

Handling Item-Level Missing Data: Simpler Is Just as Good

The Counseling Psychologist

41(4) 568-600

© The Author(s) 2013

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0011000012445176

<http://tcp.sagepub.com>



Mike C. Parent¹

Abstract

The topic of missing data has been receiving increasing attention, with calls to apply advanced methods of handling missingness to counseling psychology research. The present study sought to assess whether advanced methods of handling item-level missing data performed equivalently to simpler methods in designs similar to those counseling psychologists typically engage in. Results of an initial preliminary analysis, an analysis using real-world data, and a series of simulation studies were used in the present investigation. Results indicated that available case analysis, mean substitution, and multiple imputation had similar results across low levels of missing data, though in data with higher levels of missing data and other problems (e.g., small sample size or scales with weak internal reliability) mean substitution produced inflation of correlation coefficients among items. The present results support the use of available case analysis when dealing with low-level item-level missingness.

Keywords

methodology, quantitative, training

The topic of missing data has garnered increasing interest in the past several years, with a small but influential literature base devoted to the topic. Within

¹University of Florida, Gainesville, FL, USA

Corresponding Author:

Mike C. Parent, University of Florida, Box 112250, Psychology, University of Florida, Gainesville, FL 32611, USA

Email: michael.parent@ufl.edu

counseling psychology research, there have been calls for the use of sophisticated and complex methods to handle missing data (Schlomer, Bauman, & Card, 2010). Such calls are perhaps spurred on by advances in computational power and software accessibility that enable analyses that may have taken hours of dedicated mainframe time in the past to now be completed in seconds on a personal computer. Calls for increasing attention to advanced methodology within counseling psychology are encouraging inasmuch as they speak to the importance of, and the eagerness of researchers to use, the best available nascent methodologies. However, there is a potential limitation if these complex methods are not necessary in many research contexts. Specifically, advanced methods of handling missing data were originally developed for, and are important to, longitudinal methodology with time-wave-level missing data but are often extrapolated to use with low-level, item-level missingness on multi-item scales (i.e., the type of missingness that most often confronts researchers in counseling psychology).

It is not clear whether these advanced techniques are necessary to deal with missingness that appears primarily in the context of individual items missing from multi-item scales. For example, in Schafer and Graham's (2002) excellent review of best practices in handling missing data, item-level, scale-level, and time-wave-level missingness are discussed, but throughout much of the article there is little distinction between these categories when describing implications of imputation methods. Furthermore, examples given by Schafer and Graham are almost entirely of scale-level missingness, blurring the differences between levels of missingness. The same blurring is present within the paper by Schlomer et al. (2010), who discuss both item- and scale-level missingness but do not make distinctions between them; indeed, results of their analysis, which focuses on very high levels of scale-level missingness in a three-variable model, are extrapolated to item-level missingness. Thus, readers may be unclear about the impact of different methods of handling missing data at different levels of missingness. Although use of advanced methods with item-level missingness is certainly not incorrect, it may simply be unnecessary to conduct the complex procedures if simple methods perform just as well. Validation of relatively simpler procedures would enable researchers to spend less time mastering a complex technique that may not be necessary to their work (as well as prevent errors if that complex technique is ultimately performed incorrectly). Indeed, Wilkinson and the American Psychological Association Task Force on Statistical Inference (1999) averred that in most circumstances, the "minimally sufficient analysis" (p. 598) should be used in research. That is, although researchers should be aware of the most up-to-date methodologies, analysis methods should be chosen to handle the intended

research question appropriately, and not to “impress [one’s] readers or to deflect criticism” (p. 598). Thus, there is a need to assess the impact of different levels of handling missing data at the item level.

In this article I address whether a computationally simple method of handling item-level missing data, available item analysis (AIA¹), functions equivalently to a traditional method (participant mean substitution²) and an advanced method (multiple imputation) for item-level missingness. Specifically, I review theoretical bases for using AIA and examine the three methods in (a) an initial preliminary study of missing data replacement, (b) an example using real-world data, and (c) an example using simulation data.

Definitional Issues

Before delving into the topic of handling item-level missing data, clarification of terminology may be useful. First, I explain what is meant by “missingness.” Then, I review levels of missingness and major methods of handling missing data. This information is provided to orient the reader to these issues, and the interested reader may consult some other excellent and detailed summaries for further information, such as Schafer and Graham (2002).

Data can be missing in three ways: missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR).³ The clearest of these definitions is MNAR: In the case of MNAR, some extraneous variable, which itself is not measured, is related to the variable of interest and influences its missingness. Poorly worded or confusing items that are intentionally omitted by participants are an example of MNAR at the item level. For example, if a researcher was conducting a study on college women’s relationship satisfaction, and a scale contained an item such as “I get along well with my boyfriend or husband,” women who are in same-gender relationships or who are not in relationships may omit this item systematically, potentially biasing the responses to the item. If data on the gender and existence of a relationship partner were not assessed in that project (say, on the demographics page), the data would be MNAR because there is no way to connect the reason for the missingness (gender and existence of partner) to the item responses. MNAR at an item level may often be best thought of as an issue in the design of research rather than a statistical issue to be dealt with post hoc, and researchers should carefully plan projects, evaluate scales at the item level, and pilot surveys with diverse testers to avoid it before a study begins.

Data are MAR if missingness is due to some other, observed variable. For example, in the preceding case, data would be MAR if partner gender and relationship status were assessed. It is possible to use the available data on

these other variables to estimate the missing data through techniques such as expectation maximization or multiple imputation.

Finally, data are MCAR (a special case of MAR) if missingness is due to a factor completely unrelated to the missing data. An example might be if a participant simply misses an item while completing a survey.⁴ Under MAR and MCAR, researchers may assume that missing data are distributed such that a standard missing data technique might be applied (provided other reasonable conditions are met, such as some limit on tolerance for missing data; Schafer & Graham, 2002). Unfortunately, it is never possible to determine with complete certainty whether data are MAR/MCAR (though Little, 1988, developed a test that may help to support or not support MCAR). Researchers typically make an assumption that data might as well be treated as MCAR so long as there is not a clear bias in the missingness (such bias might be evident, for example, if a single item had an abnormally large number of missing values).

Levels of Missingness

The level at which missingness occurs is important to understanding the concept of missingness and considering strategies for handling it. The broadest level of missingness is exclusion from the study. For example, online studies exclude persons without Internet access, studies in the United States exclude persons in other countries, and studies conducted in English exclude people who do not speak, write, or read English. Like MNAR, missingness at this level is best addressed through design and interpretation of results, and is not a truly statistical issue.

Participants may also be missing data at the scale level (in the present article, *scale* is used to mean the level at which most analyses of questionnaire data takes place—i.e., mean scores⁵ across multiple items all intended to measure some common construct. In many cases this would be “subscale”-level missingness, as in the case of a single scale with multiple subscales that are scored separately, such as domains of a Five Factor inventory). Participants may have begun participation but dropped out of the study before finishing and left one or more scales completely missing. This level of missingness may be handled with advanced methods such as multiple imputation that can take into account the available data and correlations among observed variables for all cases.

Finally, there is the level of missingness that most frequently confronts counseling psychology researchers: item-level missingness. At this level, participants have omitted one or more items from one or more multi-item scales in the study without completely missing any scales. It is this level of missingness with which many researchers are often forced to deal, and for which advanced methodologies are now being recommended.

Methods of Handling Missing Data

There are a number of methods of dealing with missing data. The oldest method is likely listwise deletion, or the deletion of participants with any missing data. Although this circumvents problems of imputation of missing data points and may be useful when there is an exceptionally tiny amount of missing data, it may often have a powerfully negative effect on sample size and be potentially biasing in samples with more than trivial missingness (Schafer & Graham, 2002). AIA, the method advocated in this article, is also known as pairwise deletion or pairwise inclusion. AIA is also an older method of handling missing data that involves using the available data for analysis and excluding missing data points only for analyses in which the missing data points would be directly involved. In most research using multiple-item scales, analysis takes place at the scale level, and thus AIA can be used to generate mean scores for scales using the available data without substituting or imputing values. In SPSS, through the "compute" command, the MEAN function is used to generate an AIA scale mean.⁶ Participant mean substitution involves substituting the mean of a single participant's nonmissing items on a scale for missing item values on that particular scale or subscale.⁷ Practically, this is typically accomplished by generating the AIA scale mean then substituting this mean line-by-line for missing data on that scale. Mean substitution would produce output very similar to AIA for means (though rounding the substituted mean to fewer decimal places or to the integer would introduce some deviation), and variability (i.e., standard deviations) would be reduced. Multiple imputation involves the generation of multiple data sets with missing value imputations that factor in correlations among variables and random error, and pools these results. This is typically accomplished in a program such as NORM (Schafer, 2000), the multiple imputation add-on module for SPSS, or the MI and MIANALYZE procedure in SAS. Many other methods of imputation exist but are used relatively less often and in any case are generally supplanted by the other methods in terms of either simplicity or lack of bias. Another method worth noting, full information maximum likelihood estimation, is a special method of handling missing data in structural equation modeling applications that, like AIA, does not involve imputation of missing data at all but rather uses the available data to carry out analyses.

