Review of Educational Research
December 2014, Vol. 84, No. 4, pp. 487–508
DOI: 10.3102/0034654314532697

A Review of Missing Data Handling Methods in Education Research

Jehanzeb R. Cheema
University of Illinois at Urbana-Champaign

Missing data are a common occurrence in survey-based research studies in education, and the way missing values are handled can significantly affect the results of analyses based on such data. Despite known problems with performance of some missing data handling methods, such as mean imputation, many researchers in education continue to use those methods as a quick fix. This study reviews the current literature on missing data handling methods within the special context of education research to summarize the pros and cons of various methods and provides guidelines for future research in this area.

Keywords: missing data, imputation, education research, listwise deletion, missing value analysis

Missing data are encountered regularly by researchers in education research. Most large-scale, especially nationally representative, education data sets in the United States contain thousands of individual cases. Unfortunately, such data are seldom complete. Presence of missing data on one or more variables of interest for a proportion of the sample has become a rule rather than an exception in large-scale survey research (Acock, 2005; McKnight, McKnight, Sidani, & Figueredo, 2007; Peng, Harwell, Liou, & Ehman, 2006). A study that contains many variables with a relatively small number of missing values can cause significant attrition in the total effective sample size. For example, a data set containing 500 observations and 10 variables with 10% of the data independently missing on each variable can reduce the effective sample size with listwise deletion to just $0.9^{10} \times 500 = 175$. For many methods of analysis such attrition in sample size can force the researcher to choose alternative methods due to the fall in power as a result of the reduction in *n*.

There are various reasons why data may be missing in surveys. Sometimes it is because the respondents intentionally ignore certain questions. For example, a respondent may not feel comfortable answering questions about his or her salary or his or her criminal record. In some cases, a respondent may genuinely forget to answer a specific question or an interviewer may forget to ask a question. Other reasons for missing data include the inapplicability of a certain question to the respondent or the inability of the respondent to answer a question, for example, due to the respondent's death in a longitudinal study (Allison, 2001; Groves et al., 2004). The current literature offers many missing data imputation methods ranging from very simple, such as mean (or median) imputation where missing data on a variable are substituted simply by the mean (or median) of nonmissing data, to relatively complex procedures, such as maximum likelihood expectation-maximization, which assigns random initial values to missing data and then proceeds to maximize the expectations formed with those initial values in an iterative sequence (Dempster, Laird, & Rubin, 1977).

Past studies that dealt with imputation of missing data go as far back as the 1930s. For example, Wilks (1932) proposed a maximum likelihood method for imputation of missing data in bivariate normal distributions. With the availability of industrial-scale computing, a surge in interest related to missing data imputation occurred during the 1950s and 1960s (Afifi & Elashoff, 1966; Buck, 1960; Edgett, 1956). During these decades several advanced methods for handling missing data, such as linear regression, were introduced. However, lack of statistical packages that could deliver advanced imputation methods and the scarcity of computing resources at the disposal of individual researchers meant that there was little progress in application during that time. There was resurgence in interest in missing data imputation in the early 1980s and 1990s because of the accessible statistical packages that could easily implement such methods and widespread access to computing resources. These developments allowed a large number of researchers unprecedented access to large-scale data sets, and as a result, there was an increase in the number of new missing data imputation methods and their applications (Brick & Kalton, 1996; Hong & Wu, 2011; Zhou, Wang, & Dogherty, 2003).

Although the best way to deal with missing data is to look at each data set individually and determine the requirements for handling its missing data based on the specific features of that data set, education researchers often do not possess the expertise required to identify and implement the best handling method applicable to their specific requirements (McKnight et al., 2007), and they often use simpler but easier-to-implement readymade methods provided by popular software packages, sometimes not even realizing that the use of an incorrect method can introduce serious bias in their estimation results (Allison, 2001; Wayman, 2003).

The reasons for not possessing sufficient expertise to understand and apply advanced statistical methods vary among education researchers. Murtonen and Lethtinen (2003), for example, identified factors such as receiving superficial instruction, difficulty in linking theory with practice, difficulty of content involved, inability to visualize an integrative picture of research, and a negative attitude toward statistics as major hurdles in the learning of statistical concepts in education and sociology. These issues are especially relevant to a majority of graduate students in education who specialize in areas other than quantitative research methods, and who may have taken only a few basic and usually compulsory courses in quantitative methods. Thus, researchers who are not familiar with quantitative methods may not have the expertise to evaluate the appropriateness of a particular missing data handling method in context of their own research (Enders, 2010).

The issue of inappropriately handling missing data was documented in a 2004 review published in the *Review of Educational Research* by Peugh and Enders (2004). These authors investigated reporting practices related to missing data using studies from 23 applied research journals in the fields of education and psychology and reported that their review of 545 studies in these journals identified 229 studies (42%) that had missing data. Of these 229 studies, only 6 studies reported using modern techniques, such as maximum likelihood and multiple imputation, whereas others relied on less sophisticated methods that are known to produce unreliable or biased estimates. These findings show that despite considerable publicity of known biases introduced by traditional missing data handling methods, they are frequently used by researchers in the fields of education and psychology. Reviews of missing data handling methods from the same and later periods as the period reviewed by Peugh and Enders (2004) suggest that the problem of reliance on missing data methods that are known to be inefficient persists among empirical researchers (Bodner, 2006; Enders, 2010; Jeličić, Phelps, & Lerner, 2009; Wood, White, & Thompson, 2004). For example, Jeličić et al. (2009) recently suggested that as much as 82% of longitudinal studies in developmental psychology use missing data handling methods that are known to be error prone.

