range across all conditions, regardless of the number of items, sample size, imputation method, and missing data mechanism.

In Figure 3, the RB values of reliability coefficients are presented for datasets with a 10% missing data rate. Similar to datasets with a 5% missing data rate and datasets without missing data, Heise & Bohrnstedt's Omega coefficient overestimated reliability when the sample size was 200, the number of items was 16, and reliability (R) was 0.70. However, under the MAR missing data mechanism, when the true reliability was 0.70, and the number of items was 8, Cronbach's Alpha, Gilmer-Feldt's reliability coefficient, Armor's Theta, Gölbaşı-Şimşek & Noyan's Theta G, and Hancock & Mueller's H coefficient underestimated the true reliability when zero imputation was preferred. In all other conditions, all reliability coefficients (except Heise & Bohrnstedt's Omega when the number of items was 16, the sample size was 200, and R was 0.70) provided reliability estimates within an acceptable RB range.

In Figure 4, the RB values of reliability coefficients are presented for datasets with a 20% missing data rate. Under the MAR missing data mechanism, when the true reliability was 0.70, the number of items was 8, and the sample size was 200, Cronbach's Alpha, Gilmer-Feldt's reliability coefficient, Armor's Theta, and Gölbaşı-Şimşek & Noyan's Theta G underestimated the true reliability when zero imputation was preferred. Similarly, under the MAR missing data mechanism, when the true reliability was 0.70, the number of items was 16. The sample size was 1000, Cronbach's Alpha, Gilmer-Feldt's reliability coefficient, Armor's Theta, Gölbaşı-Şimşek & Noyan's Theta G, and Hancock & Mueller's H coefficient underestimated the true reliability when zero imputation was preferred. In all other conditions, all reliability coefficients (except Heise & Bohrnstedt's Omega when the number of items was 16, the sample size was 200, and R was 0.70) provided reliability estimates within an acceptable RB range.