AIA offers a number of benefits over mean substitution and multiple imputation as applied to item-level missingness. The primary advantage is that imputation may not be at all necessary for most analyses when dealing with item-level missing data on multi-item scales. Variables are typically entered into analyses at the scale level after all item responses have been averaged. Thus, a complete (at the item level) data set is unnecessary. Even analyses

that are sometimes thought of as requiring complete item-level data, such as calculation of Cronbach's alpha or structural equation modeling, do not necessarily require complete data (AIA calculation of Cronbach's alpha will be discussed shortly, and analyses in the structural equation modeling family, including path analysis, latent variable modeling, and confirmatory factor analysis, may use full information maximum likelihood estimation and bypass the need for any imputation).

Another advantage is that AIA is a simple procedure that can be implemented easily and is not limited by technological savvy of researchers, availability of software, or suitability of the collected data to the analysis technique. Multiple imputation, though useful in many contexts, is complex to learn and more complex to fully understand, requires learning how to navigate a program such as NORM (Schafer, 2000), and even after mastery data may still produce issues such as nonpositive definite matrices that complicate or preclude completion of the process.

Also notable is that several scales familiar to many counseling psychology research contexts exist for which the preferred scoring method is AIA. The Objectified Body Consciousness Scale (McKinley & Hyde, 1996) contains "not applicable" responses that are not included in mean score calculations, and missed items on the Lesbian, Gay, and Bisexual Knowledge and History Scale (Worthington, Dillon, & Becker-Schutte, 2005) are not counted in mean scale scores. Thus, some scale authors explicitly allow for the existence of nonresponse in their scales and recommend that AIA be used in creating a mean scale score.

The Present Studies

The present studies undertake an investigation of the differences between AIA, mean substitution, and multiple imputation on artificially constructed data sets simulating single measures (Studies 1 and 3) and on a real data set with multiple measures (Study 2). Participant mean substitution is included in the present analysis because it has been used frequently in the past, and evaluating it may be helpful in understanding the impact of this method in past research. Overall, I expected no meaningful difference to emerge among the methods. Throughout the results, I describe data to the third decimal place rather than the second to avoid confusing substantive differences with rounding.

Study 1

Study 1 used an artificial data set created specifically for this study as a preliminary investigation. The differences in values among the three methods of handling

Table 1. Correlations Among the Original Data Set Items and Total Score, and Missing Value Items and Total Score.

	Item 1	Item 2	Item 3	Item 4	Item 5	Total	Original Data		Missing Value	
							M	SD	M	SD
Item 1		.290**	.283**	.338**	.206**	.651**	2.905	2.305	2.889	2.294
Item 2	.278**		.301**	.300**	.237**	.665**	3.225	2.411	3.172	2.417
	176									
Item 3	.325**	.292**		.348**	.241**	.682**	3.105	2.440	3.096	2.408
	177	173								
Item 4	.316**	.269**	.374**		.118	.654**	2.875	2.370	2.927	2.359
	182	178	179							
Item 5	.201**	.247**	.261**	.101		.562**	2.975	2.388	2.969	2.385
	184	180	181	186						
Total	.657**	.656**	.711**	.649**	.574**		3.017	1.531	3.014	1.544
	190	186	187	192	194					

Correlations above the diagonal represent correlations for the original data set (before generating missingness); correlations below the diagonal are the missing-value data. For the missing-value data, the number below the correlation is the *n* for that bivariate relation.

***p* < .01.

missing data were compared for mean differences, scale variability differences, differences in correlations among the items, and reliability coefficients. I hypothesized that no statistically or practically significant differences would emerge from comparisons and that the effect size of differences would be small.

Method

I constructed the artificial data set using a single run of a Monte Carlo simulation in Mplus (Muthén & Muthén, 2010). Responses to five variables were generated on a multivariate normal distribution then constrained to a 7-point scale (0-6). I constrained items to load onto a latent variable in the simulation to produce correlated item data, though the association was intentionally not strong to create an artificial data set that resembled a scale of, at best, average quality. Two hundred cases were generated. Table 1 provides summary data on the initial data set (before adding missing data) for reference.

I generated missingness manually by creating a random number matrix of the same dimensions as the data matrix (5 × 200) with cell values between 1 and 100. If a cell value in the random matrix was equal to or less than 5,

Table 2. Correlations Among Mean Substitution and Multiple Imputation Item and Total Scores.

	Item 1	Item 2	Item 3	Item 4	Item 5	Total	Mean Substitution		Multiple Imputation	
							M	SD	M	SD
Item 1		.312**	.353**	.340**	.219**	.667**	2.900	2.262	2.908	2.347
Item 2	.270**		.329**	.298**	.264**	.671**	3.185	2.377	3.167	2.390
Item 3	.317**	.285**		.401**	.292**	.722**	3.130	2.369	3.115	2.425
Item 4	.308**	.269**	.358**		.129	.656**	2.895	2.328	2.909	2.390
Item 5	.203**	.243**	.265**	.098		.581**	2.985	2.365	2.980	2.385
Total	.649**	.649**	.698**	.634**	.568**		3.019	1.543	3.016	1.502

Data for mean substitution are above the diagonal and pooled data for multiple imputation are below the diagonal.

** $p < .01$.

I replaced the corresponding cell in the original data set with a missing value code. In no instance was a participant allowed to have more than one missing value (thus, each participant was missing either one or no data points on the five-item scale). Ultimately, I replaced 51 values with missing values for as many participants out of a total of 1,000 data points, or 5.1% missingness. Though no systematic analysis of prevalence of missing data has been conducted, this level of missingness approximates a substantial level of item-level missingness; in prior work conducted by me, total item-level missingness has been under 0.2% (e.g., Parent & Moradi, 2010, 2011a, 2011b) in studies with designs similar to those employed by many counseling psychologists.

The missing-value data set was identical to the AIA data set. Means for the AIA data set (generated as the mean for the number of nonmissing items: If a case had five data points, the mean is the sum of responses divided by five; if four, then divided by four) and standard deviations are presented in Table 1. I generated mean substitution values by calculating the AIA means for each participant, rounded these values to the nearest integer, and substituted them the missing data points. Means and standard deviations for mean substitution are presented in Table 2. I conducted multiple imputation in SPSS using 1,000 iterations to generate five imputed data sets. In the multiple imputation data set, values were rounded to the nearest integer and constrained to be within the original 0-6 range; means and standard deviations are presented in Table 2. Reported results are for the pooled multiple imputation data set, with the exception of Cronbach's alpha.

Table 3. Mean Score Comparisons Across Methods

Item	AIA		Mean		MI		Comparisons		
	M	SD	M	SD	M	SD	AIA: Mean	AIA: MI	Mean: MI
1	2.889	2.294	2.900	2.262	2.908	2.347	0.011	0.019	0.008
2	3.172	2.417	3.185	2.377	3.167	2.390	0.013	0.005	0.018
3	3.096	2.408	3.130	2.369	3.115	2.425	0.034	0.019	0.015
4	2.927	2.359	2.895	2.328	2.909	2.390	0.032	0.018	0.014
5	2.969	2.385	2.985	2.365	2.980	2.385	0.016	0.011	0.005
Total	3.014	1.544	3.019	1.543	3.016	1.502	0.005	0.002	0.003

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation.

In SPSS, Cronbach's alpha is calculated with listwise deletion to handle missing data. However, it is possible to generate an AIA Cronbach's alpha using syntax (see the appendix⁸). The resultant alpha is reported for the AIA data. Similarly, calculation of Cronbach's alpha in the multiple imputation data set also presented a problem; Cronbach's alpha does not have a standard error in SPSS multiple imputation and so a pooled estimate could not be generated. Thus, Cronbach's alpha values for the five imputed data sets are reported. The 95% confidence intervals are also reported for all alphas (using syntax from Iacobucci & Duhachek, 2003).

Results

First, I compared differences in mean values for items and total scores across the three data sets. Absolute values of the differences among the scores are reported in Table 3. The largest difference, between the AIA and mean substitution methods on Item 3, was 0.034.

Next, differences among variabilities were assessed. Emergent differences were very small, with an average mean of absolute differences in standard deviations of .035 ($SD = .122$). The largest observed absolute difference in standard deviations was .085, for the difference between mean substitution and multiple imputation on Item 1.