The present study builds on more recent research (e.g., Young, Weckman, & Holland, 2011) with the aim of providing an improved understanding of the barriers and complexities involved in selection of missing data handling methods that may have discouraged their widespread acceptance within the community of empirical researchers in social science areas. A major aim of this article is to provide recommendations for future research in this area that may help develop clearer guidelines in the choice of missing data handling methods.

Sources and Consequences of Missing Data

Brick and Kalton (1996) and Groves et al. (2004) identified three principal sources of missing data in survey research: noncoverage, total nonresponse, and item nonresponse. Missing data due to noncoverage occur when some population units are left outside the sampling frame and thus have no chance of being selected in the sample. Missing data due to total nonresponse occur when a respondent refuses to respond to any item on the survey (i.e., the entire row in the data set representing that respondent has missing data). Missing data due to item nonresponse occur when a respondent responds to only some of the items on the survey. Of these three sources of missing data, noncoverage and total nonresponse can be addressed by using appropriate sampling weights that are designed to make the sample accurately represent the target population.

However, missing values due to item nonresponse cannot be fixed by using weights. The choice is typically between listwise deletion of cases with item nonresponse, which results in loss of some of the valuable information that those cases did provide, and missing value imputation, which introduces an additional layer of error in parameter estimation because such imputed data, however precisely imputed, is unlikely to exactly match the missing information. Brick and Kalton (1996) identified another source of missing data, partial nonresponse, where respondents provide responses to only a very small number of items. This kind of nonresponse falls in between total nonresponse and item nonresponse and can be corrected by either using weights or missing data imputation, depending on how the researcher wants to treat such nonresponse. Whenever missing data due to item nonresponse are present, the researcher needs to compare the net benefit of more precision at the expense of losing some information with that of using imputed data at the cost of a potentially larger measurement error.

Although one of the concerns with missing data is the attrition in sample size, even in cases where such attrition is not large, the concern remains about whether the sample with complete data is still representative of the target population (Roth, 1994). To see this clearly, consider an extreme example of a sample of 100 employees of whom 5 are managers. If data on all managers are missing, at 5%, the overall sample attrition is small, but the reduced sample that has full information is clearly not representative of its target population because it does not contain any of the managers.

Because the nature and properties of missing data can be very different from the originally observed data, it is important to analyze various methods of treating missing data in order to determine which methods work best under a given set of conditions. The simplest situation is when the missing data can be completely ignored. This strategy is legitimate when the reduced sample size due to missing data still accurately represents the target population. Rubin (1976) termed such missing data as *missing completely at random* (MCAR). An example of MCAR data is when missing data on a variable Y (say, self-efficacy) is unrelated to any of the predictors in the data set (say, gender and race), and when missing data on Y is not related to the value of Y itself (i.e., we do not have a situation where more people with either high or low self-efficacy values are nonrespondents). When data are MCAR, the researcher can perform data analysis procedures on the reduced sample as if it were the full sample without any lack of generalizability. However, if the sample after discarding missing data no longer represents the population of interest, then any findings based on that sample will not be generalizable to the target population, thus restricting the usefulness of such findings. In such a situation, discarding missing data from the original sample will obviously not suffice, and in order to have parameter estimates that are consistent and unbiased, the researcher must resort to missing data imputation.

Finally, an important point to remember is that even in situations where case deletion methods, such as listwise deletion, produce samples that represent their corresponding population well, the power of the analysis is reduced due to the positive relationship between sample size and power. Thus, even in cases where listwise deletion may produce consistent and unbiased estimates of parameters, it may be desirable to use a more sophisticated missing data handling method in order to conserve power.

Missing Data Mechanisms: MCAR, MAR, and NMAR

The appropriateness of a missing data handling method is contextual and depends on the missing data mechanism. One such mechanism, MCAR, has already been discussed. When data are not MCAR, they can either be missing at random (MAR) or not missing at random (NMAR). Data are MAR when the probability of missing data on a variable is unrelated to the value of that variable,