Figure 2.
Relative Bias Values for Datasets with %5 Missing

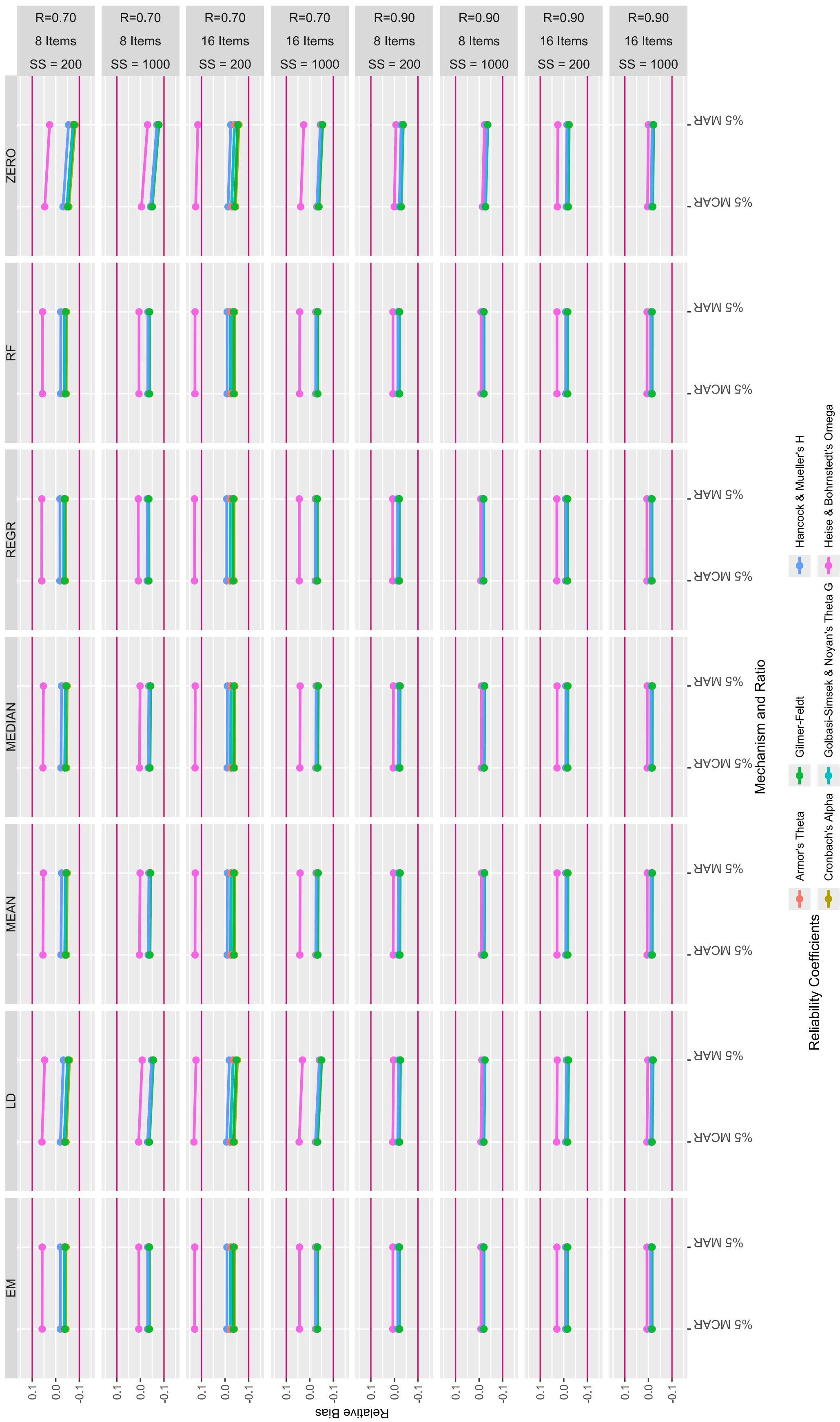


Figure 3.