Next, differences in the correlation matrices were assessed. Absolute values were calculated for the differences between each bivariate pair of data sets (presented in Tables 1 and 2). The largest difference among correlations was

.044 on Items 2 and 3 when comparing mean substitution and multiple imputation.

Finally, Cronbach's alphas were calculated for the three data sets. Alpha for items on the AIA data set (using the aforementioned modification) was .643, 95% CI [.558, .716]; alpha for items on the mean substitution data set was .675, 95% CI [.598, .741]; and alphas for the five multiple imputation data sets were .630, 95% CI [.542, .705]; .634, 95% CI [.547, .708]; .649, 95% CI [.566, .720]; .653, 95% CI [.571, .724]; and .627, 95% CI [.539, .703].

Discussion

In sum, across the artificial data sets, there was no major difference between the methods of missing data replacement in terms of mean values, variability, correlations, or Cronbach's alpha internal reliability coefficients. These results support the equivalence of missing data methods across these artificial data sets in this preliminary investigation; to build on this finding, another analysis using real data was undertaken.

Study 2

The purpose of Study 2 was to further investigate the effects of the three methods of handling missing data using an existing data collection set. This investigation provides an assessment of the methods using data more akin to what researchers are actually accustomed to using. Once again, I expected that no meaningful differences would emerge.

Method

Data were taken from a larger data set previously collected by the present author. The data are explained in greater detail in Parent and Moradi (2011a). The included data consisted of six scales: the Sociocultural Attitudes Toward Appearance Questionnaire (SATAQ; Heinberg, Thompson, & Stormer, 1995), the Objectified Body Consciousness Scale Surveillance and Shame subscales (OBCS-Surv and OBCS-Shame; McKinley & Hyde, 1996), the Drive for Muscularity Scale (DMS; McCreary & Sasse, 2000), and the Outcome Expectations for Anabolic Androgenic Steroid Use and Intentions for Anabolic Androgenic Steroid Use scale (OE-AAS and I-AAS; Parent & Moradi, 2011a, and 276 cases. There were 33 missing data points distributed over 23 cases. (There are slightly more missing data in this data set than the published article because missing data on the OBCS-Surv and OBCS-Shame were treated using AIA in the original study, as per their

Table 4. Mean Score Comparisons Across Methods.

Scale	AIA		Mean		MI		Comparisons		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	AIA: Mean	AIA: MI	Mean: MI
SATAQ	1.776	0.733	1.776	0.733	1.777	0.733	0.000	0.001	0.001
OBCS-Surv	2.776	0.827	2.776	0.826	2.775	0.827	0.000	0.001	0.001
OBCS-Shame	1.803	0.799	1.803	0.780	1.803	0.800	0.000	0.000	0.000
DMS	1.872	0.863	1.879	0.864	1.878	0.862	0.007	0.006	0.001
OE-AAS	0.823	1.412	0.823	1.411	0.823	1.412	0.000	0.000	0.000
I-AAS	0.561	1.119	0.562	1.120	0.562	1.119	0.001	0.001	0.000

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation.

instructions. In the present study they were treated as missing to allow more missing data points with which to assess the impact of those methods. In addition, there were 6 extra participants who were identified as multivariate outliers in the original study whose data were retained for the present analysis, resulting in fit statistics and scale scores differing slightly from the original model.) All of the scales had at least 1 missing data point. AIA, mean substitution, and multiple imputation were all conducted in the same manner as in Study 1.

Analyses are similar to the previous study. However, a replication of the path analysis from the original research is presented to examine the impact of the methods of handling missing data on analyses that might be undertaken by researchers. I present the same statistics as in Study 1 (means, standard deviations, mean comparisons, variabilities, correlation matrices and correlation matrix comparisons, and Cronbach's alpha reliability coefficients). The path analysis undertaken in Parent and Moradi (2011a) is replicated using each of the three methods of handling missing data, and I present path coefficients and fit indices. Again, for all comparisons I expected no meaningful differences to emerge.

Results and Discussion

First, differences in mean values for items and total scores were compared across the three data sets. Mean scores and absolute values of the differences among the scores are reported in Table 4. The largest difference, between AIA and mean substitution on DMS, was 0.007.

Next, differences among variabilities were assessed. Again, emergent differences were very small, with an average mean of absolute differences in

Table 5. Cronbach's Alpha for Scales by Missing Data Method.

Scale	AIA	Mean	MI				
			1	2	3	4	5
SATAQ	0.850	0.852	0.852	0.852	0.851	0.851	0.851
OBCS-Surv	0.847	0.848	0.847	0.848	0.848	0.848	0.848
OBCS-Shame	0.780	0.784	0.782	0.782	0.784	0.784	0.785
DMS	0.887	0.886	0.885	0.885	0.885	0.885	0.885
OE-AAS	0.950	0.950	0.950	0.950	0.950	0.950	0.950
I-AAS	0.868	0.868	0.868	0.867	0.867	0.868	0.868

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation.

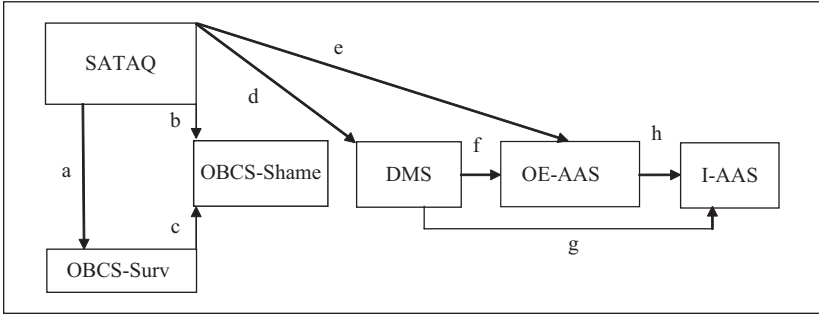


Figure 1. Final model in Parent & Moradi (2011a).

standard deviations of .003 ($SD = .007$). The largest observed absolute difference in standard deviations was .020 for the difference between mean substitution and multiple imputation on OBCS-Shame.

Next, differences in the correlation matrices were assessed. Absolute values were calculated for the differences between each bivariate pair of data sets, and the matrices are presented in Table 5. The largest difference among correlations was .002, which occurred for several of the comparisons.

Next, Cronbach's alphas were calculated for the three data sets. Alpha for the items are presented in Table 5 (confidence intervals omitted as the homogeneity of the results minimizes their utility).

Finally, the original path analyses were replicated. Figure 1 presents the path diagram and Table 6 presents path coefficients and fit indices for each of the three methods of data collection.

Table 6. Comparison of Standardized Path Coefficients (With Standard Errors) and Fit Indices Across Methods.

	Paths										Fit Indices			
	a	b	c	d	e	f	g	h	χ^2	CFI	RMSEA	SRMR		
AIA	0.511 (0.044)	0.174 (0.062)	0.327 (0.060)	0.530 (0.043)	0.224 (0.061)	0.338 (0.059)	0.141 (0.052)	0.551 (0.046)	8.787 (0.046)	0.996 (0.030)	0.030 (0.000, 0.084)	0.029		
Mean	0.511 (0.044)	0.175 (0.062)	0.326 (0.060)	0.530 (0.043)	0.224 (0.061)	0.338 (0.059)	0.142 (0.052)	0.551 (0.046)	8.798 (0.046)	0.996 (0.031)	0.030 (0.000, 0.084)	0.029		
MI	0.511 (0.044)	0.174 (0.062)	0.327 (0.060)	0.531 (0.043)	0.226 (0.061)	0.336 (0.059)	0.141 (0.052)	0.551 (0.046)	8.748 (0.046)	0.996 (0.030)	0.030 (0.000, 0.084)	0.029		

Note. Values in the Paths columns are the standardized path coefficients and standard errors (in parentheses).

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation; CFI = comparative fit index; RMSEA = residual mean square error of approximation (with 95% confidence interval); SRMR = standardized root mean residual.

Study 3

The purpose of Study 3 was to investigate the methods of handling missing data in a series of simulated data sets that varied by specific parameters. This investigation was intended to provide additional detail on the impact of the three methods of handling missing data as a function of number of items, sample size, association among items, and level of missingness.