Relative Bias Values for Datasets with %10 Missing

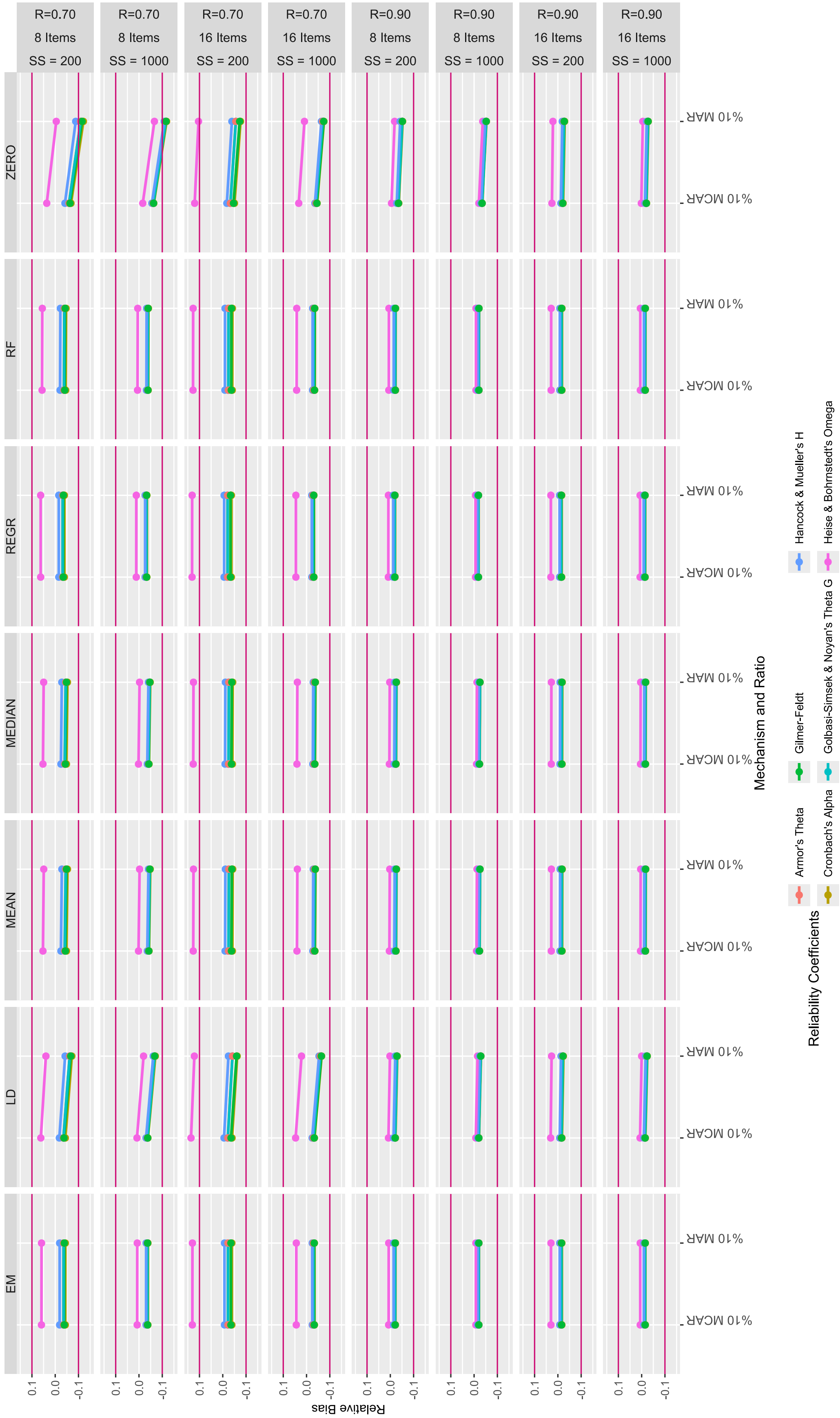
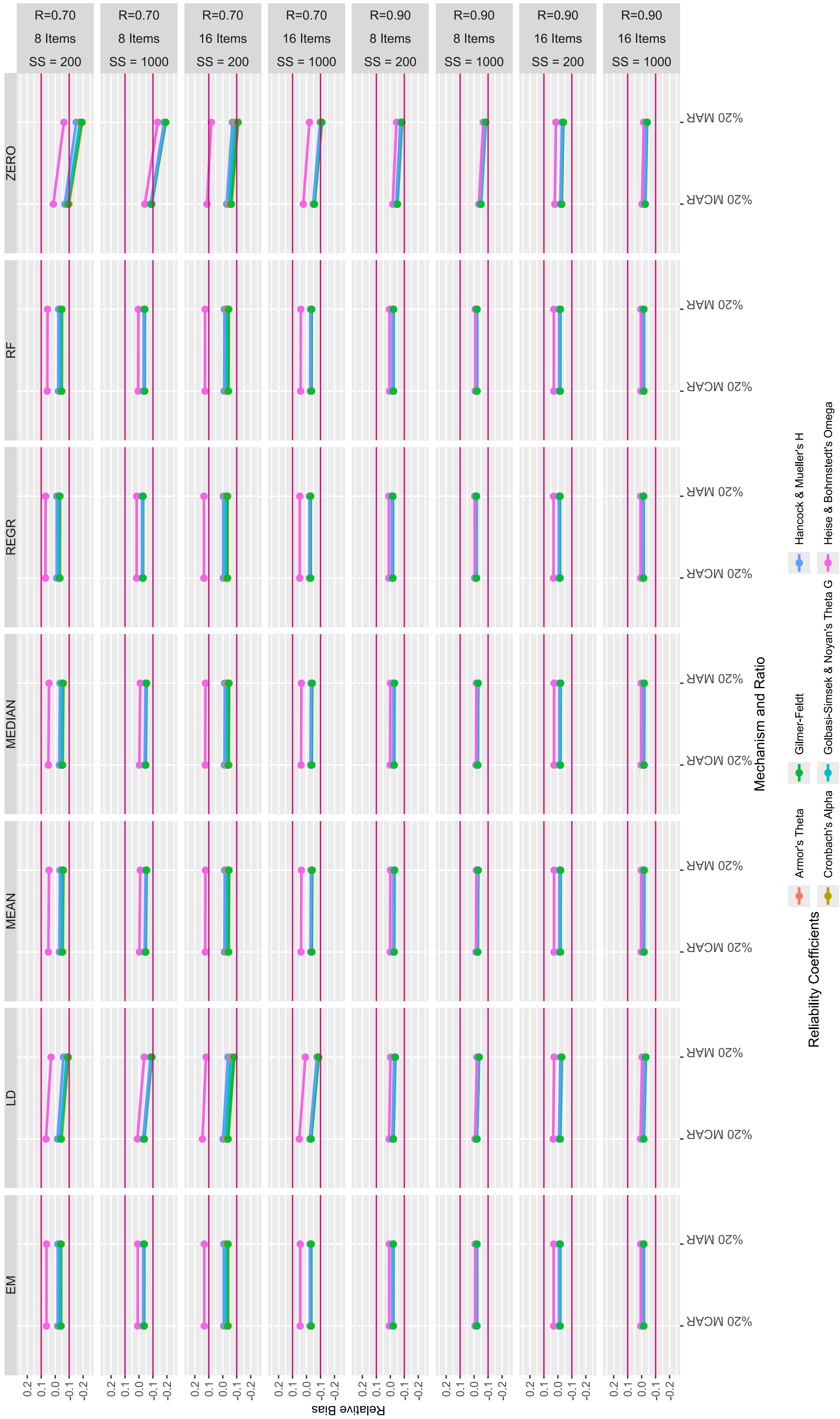


Figure 4.
Relative Bias Values for Datasets with %20 Missing



Upon examination of Figures 1, 2, 3, and 4, it was observed that Heise & Bohrnstedt's Omega coefficient tends to overestimate reliability, especially in conditions of low reliability (0.70), small sample size (200), and 16 items, regardless of missing data and imputation methods. Therefore, under these specific conditions, Heise & Bohrnstedt's Omega coefficient can be considered an upper limit for estimating reliability.

Using the zero-imputation method to handle missing data can yield misleading results, particularly when the missing data mechanism is MAR. When preferred alongside commonly used imputation methods such as mean imputation, median imputation, and regression imputation, as well as Random Forests imputation and expectation maximization methods, reliability coefficients generally provide estimations within acceptable limits ($|RB|<0.10$).

Discussion

This study compares the relative bias values of reliability coefficients in unidimensional structures consisting of five-category ordinal indicators regarding the impact of missing data handling methods. Findings show that zero imputation method may cause to underestimation of reliability coefficients if missing data mechanism is MAR and when missing data ratio increases.

When the sample size gets lower for the complete data sets obtained with the listwise deletion method, it is seen that the reliability coefficients do not give biased results. This is in line with Enders' (2003) findings that when LD is preferred as a method of dealing with missing data in data sets with 15% missing data. Similar to Enders' (2003) study, the level of biased estimation for Cronbach's Alpha coefficient was within acceptable limits, and the bias decreased as the sample size increased. In the MAR and MCAR mechanism, the BM algorithm estimated Cronbach's Alpha coefficient with acceptable bias, similar to Enders' (2004) study. For the MCAR mechanism, LD, RF, ME, MD, RI, and EM algorithms gave unbiased results, which are theoretically and practically compatible. Analogous to these findings, in their study, Sheng and Shen (2012) observed that multiple imputation is efficient in enhancing dependability estimations, even when dealing with data distributions that are skewed. Franco-Martínez et al. (2022) highlighted the compensating nature of multiple imputation, indicating its ability to effectively handle intricate missing data patterns. Enders (2010) emphasized the effectiveness of regression imputation in producing unbiased estimates in several situations where data is missing.

When the effect size on the RB values obtained from the data sets with and without missing data is analyzed, it is seen that the largest effect size is in the reliability coefficients. Accordingly, the reliability coefficient is the most influential factor on the RB. In other words, the bias values obtained from the reliability coefficients differ from each other at a statistically significant level. In addition to this situation, reliability coefficients make unbiased predictions in general terms. The different values used in the calculation formulae of the reliability coefficients (Table 1) may have caused this. The "number of items," which is common in all formulae, has a moderate effect size in the data sets with and without missing data. These findings are also theoretically expected. Increasing the number of test items generally improves reliability estimates, reducing standard errors and bias (A. Hayes & Coutts, 2020; Sheng & Sheng, 2012; Trizano-Hermosilla & Alvarado, 2016; Turner et al., 2017).

The true reliability values (0.70 and 0.90) were obtained in line with McDonald's Omega coefficient formula. Since McDonald's Omega coefficient is calculated with reference to factor loadings, factor loadings vary at each true reliability level. The effect size on reliability levels estimated from sets with missing data is small ($\eta2=0.05$). This may be related to the quality of the items (factor loadings).