Method

Monte Carlo simulations were conducted in R (R Development Core Team, 2008). Data for 1,000 Monte Carlo runs for each condition were generated on multivariate normal distributions with a mean of 4.00 and a standard deviation of 1.25. The data were recategorized to create a 7-point integer response scale (i.e., 0 = *values* < 0-1, 1 = 1-2, etc.). In each run, data sets varied by number of items, sample size, association among items, and level of missingness. Number of items was set to 3, 5, or 10, representing three common scale lengths (3 items being suggested as the minimum to conduct structural equation modeling using latent variable modeling, though such short scales are probably generally unreliable and suboptimal for use as indicators in structural equation modeling; Tabachnick & Fidell, 2007; Worthington & Whittaker, 2006). Sample size was set to either 50 (representing a small-scale study) or 200 (representing the lower conventional bound for structural equation modeling studies; Tabachnick & Fidell, 2007). Associations among items were set to approximately .30 (representing a relatively weak scale) and .50 (representing a standard scale). Level of missingness was set to 1%, 5%, or 10%. The levels chosen for investigation were intended to represent data that might commonly occur in counseling psychology research; a literally infinite number of other levels and combinations of each of the variables is possible, and future work may continue to explore such variables.

Missingness was imposed quasi-randomly within the R runs, with the limitation that no row could be completely missing (this would be akin to a participant missing the scale entirely). Multiple imputation was conducted within R using the AMELIA program (Honaker, King, & Blackwell, 2011) and the multiple imputation runs were transferred to SPSS version 18 (SPSS Inc., 2010) for pooling as in Studies 1 and 2.

Results are presented in Tables 7, 8, and 9. I present the means, standard errors, and Cronbach's alphas with 95% confidence intervals for the data sets.

Table 7. Monte Carlo Results for Three-Item Simulations.

Method	Three Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
<i>n</i> = 50						
AIA						
M	4.500, 4.388, 4.383	4.200, 3.729, 3.957	3.673, 3.915, 3.884	4.140, 3.917, 3.792	3.830, 3.979, 3.761	4.091, 4.000, 3.822
SE	0.203, 0.175, 0.176	0.181, 0.210, 0.163	0.164, 0.181, 0.186	0.181, 0.177, 0.177	0.168, 0.165, 0.161	0.180, 0.157, 0.194
Alpha	.526 (.242-.715)	.553 (.285-.732)	.370 (-.007-.622)	.684 (.495-.810)	.725 (.560-.835)	.632 (.412-.779)
Mean						
M	4.500, 4.420, 4.380	4.200, 3.760, 3.900	3.680, 3.900, 3.860	4.140, 3.920, 3.800	3.860, 3.960, 3.740	4.100, 3.940, 3.840
SE	0.203, 0.176, 0.171	0.181, 0.207, 0.167	0.163, 0.177, 0.183	0.181, 0.183, 0.174	0.164, 0.162, 0.156	0.188, 0.163, 0.195
Alpha	.555 (.288-.733)	.623 (.397-.774)	.467 (.148-.680)	.714 (.543-.828)	0.724 (.559-.834)	.808 (.693-.885)
MI						
M	4.506, 4.395, 4.392	4.200, 3.733, 3.945	3.672, 3.964, 3.838	4.140, 3.915, 3.771	3.843, 3.979, 3.743	4.155, 3.992, 3.835
SE	0.204, 0.178, 0.182	0.181, 0.209, 0.170	0.165, 0.190, 0.213	0.181, 0.178, 0.181	0.171, 0.181, 0.164	0.189, 0.176, 0.197
Alpha (M, range)	.509 (.215-.705); (.486 [.178-.691]- .529 [.247-.717])	.557 (.292-.734); (.515 [.224-.709]- .593 [.349-.756])	.399 (.039-.639); (.226 [-.238-.535]- .492 [-.188-.695])	.689 (.503-.813); (.656 [.450-.793]- .713 [.541-.828])	.698 (.517-.819); (.664 [.463-.798]- .724 [.559-.834])	.739 (.583-.843); (.663 [.461-.796]- .777 [.643-.866])
<i>n</i> = 200						
AIA						
M	3.980, 3.980, 3.903	4.011, 4.064, 4.011	4.017, 3.954, 4.043	3.995, 3.924, 3.780	4.000, 3.921, 3.892	3.872, 3.843, 4.071
SE	0.101, 0.097, 0.090	0.089, 0.086, 0.093	0.101, 0.093, 0.089	0.089, 0.087, 0.091	0.087, 0.087, 0.089	0.088, 0.087, 0.091
Alpha	.601 (.495-.688)	0.561 (.444-.657)	.601 (.495-.688)	.685 (.601-.754)	.752 (.686-.806)	.757 (.692-.810)

(continued)

Table 7. (continued)

Method	Three Items						
	Weak Association			Moderate Association			
	Missing = 1 %	Missing = 5 %	Missing = 10 %	Missing = 1 %	Missing = 5 %	Missing = 10 %	
<i>n</i> = 50							
Mean							
<i>M</i>	3.985, 3.975, 3.900	3.980, 4.015, 3.975	4.040, 3.975, 4.045	3.995, 3.930, 3.780	4.015, 3.945, 3.890	3.895, 3.870, 4.055	
<i>SE</i>	0.099, 0.097, 0.090	0.089, 0.086, 0.093	0.097, 0.095, 0.087	0.088, 0.087, 0.091	0.085, 0.086, 0.088	0.087, 0.087, 0.090	
Alpha	.613 (.510-.697)	0.642 (.547-.720)	.687 (.604-.755)	.691 (.608-.758)	0.761 (.697-.813)	.780 (.721-.828)	
MI							
<i>M</i>	3.985, 3.984, 3.893	4.000, 4.052, 3.996	3.995, 3.980, 4.045	3.999, 3.927, 3.780	4.008, 3.935, 3.893	3.884, 3.841, 4.085	
<i>SE</i>	0.099, 0.099, 0.091	0.093, 0.089, 0.094	0.110, 0.099, 0.090	0.089, 0.087, 0.091	0.087, 0.087, 0.091	0.088, 0.089, 0.094	
Alpha (<i>M</i> , range)	.598 (.491-.686); (.587 [.477-.677]- .607 [.502-.693])	.562 (.445-.658); (.548 [.428-.647]- .588 [.478-.678])	.604 (.498-.690); (.561 [.444-.657]- .624 [.524-.706])	.688 (.605-.756); (.685 [.601-.754]-.691 [.609-.758])	.739 (.669-.796); (.734 [.663-.792]- .745 [.677-.801])	.745 (.677-.801); (.723 [.649-.783]- .757 [.692-.810])	

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation. Correlations for AIA data use pairwise deletion. Alphas for MI are summarized across the five imputation runs. Values in *M* and *SE* rows represent item values.

Table 8. Monte Carlo Results for Five-Item Simulations.

Method	Five Items				
	Weak Association		Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%
	<i>n</i> = 50				
AIA					
M	4.340, 4.104, 4.143, 4.104, 4.245, 4.021, 4.240, 3.898	3.978, 3.957, 3.979, 3.660, 3.780, 3.918, 4.312, 4.208	3.978, 3.957, 3.979, 3.660, 3.780, 3.918, 3.812, 3.980, 3.980, 4.043, 3.761, 3.857, 4.240, 3.898	3.978, 3.957, 3.979, 3.660, 3.780, 3.900, 4.082, 3.755	3.979, 3.977
SE	0.209, 0.163, 0.171, 0.196, 0.151, 0.176, 0.168, 0.176	0.178, 0.193, 0.173, 0.189, 0.186, 0.191, 0.148, 0.164	0.178, 0.193, 0.173, 0.189, 0.186, 0.191, 0.167, 0.172, 0.165, 0.169, 0.169, 0.204, 0.168, 0.176	0.178, 0.193, 0.173, 0.189, 0.186, 0.191, 0.167, 0.172, 0.165, 0.169, 0.169, 0.204, 0.168, 0.176	0.179, 0.181
Alpha	.631 (.440-.771)	.675 (.507-.798)	.591 (.380-.746)	.857 (.783-.911)	.771 (.652-.858)
Mean					
M	4.340, 4.080, 4.120, 4.040, 4.240, 4.020, 4.240, 3.900	3.960, 3.920, 4.020, 3.660, 3.780, 3.880, 4.300, 4.200	3.960, 3.920, 4.020, 3.660, 3.780, 3.880, 3.820, 3.980, 3.980, 3.960, 3.800, 3.880, 4.240, 3.900	3.960, 4.400, 3.980, 3.900	4.000, 3.940
SE	0.209, 0.161, 0.171, 0.198, 0.150, 0.173, 0.168, 0.174	0.171, 0.189, 0.173, 0.189, 0.186, 0.193, 0.146, 0.171	0.171, 0.189, 0.173, 0.189, 0.186, 0.193, 0.166, 0.170, 0.165, 0.171, 0.171, 0.199, 0.168, 0.174	0.171, 0.189, 0.173, 0.189, 0.186, 0.193, 0.166, 0.170, 0.165, 0.171, 0.171, 0.199, 0.168, 0.174	0.176, 0.188
Alpha	.679 (.513-.801)	.689 (.528-.807)	.638 (.451-.775)	.860 (.788-.913)	.833 (.896-.747)
MI					
M	4.340, 4.0823, 4.141, 4.240, 3.894	4.050, 4.241, 4.026, 4.302, 4.218	3.963, 3.956, 3.990, 3.660, 3.780, 3.884, 3.827, 3.992, 3.980, 3.991, 3.728, 3.818, 4.141, 4.240, 3.894	3.956, 4.394, 3.980, 3.900	4.028, 3.901
SE	0.209, 0.168, 0.175, 0.200, 0.151, 0.175, 0.168, 0.179	0.178, 0.197, 0.182, 0.189, 0.186, 0.193, 0.162, 0.178	0.178, 0.197, 0.182, 0.189, 0.186, 0.193, 0.168, 0.173, 0.165, 0.176, 0.178, 0.208, 0.168, 0.179	0.178, 0.197, 0.182, 0.189, 0.186, 0.193, 0.168, 0.173, 0.165, 0.176, 0.178, 0.208, 0.168, 0.179	0.181, 0.190