Results and Suggestions

This study demonstrates that due to their consistently unbiased outcomes across simulation conditions, the EM, RF, RI, $Mean_{Person}$, and $Median_{Item}$ imputation methods may be more beneficial than zero imputation and listwise deletion (LD) methods when estimating Cronbach's Alpha, Hancock & Mueller's H, Armor's Theta, Gölbaşı-Şimşek & Noyan's Theta G, and Gilmer-Feldt's reliability coefficient. It is noted that Heise & Bohrnstedt's Omega coefficient may overestimate reliability independently of missing data issues. Therefore, although Heise & Bohrnstedt's Omega coefficient is based on factor loadings, indicating a congeneric measurement, it is recommended to assess its model-data fit at low-reliability levels due to low factor loadings before reporting. Considering that measurements in scale development and adaptation studies are often congeneric, Cronbach's Alpha coefficient has provided unbiased results even in datasets that do not meet this assumption. Despite this, researchers may still prefer Cronbach's Alpha coefficient in situations similar to simulation conditions when it yields similar estimates of true reliability. Still, this recommendation pertains solely to the impact of missing data imputation methods. It is crucial to note that there are simulation studies from various perspectives suggesting that Alpha tends to underestimate reliability (McNeish, 2018; Edwards et al., 2021). Furthermore, zero imputation directly affects the assumption of normal distribution, particularly for variables identified as normally distributed. Therefore, if zero imputation is preferred for such variables, a reevaluation of the normal distribution assumption is advisable. Additionally, the $\text{Mean}_{\text{Item}}$ and RI methods, despite dealing with categorical and discrete variables, perform continuous imputation, which alters the structure of the dataset. Given that educational sciences often work with categorical data, attention should be paid to this aspect in planned analyses following the reporting of reliability coefficients. It should be noted that this study is a Monte Carlo simulation, providing valid results for simulation conditions. Still, the validity of the results obtained decreases as the variables in real datasets deviate from these conditions. Therefore, the obtained results should be interpreted within these limitations. Future studies may manipulate the number and skewness of categories of variables to conduct simulation studies.

Declarations

Conflict of Interest: No potential conflict of interest was reported by the author(s).

Ethical Approval: We declare that all ethical guidelines for authors have been followed by all authors. Ethical approval is not required as the datasets in the study were simulated.

References

Allison, P. D. (2002). Missing data. Sage Publications.

Armor, D. J. (1974). Theta reliability and factor scaling. In H. Costner (Ed.), Sociological Methodology (pp. 17–50). Jossey-Bass.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. Journal of School Psychology, 48(1), https://doi.org/10.1016/j.jsp.2009.10.001

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Structural Equation Modeling: A Multidisciplinary Journal, 13(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Béland, S., Jolani, S., Pichette, F., & Renaud, J.-S. (2018). Impact of simple substitution methods for missing data on Classical test theory difficulty and discrimination. The Quantitative Methods for Psychology, 14(3), 180–192. https://doi.org/10.20982/tqmp.14.3.p180

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. https://doi.org/10.1111/j.1467-842X.2001.tb00294.x

Brown, T. A. (2006). Confirmatory factor analysis for applied research. Guilford Press.

Cheema, J. R. (2014). Some general guidelines for choosing missing data handling methods in educational research. Journal of Modern Applied Statistical Methods, 13(2), 53–75. https://doi.org/10.22237/jmasm/1414814520

Cho, E. (2023). reliacoef: Compute and compare unidimensional and multidimensional reliability coefficients (1.0.0) [R].

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), https://doi.org/10.1007/BF02310555

Dai, S. (2021). Handling missing responses in psychometrics: Methods and software. Psych, 3(4), 673–693. https://doi.org/10.3390/psych3040043

Dai, T., Du, Y., Cromley, J., Fechter, T., & Nelson, F. (2024). Analytic approaches to handle missing data in simple matrix sampling planned missing designs. The Journal of Experimental Education, 92(3), 531–558. https://doi.org/10.1080/00220973.2023.2196678

Doove, L. L., Van Buuren, S., & Dusseldorp, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. Computational Statistics & Data Analysis, 72, 92–104. https://doi.org/10.1016/j.csda.2013.10.025

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105 3, 399–412. https://doi.org/10.1111/bjop.12046

Edwards, A. A., Joyner, K. J., & Schatschneider, C. (2021). A simulation study on the performance of different reliability estimation methods. Educational and Psychological Measurement, 81(6), 1089–1117. https://doi.org/10.1177/0013164421994184

Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. Psychological Methods, 8(3), 322–337. https://doi.org/10.1037/1082-989X.8.3.322

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

Enders, C. K. (2010). Applied missing data analysis. Guilford Press.

Fan, J., & Wu, W. (2022). A comparison of multiple imputation strategies to deal with missing nonnormal data in structural equation modeling. Behavior Research Methods. https://doi.org/10.3758/s13428-022-01936-y

Feldt, L. S., & Charter, R. A. (2003). Estimation of internal consistency reliability when test parts vary in effective length. Measurement and Evaluation in Counseling and Development, 36(1), 23–27. https://doi.org/10.1080/07481756.2003.12069077

Finch, W. H. (2016). Missing data and multiple imputation in the context of multivariate analysis of variance. The Journal of Experimental Education, 84(2), 356–372. https://doi.org/10.1080/00220973.2015.1011594

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods, 9(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Gölbaşı-Şimşek, G., & Noyan, F. (2013). McDonald's ω t , Cronbach's α, and Generalized θ for composite reliability of common factors structures. Communications in Statistics - Simulation and Computation, 42(9), 2008–2025. https://doi.org/10.1080/03610918.2012.689062

Goretzko, D. (2021). Factor retention in exploratory factor analysis with missing data. Educational and Psychological Measurement, 82, 444–464. https://doi.org/10.1177/00131644211022031

Goretzko, D., Heumann, C., & Bühner, M. (2020). Investigating parallel analysis in the context of missing data: A simulation study comparing six missing data methods. Educational and Psychological Measurement, 80, 756–774. https://doi.org/10.1177/0013164419893413

Gorsuch, R. L. (2015). Factor analysis (Classic edition). Routledge, Taylor & Francis Group.

Graham, J. W. (2012). Missing data. Springer New York. https://doi.org/10.1007/978-1-4614-4018-5

Graham, J. W., Cumsille, P., & Shevock, A. E. (2013). Methods for handling missing data. In I. B. Weiner (Ed.), Handbook of Psychology, Second Edition (pp. 109–141). http://doi.org/https://doi.org/10.1002/9781118133880.hop202004

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), Structural equation modeling: Present and future—A festschrift in honor of Karl Jöreskog (pp. 195–216). Scientific Software International.

Hayes, A., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But... Communication Methods and Measures, 14, 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hayes, T., & McArdle, J. J. (2017). Should we impute or should we weight? Examining the performance of two CART-based techniques for addressing missing data in small sample research with nonnormal variables. Computational Statistics & Data Analysis, 115, 35–52. https://doi.org/10.1016/j.csda.2017.05.006

Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. Sociological Methodology, 2, 104. https://doi.org/10.2307/270785

Howell, D. C. (2007). The treatment of missing data. In W. Outhwaite & S. Turner, The SAGE Handbook of Social Science Methodology (pp. 212–226). SAGE Publications Ltd. https://doi.org/10.4135/9781848607958.n11

Lee, D. Y., Harring, J. R., & Stapleton, L. M. (2019). Comparing methods for addressing missingness in longitudinal modeling of panel data. The Journal of Experimental Education, 87(4), 596–615. https://doi.org/10.1080/00220973.2018.1520683

Lee, H. J., & Huber, J. C. Jr. (2021). Evaluation of multiple imputation with large proportions of missing data: How much is too much? Iranian Journal of Public Health. https://doi.org/10.18502/ijph.v50i7.6626

Lei, P.-W., & Shiverdecker, L. K. (2020). Performance of estimators for confirmatory factor analysis of ordinal variables with missing data. Structural Equation Modeling: A Multidisciplinary Journal, 27(4), 584–601. https://doi.org/10.1080/10705511.2019.1680292

Leite, W., & Beretvas, S. N. (2010). The performance of multiple imputation for likert-type items with missing data. Journal of Modern Applied Statistical Methods, 9(1), 64–74. https://doi.org/10.22237/jmasm/1272686820