(continued)

Table 8. (continued)

Method	Five Items					
	Weak Association			Moderate Association		
	Missing = 1 %	Missing = 5 %	Missing = 10 %	Missing = 1 %	Missing = 5 %	Missing = 10 %
	<i>n</i> = 50					
Alpha	.627 (.434-.768);	.678 (.512-.800);	.567 (.343-.731);	.859 (.786-.912);	.783 (.671-.865);	.808 (.709-.881);
(M,	.615 [.416-.761]-	.658 [.481-.788]-	.542 [.305-.716]-	.858 [.785-.912]-	.773 [.656-.859]-	.797 [.692-.874]-
range)	.642 [.457-.778]]	.687 [.525-.806]]	.595 [.386-.749]]	.860 [.788-.913]]	.790 [.681-.870]]	.821 [.728-.889]]
	<i>n</i> = 200					
AIA						
M	4.015, 3.872, 4.095, 3.947, 3.802, 3.938, 3.995, 4.027, 4.048, 4.172, 4.211, 3.965, 3.886, 3.948, 3.923, 3.973, 3.927, 4.045, 3.960, 4.025	4.011, 3.923	4.176, 3.920	4.201, 4.005	4.000, 3.995	3.961, 3.813
SE	0.092, 0.089, 0.093, 0.089, 0.086, 0.093, 0.095, 0.095, 0.089, 0.087, 0.086, 0.095, 0.095, 0.092, 0.100, 0.089, 0.085	0.089, 0.095	0.084, 0.086	0.086, 0.093	0.087, 0.086	0.085, 0.088
Alpha	.722 (.661-.776)	.700 (.634-.758)	.657 (.582-.724)	.820 (.780-.855)	.842 (.807-.873)	.832 (.795-.865)
Mean						
M	4.025, 3.875, 4.095, 3.930, 3.810, 3.950, 3.975, 4.020, 4.050, 4.175, 4.210, 3.965, 3.880, 3.935, 3.980, 3.950, 4.060, 3.960, 4.025	4.000, 3.900	4.165, 3.930	4.190, 4.005	3.935, 3.985, 4.005	3.970, 3.845
SE	0.093, 0.088, 0.093, 0.088, 0.085, 0.092, 0.094, 0.094, 0.088, 0.087, 0.086, 0.094, 0.095, 0.091, 0.102, 0.091, 0.095, 0.097, 0.088, 0.091	0.087, 0.092	0.082, 0.083	0.085, 0.093	0.087, 0.085	0.084, 0.088

(continued)

Table 8. (continued)

Method	Five Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
	<i>n</i> = 50					
Alpha	.724 (.663-.778)	.718 (.656-.773)	.710 (.646-.766)	.819 (.779-.854)	.853 (.821-.882)	.855 (.823-.883)
MI						
M	4.021, 3.875, 4.092, 3.960, 4.025	3.936, 3.799, 3.946, 4.007, 3.912	3.948, 4.023, 4.068, 4.165, 3.920	4.176, 4.211, 3.966, 4.209, 4.005	3.878, 3.947, 3.935, 4.002, 4.004	3.988, 3.943, 4.084, 3.964, 3.823
SE	0.093, 0.090, 0.094, 0.089, 0.091	0.093, 0.088, 0.096, 0.084, 0.105	0.103, 0.100, 0.089, 0.093, 0.094	0.088, 0.086, 0.095, 0.087, 0.093	0.096, 0.093, 0.104, 0.088, 0.086	0.097, 0.105, 0.107, 0.087, 0.085
Alpha	.721 (.660-.775);	.699 (.633-.758);	.667 (.594-.732);	.817 (.777-.853);	.843 (.808-.874);	.831 (.794-.864);
(M,	.719 [.657-.774]-	.690 [.622-.750]-	.656 [.580-.723]-	.815 [.774-.851]-	.840 [.805-.871]-	.829 [.791-.862]-
range)	.725 [.664-.778]	.716 [.654-.771]	.685 [.616-.746]	.819 [.779-.854]	.847 [.813-.877]	.833 [.796-.865]

AIA = available item analysis; mean = mean substitution; MI = multiple imputation. Correlations for AIA data use pairwise deletion. Alphas for MI are summarized across the five imputation runs. Values in *M* and *SE* rows represent item values

Table 9. Monte Carlo Results for Ten-Item Simulations

Method	Ten Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
<i>n</i> = 50						
AIA						
M	4.340, 3.920, 4.000, 4.083, 4.000, 4.163, 3.591, 3.533, 3.581, 4.000, 3.960, 3.796, 3.809, 4.000, 4.174, 4.191, 4.114, 4.227, 4.020, 4.400, 3.880, 4.170, 4.143, 4.067, 3.913, 3.818, 3.889, 4.083, 3.979, 3.880, 4.042, 4.104, 3.917, 3.841, 3.929, 4.163, 4.520, 3.898, 4.347, 3.860, 4.085, 4.022, 3.652, 3.870, 3.617, 4.000, 4.040, 3.940, 3.780, 3.711, 4.000, 3.915, 4.130, 3.651, 4.200 4.174 3.844 4.000 3.978 3.667					
SE	0.158, 0.173, 0.161, 0.177, 0.185, 0.159, 0.194, 0.206, 0.201, 0.157, 0.190, 0.191, 0.190, 0.176, 0.164, 0.163, 0.196, 0.204, 0.175, 0.194, 0.211, 0.190, 0.185, 0.158, 0.181, 0.203, 0.222, 0.142, 0.153, 0.168, 0.196, 0.163, 0.170, 0.143, 0.206, 0.174, 0.179, 0.185, 0.170, 0.214, 0.195, 0.175, 0.198, 0.188, 0.176, 0.162, 0.166, 0.168, 0.190, 0.170, 0.180, 0.176, 0.183, 0.180, 0.182 0.191 0.212 0.168 0.180 0.212					
Alpha	.813 (.725-.882)	.813 (.725-.882)	.829 (.749-.892)	.885 (.831-.927)	.894 (.844-.933)	.902 (.856-.938)
Mean						
M	4.340, 3.920, 4.000, 4.040, 3.980, 4.160, 3.640, 3.540, 3.660, 4.040, 3.960, 3.800, 3.880, 4.000, 4.160, 4.180, 4.140, 4.240, 4.020, 4.400, 3.880, 4.160, 4.140, 4.080, 3.840, 3.780, 3.880, 4.040, 3.960, 3.880, 4.020, 4.080, 4.000, 3.900, 3.980, 4.100, 4.520, 3.880, 4.360, 3.860, 4.080, 4.040, 3.720, 3.880, 3.640, 4.000, 4.040, 3.940, 3.780, 3.740, 3.980, 3.920, 4.080, 3.720, 4.200 4.160 3.860 4.040 3.960 3.720					

(continued)

Table 9. (continued)