Li, J., & Lomax, R. G. (2017). Effects of missing data methods in SEM under conditions of incomplete and nonnormal Data. The Journal of Experimental Education, 85(2), 231–258. https://doi.org/10.1080/00220973.2015.1134418

Little, R., & Rubin, D. (2002). Statistical analysis with missing data (1st ed.). Wiley. https://doi.org/10.1002/9781119013563

Little, R., & Rubin, D. (2019). Statistical analysis with missing data (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260

McAllister, D. J., & Bigley, G. A. (2002). Work context and the definition of self: How organizational care influences organization-based self-esteem. Academy of Management Journal, 45(5), 894–904. https://doi.org/10.2307/3069320

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. British Journal of Mathematical and Statistical Psychology, 23(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x

McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates.

McNeish, D. M. (2017). Exploratory factor analysis with small samples and missing data. Journal of Personality Assessment, 99(6), 637–652. https://doi.org/10.1080/00223891.2016.1252382

McNeish, D. M. (2018). Thanks coefficient alpha, we'll take it from here. Psychological Methods, 23(3), 412–433. https://doi.org/10.1037/met0000144

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. The R Journal, 9(1), 207. https://doi.org/10.32614/RJ-2017-009

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. Communication Methods and Measures, 5(4), 297–310. https://doi.org/10.1080/19312458.2011.624490

Newman, D. A. (2014). Missing data. Organizational Research Methods. https://doi.org/10.1177/1094428114548590

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3nd ed.). McGraw-Hill.

Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. The Counseling Psychologist, 41(4), 568–600. https://doi.org/10.1177/0011000012445176

Revelle, W. (2024). psych: Procedures for psychological, psychometric, and personality research (R package version 2.4.1) [Computer software]. https://CRAN.R-project.org/package=psych

Rockel, T. (2022). missMethods: Methods for missing data (0.4.0) [R]. https://github.com/torockel/missMethods

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A monte carlo analysis of missing data techniques. Organizational Research Methods, 2(3), 211–232. https://doi.org/10.1177/109442819923001

Şahin Kürşad, M., & Nartgün, Z. (2015). Kayıp veri sorununun çözümünde kullanılan farklı yöntemlerin ölçeklerin geçerlik ve güvenirliği bağlamında karşılaştırılması. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6(2), 254–267. https://doi.org/10.21031/epod.95917

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

Scheffer, J. (2002). Dealing with missing data (1st ed.). Massey University. https://mro.massey.ac.nz/handle/10179/4355

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. American Journal of Epidemiology, 179(6), 764–774. https://doi.org/10.1093/aje/kwt312

Sheng, Y., & Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? Frontiers in Psychology, 3(34). https://doi.org/10.3389/fpsyg.2012.00034

Sigal, M. J., & Chalmers, R. P. (2016). Play it again: Teaching statistics with monte carlo simulation. Journal of Statistics Education, 24(3), 136–156. https://doi.org/10.1080/10691898.2016.1246953

Sijtsma, K., & Van Der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. Multivariate Behavioral Research, 38(4), 505–528. https://doi.org/10.1207/s15327906mbr3804_4

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's Alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. Frontiers in Psychology, 7. https://doi.org/10.3389/fpsyg.2016.00769

Turner, H. J., Natesan, P., & Henson, R. K. (2017). Performance Evaluation of confidence intervals for ordinal coefficient alpha. Journal of Modern Applied Statistical Methods, 16(2), 157–185. https://doi.org/10.22237/jmasm/1509494940

Uysal, İ., & Kılıç, A. (2022). Normal dağılım ikilemi. Anadolu Journal of Educational Sciences International, 12(1), 220–248. https://doi.org/10.18039/ajesi.962653

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3). https://doi.org/10.18637/jss.v045.i03

Van Ginkel, J. R., Van Der Ark, L. A., & Sijtsma, K. (2007). Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results. Multivariate Behavioral Research, 42(2), 387–414. https://doi.org/10.1080/00273170701360803

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing value imputation approach for mass spectrometry-based metabolomics data. Scientific Reports, 8(1), 663. https://doi.org/10.1038/s41598-017-19120-0

Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. Annals of Translational Medicine, 4(1), 1–8. https://doi.org/10.3978/j.issn.2305-5839.2015.12.38

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. https://doi.org/10.1177/0013164415594658