Method	Ten Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
	n = 50					
SE	0.158, 0.173, 0.159, 0.176, 0.184, 0.157, 0.187, 0.196, 0.193, 0.156, 0.190, 0.190, 0.191, 0.174, 0.170, 0.158, 0.185, 0.197, 0.175, 0.194, 0.211, 0.184, 0.183, 0.156, 0.179, 0.192, 0.213, 0.146, 0.151, 0.168, 0.193, 0.161, 0.176, 0.152, 0.193, 0.172, 0.179, 0.184, 0.168, 0.214, 0.191, 0.172, 0.194, 0.182, 0.178, 0.162, 0.166, 0.168, 0.190, 0.166, 0.180, 0.171, 0.178, 0.178, 0.182, 0.185, 0.205, 0.209					
Alpha	.813 (.725-.882)	.823 (.740-.888)	.851 (.781-.906)	.889 (.837-.930)	.909 (.866-.942)	.911 (.869-.944)
MI						
M	4.340, 3.920, 4.001, 4.060, 3.956, 4.164, 3.645, 3.574, 3.684, 4.079, 3.960, 3.809, 3.783, 3.989, 4.167, 4.185, 4.162, 4.250, 4.020, 4.400, 3.880, 4.136, 4.115, 4.092, 3.903, 3.746, 3.945, 4.053, 3.974, 3.880, 4.001, 4.070, 3.964, 3.923, 3.937, 4.199, 4.520, 3.873, 4.353, 3.860, 4.093, 4.059, 3.621, 3.875, 3.689, 4.000, 4.040, 3.940, 3.780, 3.626, 3.972, 3.957, 4.080, 3.613, 4.200, 3.863, 4.178, 4.028, 3.950, 3.716					
SE	0.158, 0.173, 0.160, 0.181, 0.200, 0.161, 0.228, 0.211, 0.236, 0.163, 0.190, 0.194, 0.191, 0.175, 0.174, 0.166, 0.223, 0.242, 0.175, 0.194, 0.211, 0.196, 0.187, 0.172, 0.194, 0.225, 0.249, 0.146, 0.153, 0.168, 0.203, 0.172, 0.174, 0.171, 0.276, 0.277, 0.179, 0.187, 0.170, 0.214, 0.201, 0.185, 0.211, 0.191, 0.185, 0.162, 0.166, 0.168, 0.190, 0.191, 0.182, 0.176, 0.233, 0.199, 0.182, 0.212, 0.225, 0.171, 0.191, 0.259					
	0.182	0.212	0.225	0.171	0.191	0.259
	(continued)					

(continued)

Table 9. (continued)

Method	Ten Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
<i>n</i> = 50						
Alpha	.778 (.674-.859);	.806 (.715-.877);	.821 (.737-.887);	.887 (.834-.928);	.898 (.850-.935);	.859 (.793-.911);
(M,	(.775 [.669-.857])-	(.787 [.687-.865])-	(.807 [.716-.878])-	(.886 [.832-.928])-	(.892 [.841-.932])-	(.701 [.560-.811])-
range)	.780 [.676-.861])	.820 [.735-.886])	.834 [.756-.895])	.889 [.837-.930])	.902 [.856-.938])	.904 [.859-.939])
<i>n</i> = 200						
AIA						
M	3.939, 4.040, 4.098, 4.079, 4.047, 3.954, 3.841, 3.989, 3.963, 4.112, 4.211, 4.136, 3.959, 3.984, 3.914, 3.926, 3.929, 4.083, 4.025, 3.878, 3.960, 3.951, 3.984, 4.005, 3.983, 4.045, 4.097, 4.020, 3.975, 4.061, 3.940, 4.164, 3.990, 3.988, 4.000, 4.000, 4.005, 3.990, 3.844, 3.973, 3.979, 4.186, 4.119, 4.073, 3.966, 3.980, 4.116, 4.313, 3.942, 3.872, 4.037, 4.058, 4.066, 4.011, 4.066, 4.021, 3.870, 4.080, 4.046, 4.054					
SE	0.089, 0.087, 0.095, 0.094, 0.089, 0.093, 0.089, 0.090, 0.086, 0.089, 0.084, 0.086, 0.087, 0.085, 0.092, 0.091, 0.087, 0.090, 0.090, 0.094, 0.093, 0.093, 0.096, 0.096, 0.088, 0.083, 0.095, 0.090, 0.097, 0.094, 0.088, 0.094, 0.090, 0.090, 0.096, 0.089, 0.088, 0.096, 0.093, 0.093, 0.091, 0.092, 0.087, 0.089, 0.086, 0.095, 0.085, 0.089, 0.088, 0.094, 0.088, 0.089, 0.089, 0.089					
Alpha	.799 (.755-.838)	.832 (.795-.865)	.798 (.754-.837)	.907 (.887-.925)	.905 (.884-.923)	.904 (.883-.923)

(continued)

Table 9. (continued)

Method	Ten Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
	<i>n</i> = 50					
Mean						
<i>M</i>	3.940, 4.030, 4.090, 4.055, 4.040, 3.965, 3.855, 3.995, 3.945, 4.110, 4.205, 4.130, 3.955, 3.990, 3.945, 3.935, 3.955, 4.055, 4.025, 3.895, 3.955, 3.970, 4.000, 4.005, 4.015, 4.045, 4.055, 4.020, 3.980, 4.070, 3.980, 4.105, 3.995, 3.965, 3.980, 3.990, 4.005, 3.990, 3.850, 3.985, 3.980, 4.185, 4.120, 4.060, 3.955, 3.985, 4.120, 4.305, 3.950, 3.890, 4.020, 4.050, 4.050, 4.015, 4.055, 4.020					
<i>SE</i>	0.089, 0.087, 0.094, 0.094, 0.088, 0.092, 0.086, 0.088, 0.084, 0.089, 0.084, 0.086, 0.087, 0.083, 0.091, 0.088, 0.086, 0.087, 0.090, 0.093, 0.092, 0.091, 0.093, 0.095, 0.085, 0.081, 0.092, 0.090, 0.097, 0.094, 0.087, 0.092, 0.089, 0.079, 0.089, 0.088, 0.094, 0.092, 0.091, 0.088, 0.090, 0.084, 0.089, 0.086, 0.094, 0.085, 0.088, 0.087, 0.091, 0.086, 0.088, 0.089					
<i>Alpha</i>	.804 (.761-.842)	.840 (.805-.871)	.830 (.793-.863)	.907 (.887-.925)	.911 (.891-.928)	.912 (.893-.929)
<i>MI</i>						
<i>M</i>	3.929, 4.033, 4.111, 4.079, 4.032, 3.959, 3.849, 4.002, 3.959, 4.111, 4.206, 4.136, 3.954, 3.985, 3.939, 3.919, 3.940, 4.052, 4.025, 3.894, 3.953, 3.943, 3.981, 4.001, 4.004, 4.048, 4.092, 4.020, 3.980, 4.060, 3.979, 4.116, 3.985, 3.986, 3.941, 3.989, 3.996, 3.985, 3.846, 3.988, 3.974, 4.183, 4.110, 4.051, 3.951, 3.985, 4.118, 4.317, 3.947, 3.864, 4.023, 4.079, 4.066, 4.012, 4.057					
		4.011	3.884	4.078	4.036	4.058

(continued)

Table 9. (continued)

Method	Ten Items					
	Weak Association			Moderate Association		
	Missing = 1%	Missing = 5%	Missing = 10%	Missing = 1%	Missing = 5%	Missing = 10%
	<i>n</i> = 50					
SE	0.090, 0.088, 0.098, 0.090, 0.095, 0.092, 0.086, 0.089, 0.088, 0.089	0.100, 0.094, 0.093, 0.097, 0.096, 0.096, 0.103, 0.093, 0.095, 0.102	0.095, 0.098, 0.086, 0.091, 0.089, 0.106, 0.094, 0.098, 0.095, 0.093	0.089, 0.084, 0.086, 0.090, 0.097, 0.095, 0.090, 0.086, 0.095, 0.091	0.088, 0.087, 0.096, 0.091, 0.095, 0.091, 0.086, 0.092, 0.092, 0.098	0.093, 0.088, 0.096, 0.101, 0.099, 0.091, 0.101, 0.090, 0.091, 0.095
Alpha (M, range)	.801 (.757-.840); (.800 [.756-.839]- .802 [.758-.841])	.831 (.794-.864); (.830 [.793-.863]- .832 [.795-.865])	.805 (.762-.843); (.802 [.758-.841]- .808 [.767-.845])	.906 (.885-.924); (.906 [.885-.924]- .907 [.887-.925])	.905 (.884-.923); (.904 [.883-.923]- .906 [.885-.924])	.901 (.879-.920); (.898 [.876-.918]- .902 [.880-.921])

AIA = available item analysis; Mean = mean substitution; MI = multiple imputation. Correlations for AIA data use pairwise deletion. Alphas for MI are summarized across the five imputation runs. Values in *M* and *SE* rows represent item values.

Results and Discussion

Results of the simulation studies in general support the equivalence of AIA and multiple imputation to handle missing data, even in suboptimal conditions (i.e., low sample size, low associations among items, and small number of items) and in the “worst” condition (i.e., $n = 50$, associations = weak, number of items = 3). Differences that did arise between AIA and multiple imputation were generally negligible and did not appear to take on a consistent pattern. Mean substitution appeared to fair adequately at low levels of missingness but produced increasingly inflated correlations among items and alphas as missingness increased to 10% and when there were other negative circumstances (a lower n , lower interitem correlations, or fewer items). Ultimately, though, within the “best” data condition (i.e., $n = 200$, associations = moderate, and number of items = 10), even 10% missing data did not produce notable difference among methods (means, standard errors, and alphas were similar across methods).

General Discussion

The present article was an investigation of the suitability of handling missing data by using AIA rather than conducting imputation methods. Given that research conducted by counseling psychologists typically involves scale- or subscale-level, not item-level, analysis, the present findings indicate that imputation of item-level missing data for most situations encountered by counseling psychologists may not be necessary. Indeed, in investigating the differences between methods of handling missing data, few notable differences were observed. In Study 1, using simulated data, there were no meaningful differences in mean values of items or total scores between methods, no meaningful differences in variances, and negligible differences in the correlation matrices. There were slight differences in Cronbach's alphas, though none of these differences were significant as evident from examination of the confidence intervals. These findings were replicated in Study 2 using real data; again, no significant differences emerged across any of the analyses. In Study 3, it became apparent that mean substitution produces inflated correlations when there is a substantive amount of missing data along with an additional challenge in the data (e.g., few items, weak interitem correlations, or low sample size). In sum, AIA and multiple imputation demonstrated approximately equivalent performance across the studies and across data conditions. Mean substitution does appear to degrade as a method of handling missing data when missingness is high and there are other nonfavorable conditions in the data.

There have been many recent calls for use of advanced methodology to handle missing data. Much of this literature is based on longitudinal research paradigms and time-wave-level missingness at the scale, subscale, and construct level, but is extrapolated to item-level missingness. Such methodology may not be necessary to handle the type and level of missingness that counseling psychologists most frequently encounter, and the results of the present article suggest that calls for use of advanced methodology in handling low-level, item-level missingness may be overgeneralized. At the very least, researchers may feel more confident in the results of older research that used participant mean substitution, despite some suggestions that such procedures are unacceptable (Schlomer et al., 2010).

Researchers may wish to consider whether any imputation is necessary for their analyses. As mentioned, because most research does not involve item-level analyses, using available data (i.e., AIA) to generate mean scores may be a fine solution in most circumstances and may circumvent problems (bias, complexity) associated with both simple methods of imputation (e.g., mean substitution) and complex methods (e.g., multiple imputation). Analyses that do involve item-level data may also use data sets with missingness; exploratory factor analysis can be run with missing data and structural equation modeling may be run using full information maximum likelihood estimation (all provided the level of missingness is low and not MNAR), provided that the missingness prevalence is low and the pattern does not disrupt the covariance matrix. The present article also provides syntax for calculating Cronbach's alpha in SPSS using AIA.

The results of the present article are specific to, and do not extend beyond, item-level missingness at the prevalence common in counseling psychology research. The next level of missingness, scale-level missingness, may require advanced procedures such as multiple imputation. However, although no detailed meta-analysis has been undertaken to examine the extent of scale-level missingness in research, in my experience the extent of missingness at the scale level has been so low that the benefit of running a scale-level imputation procedure would be minimal (perhaps a gain of one or two dozen participants to a sample of several hundred; Parent & Moradi, 2010, 2011a, 2011b). Finally, although the present analyses do not extend to longitudinal analysis, it is important to note that innovations in analysis allow for many longitudinal analyses to be conducted with time-wave level missing data (Duncan, Duncan, & Strycker, 2006) using available data, again circumventing any need to conduct any imputation.

Further research may address some of the implications of this article and advance our understanding of handling missing data. For example,

additional Monte Carlo analyses could be conducted to assess the extent of differences between methods of handling missing data as data sets vary in sample size, number of items, correlations among items, and level of missingness (the possible combinations of these variables is literally infinite, and the present analyses undertook only one particular subset of the infinite possible analyses). Monte Carlo simulations could also be set up with artificial independent and dependent variables, which themselves have differing levels of items, intercorrelations, and missingness, analyzed using a common method such as regression, to further explore the methods. Archival data sets could also be examined to assess the impact of multiple ways of handling item-level missing data in real-world scenarios.

Recommendations

Overall, the present study suggests a number of considerations for counseling psychologists. First, a number of issues related to missing data can be handled before the study even begins. Before collecting data it would behoove researchers to carefully assess scales at the item level and pilot these items with diverse assistants to check for items that may be unanswerable to particular persons (e.g., relationship-related items that contain heterosexist assumptions) to avoid issues of MNAR before they arise. Given the increasing diversity of institutions of higher education, it should not be difficult to recruit a small number of pilots who vary in gender, gender identity/expression, race/ethnicity, sexual orientation, ability status, age, relationship status, or other potentially confounding individual difference characteristics that may affect interpretation of or responses to items. The feedback of these testers can be used to inform scale selection or careful adjustment of items to be inclusive and understandable. As well, researchers may time these pilots to ensure that their survey protocol is not unreasonably long.

Second, after data are collected, researchers may examine their missing data to choose an appropriate level of overall tolerance for missing data (say, 20% missingness on items on any given subscale per participant) and then evaluate whether it is really going to be necessary to analysis to impute missing values or if AIA may be used instead.⁹ Importantly, this decision is not to be made unthinkingly. For instance, imagine a researcher conducting a study using three scales; one has 4 items and the other two have 10 items each. Were the researcher to set an arbitrary tolerance of 20%, this would mean that any participant missing any responses on the first scale would be excluded. This lopsided deletion could be potentially biasing to the results.

Third, researchers should clearly and concisely detail the level of missing data in their studies. These data must be specific; Schlomer et al. (2010) recommended that a “statement giving the range of missing data [be provided]. For example, one could say, ‘Missing data ranged from a low of 4% for Attachment Anxiety to a high of 12% for Depression’” (p. 9). But such a statement is unclear—what level of missingness is being written about? Does the writer mean that *participants* missed 4%-12% of data (as in, e.g., Attachment Anxiety was a 25-item scale and participants missed at most one item)? Or, does it mean that of *all* data points in the $n \times k$ matrix for Attachment Anxiety, 4% were missing (which would mean that the percent is interpretable only by referring to sample size and number of items for each scale, and also omits the maximum amount of missing data by participant)? Instead, I recommend that authors (a) state their tolerance level for missing data by scale or subscale (e.g., “We calculated means for all subscales on which participants gave at least 75% complete data”) and then (b) report the individual missingness rates by scale per data point (i.e., the number of missing values out of all data points on that scale for all participants) and the maximum by participant (e.g., “For Attachment Anxiety, a total of 4 missing data points out of 100 were observed, with no participant missing more than a single data point”). Researchers must also verify that before conducting AIA, they manually inspected missing data for obvious patterns, such as abnormally high missing rates for only one or two items. This can be accomplished easily by simply requesting frequencies output for the items and checking the observed (i.e., not missing) data points for each scale, and ensuring that there are no abnormal spikes in missingness. Although Schlomer et al. (2010) recommended missing data patterns be presented in a table, we typically take researchers at their word about other substantial issues (e.g., meeting normality or homogeneity of variance or linearity assumptions) and so the above seems sufficient.

Fourth, when researchers encounter item-level missing data they may choose to consider using AIA rather than using participant mean substitution or multiple imputation, presuming that missing data levels are assumable to be not MNAR, are at normal levels (e.g., below 10% of all data on each scale, and ideally much lower), and there are no other major complicating concerns (e.g., low sample size, poor internal reliability of scales, scales with fewer than five items). In most circumstances, research does not use item-level responses as variables but rather scale mean scales. For example, most regressions or group difference testing analyses are conducted with scale scores. Thus, AIA means can be generated (e.g., by using the MEAN function in the “compute” command of SPSS) and those are the variables that would be entered into an analysis. Many other analyses take place at the item level but do not require

complete data; factor analysis can be run with missing data, and structural equal modeling using item indicators does not need complete data if full information maximum likelihood estimation is used. Cronbach's alpha is calculated by default using listwise deletion in SPSS; however, it is possible to compute an AIA Cronbach's alpha using a covariance matrix obtained using pairwise deletion (see the appendix). Thus, researchers should think carefully about whether they actually need a complete data set at the item level to carry out their analyses; if they do not, they may simply use AIA.

Fifth, researchers may feel more confident in interpreting studies that have used older methods of handling missing data, such as participant mean substitution, at low-level, item-level missingness, at least in situations where researchers have a sizable sample and reliable scales.

Appendix

* This file contains the syntax capable of generating a Cronbach's alpha internal reliability coefficient for data files with missing data. Presently the syntax is written such that alpha is calculated one "measure" (i.e., the intended unit of analysis, whether it be scale or subscale).

* Differences may arise between version of SPSS; the present syntax were verified to work only in SPSS v.18.

* #####BEGIN#####

* This syntax calls up a correlation matrix with Covariates also displayed (i.e., the CORRELATIONS command), handling data with pairwise deletion. The raw generated file contains extra information (e.g., Correlations, Means, SDs, Ns). These have to be deleted to create a matrix composed of only covariances. The OMS commands accomplish this. The syntax produces the alpha value and c-bar and v-bar for verification purposes.

* THINGS TO CHANGE:

* -Replace "var1...var10" with a list of the variables for which you wish to calculate an alpha

OMS

/select tables

/if commands = ['Correlations'] SUBTYPES = ['Correlations']

/destination format = sav OUTFILE = 'alpha1.sav'.

CORRELATIONS

/VARIABLES= var1 var2 var3 var4 var5 var6 var7 var8 var9 var10

/STATISTICS XPROD

/MISSING=PAIRWISE .

OMSEND .

```

GET FILE "alpha1.sav".
SELECT IF (Var2='Covariance').
SAVE OUTFILE 'alpha2.sav'
  /DROP = Command_ Subtype_ Label_ Var1 Var2 .

GET FILE 'alpha2.sav'.
MATRIX .
GET X /FILE=* /VARIABLES ALL.
COMPUTE k = nrow(x) .
COMPUTE SUMVAR=TRACE(X).
COMPUTE SUMCOV=(MSUM(X)-SUMVAR)/2.
COMPUTE VBAR = SUMVAR/k .
COMPUTE CBAR = SUMCOV/((k*k-k)/2) .
PRINT VBAR.
PRINT CBAR.
COMPUTE AIAALPH = (k*CBAR)/(VBAR+(k-1)*CBAR) .
PRINT AIAalph .
END MATRIX .

```

Note: Please note that if copying and pasting the syntax directly from this manuscript as a pdf, there may be formatting issues that need to be resolved (such as accidental copying of header or footer text, or retention of formatting from the article that is not readable in SPSS syntax [i.e., double and single quotation marks may be pasted as curly-quotes, which will need to be deleted and retyped]).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Available item analysis (AIA) has sometimes been referred to as available case analysis in the past. However, *case* is often used to mean *participant*, and thus the term may be confused with *complete case analysis* (i.e., listwise deletion). AIA is used in the present study to avoid such confusion.
2. There are several methods of mean substitution, participant mean substitution and item mean substitution being two common versions. In participant mean

substitution, each *individual's* mean score for nonmissing items on a scale is substituted for the missing item values on that scale. In item mean substitution, the *sample's* mean score for a particular item is substituted for any missing values on that item. Some other methods of mean substitution, such as scale mean substitution (in which the sample's mean on a scale is used to substitute participants' missing item values on that scale), also exist. In the present analysis, mean substitution refers to participant mean substitution unless otherwise specified. Numerous other reviewers have commented on the unsuitability of item mean substitution; it offers no advantages in terms of computational simplicity over participant mean substitution, and if there is a logical basis to its use, that basis eludes the present author. Unfortunately, many authors do not clearly distinguish which type of mean substitution they mean when they discuss it, blurring discussion on the topic. (This ambiguity about mean substitution [participant vs. scale] is really only a problem at the item level, as mean substitution at the scale level is only possible using the scale mean to substitute for missing values.)

3. The use of the term *random* in the nomenclature is often confusing to newcomers to missing data issues. Schafer and Graham (2002) provide a detailed, thoughtful, and accessible explanation of the term as it applies to types of missingness that need not be reproduced here.
4. In this example, strictly speaking it would be necessary that the survey items be randomized because if they were not, participants are probably more likely to miss items toward the end of the survey because of fatigue, or toward the middle of the survey if the participant accidentally skips a line (and thus missingness would not be random but would be related to item order). A clear example of data missing completely at random (MCAR) when considering item-level missingness is less obvious than when considering longitudinal analysis (a clear example from the latter would be if a single participant were randomly abducted by space aliens on the way to the testing session one year; such a cause of missingness would have nothing to do with the participant her- or himself). In practice, because randomizing every item is often not possible, researchers are probably justified in treating missing data as MCAR so long as no patterns of missingness in item responses emerge.
5. Sometimes for multi-item scales, total scores are created by summing responses rather than calculating a mean for responses. For reasons that will be clear as we proceed, total scale scores are completely incompatible with AIA and mean scores must be used. In any case, mean scores seem to be better anyway, as the value can be intuitively understood as the average response on the scale's response interval without the mental division necessary to understand a summed score.
6. Notably, it does *not* work to generate the average manually, as in typing "(item1+item2+...item_k)/k" into the compute command or syntax, where k is the number of items, as, at least in SPSS, this command would return an empty entry for participants missing any data.

7. The status of (any sort of) mean substitution is debatable, with most commentators preferring other methods. However, very often mean substitution has been discussed without specifying what kind of mean substitution is meant, what level of missingness is being addressed, and the implications of these differences. Dodeen (2003) provided an analysis that indicates that participant mean substitution functions well with item-level missing data compared to multiple imputation, even with fairly substantial amounts of missing data.
8. It is important to note that the AIA alpha is completely dependent on missingness not distorting the covariance matrix; that is, the prevalence of missingness should be fairly low, and it should be assumable to be MCAR.
9. It should be mentioned again here, for emphasis, that the issue of scale-level missingness, including participants who are over the item-level missingness tolerance threshold, is entirely separate from these recommendations and requires different strategies. Consistent with prior work (e.g., Schafer and Graham, 2002), I recommend listwise deletion for situations in which there will be a negligible loss of power (e.g., losing only a small number [say, less than 5%] of participants) and conducting scale-level imputation would result in only a small boost in n , and multiple imputation for all other situations.

References

- Dodeen, H. M. (2003). Effectiveness of valid mean substitution in treating missing data in attitude assessment. *Assessment & Evaluation in Higher Education*, 28, 505-513. doi:10.1080/0260293032000120370
- Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.
- Heinberg, L. J., Thompson, J. K., & Stormer, S. (1995). Development and validation of the Sociocultural Attitudes Toward Appearance Questionnaire. *International Journal of Eating Disorders*, 17, 81-89. doi:10.1002/1098-108X(199501).
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data (Version 1.5-4) [Software]. Retrieved from <http://gking.harvard.edu/amelia/>
- Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology*, 13, 478-487.
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198-1202.
- McCreary, D. R., & Sasse, D. K. (2000). Exploring the drive for muscularity in adolescent boys and girls. *Journal of American College Health*, 48, 297-304. doi:10.1080/07448480009596271
- McKinley, N. M., & Hyde, J. S. (1996). The Objectified Body Consciousness Scale: Development and validation. *Psychology of Women Quarterly*, 20, 181-215. doi:10.1111/j.1471-6402.1996.tb00467.x

- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Parent, M. C., & Moradi, B. (2010). Confirmatory factor analysis of the conformity to feminine norms inventory and development of an abbreviated version: The CFNI-45. *Psychology of Women Quarterly*, 34, 97-109. doi: 10.1111/j.1471-6402.2009.01545.x
- Parent, M. C., & Moradi, B. (2011a). His biceps become him: A test of objectification theory's application to drive for muscularity and propensity for steroid use in college men. *Journal of Counseling Psychology*, 58, 246-256. doi:10.1037/a0021398
- Parent, M. C., & Moradi, B. (2011b). An abbreviated tool for assessing conformity to masculine norms: Psychometric properties of the conformity to masculine norms inventory-46. *Psychology of Men & Masculinity*, 12, 339-353. doi: 10.1037/a0021904
- R Development Core Team. (2008). R: A language and environment for statistical computing [Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Schafer, J. L. (2000). NORM (Version 2.03) [Software].
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147-177. doi:10.1037/1082-989x.7.2.147
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 1, 1-10. doi:10.1037/a0018082
- SPSS Inc. (2010). Statistical Package for the Social Sciences (Version 19) [Software]. Retrieved from <http://www.spss.com>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.
- Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals guidelines and explanations. *American Psychologist*, 54, 594-604.
- Worthington, R. L., Dillon, F. R., & Becker-Schutte, A. M. (2005). Development, reliability, and validity of the Lesbian, Gay, and Bisexual Knowledge and Attitudes Scale for Heterosexuals (LGB-KASH). *Journal of Counseling Psychology*, 52, 104-118. doi:10.1037/0022-0167.52.1.104
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34, 806-838. doi:10.1177/0011000006288127

Bio

Mike C. Parent, MS, is a doctoral candidate at the University of Florida. He earned a master's degree in counseling psychology at the University of Florida and a bachelor's of arts from the University of Manitoba. His research interests focus on intersections of gender, sexuality, and behavioral health, as well as graduate psychology education.