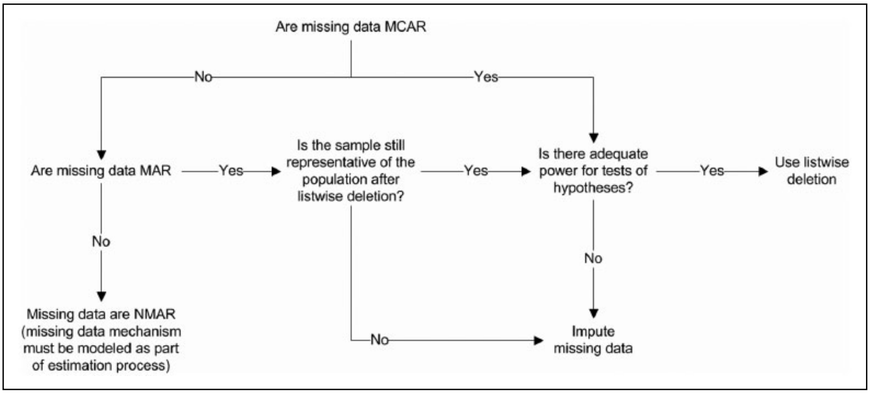

FIGURE 1. *Decision process governing the choice between missing data imputation and listwise deletion.*

itself, but may be related to the values of other variables in the data set. For example, under the MAR assumption, the missing data for Y (say, self-efficacy) may depend on another variable X (say, race) but is not related to the value of Y when X is controlled for. A counterexample is that of salary where salary may be related to race, but even after controlling for race, missing data on salary may still be related to the value of salary, itself (e.g., when individuals with higher salaries are reluctant to report their salaries).

Data are NMAR when the probability of missing data on a variable is a function of the value of that variable, itself (Allison, 2001; Rubin, 1976). An example of an NMAR variable is salary. Individuals with high salaries tend to not report their salaries as compared with those who have lower salaries. Thus, the probability of a missing value for salary is a function of the value of salary, itself. The missing data in this case are thus NMAR. When data are either MCAR or MAR, there is no need to model the missing data mechanism as a part of the estimation process (see Figure 1). In other words, once the missing data handling method has been applied to MCAR and MAR data, any method of analysis can be used with the resulting data set as if it were complete. When data are NMAR, the missing data mechanism needs to be specifically modeled as a part of the estimation process due to the fact that for NMAR data, the parameter estimates of the method of analysis are not independent of the process through which data are missing. Imputation of NMAR data requires extensive a priori knowledge of the missing data process, as the process by which the data are missing cannot be determined from the observed data. For these reasons, missing data handling methods for NMAR data must be tailored to context of the missing data process and cannot be used to construct general guidelines that are applicable under the relatively stronger assumptions of MCAR and MAR (Allison, 2001).

What happens if we treat these mechanisms incorrectly? When NMAR data are incorrectly treated as MCAR or MAR, it means that the missing data process is not being modeled corrected, and parameter estimates will not be accurate.

Similarly, when MCAR and MAR data are incorrectly treated as NMAR, it means that the researcher is introducing unnecessarily more complexity into the handling of missing data. Finally, when MAR data are incorrectly treated as MCAR, the researcher is oversimplifying the handling of missing data and will generate parameter estimates that are not generalizable to the population (Allison, 2001).

It is not always possible to identify statistically the mechanism underlying missing data. One exception is MCAR data for which Little's MCAR test (Little, 1988) exists. This test, which is based on the missing data patterns identifiable from observable data, is implemented as a chi-squared test in general statistical packages with the null hypothesis that missing data are MCAR. The term *missing data pattern* here refers to sorting the data set into groups based on whether a case has a missing or a nonmissing value on a certain variable. For example, in a bivariate data set containing missing data with variables *X* and *Y*, four patterns are possible for a given case: both *X* and *Y* values are missing, only *X* value is missing, only *Y* value is missing, and neither *X* nor *Y* is missing. Take another example where *Y* is achievement score and *X* is gender with equal representation in the population. One can analyze missing data with respect to gender in order to see whether the proportion of missing data vary between males and females. If 20% of achievement scores for males are missing as compared with 80% missing for females, then the researcher cannot consider the data to be MCAR. If, after controlling for gender, we observe that missing data on achievement score are not related to the value of achievement score itself, then the data would be considered MAR.

The MCAR assumption is easier to test than the MAR assumption, the latter being based on the assumption that missing data on a variable are not associated in any way to the values of that variable, itself. In some cases, this assumption is known to be not valid. Again, an example is salary, as individuals with higher salaries tend not to report their salaries (Groves et al., 2004). Thus, in this case the missing data (on salary) are related to the value of missing data (with high salaries missing more often than low salaries), and the missing data mechanism is neither MCAR nor MAR.

Missing Data Handling Methods

The missing data handling methods included in this section have been individually discussed extensively in the literature spanning the past 30 years. As any attempt to reproduce that discussion in its entirety is an endeavor worthy of a textbook, only nontechnical descriptions of missing data handling methods are provided here. Readers who are interested in detailed technical aspects of missing data handling, including mathematically intensive proofs and theorems, and application of these methods in various fields including education are referred to Madow, Nisselson, and Olkin (1983); Madow and Olkin (1983); Madow, Olkin, and Rubin (1983); Rubin (1987); Jones (1996); Groves, Dillman, Eltinge, and Little (2002); and Peugh and Enders (2004).

Missing data handling methods can be divided into two broad categories. The first category includes methods that rely on discarding a portion of the sample whereas the second category includes methods that replace missing data with imputed values.

Case Deletion Methods

Two commonly used methods that work by discarding cases with incomplete information are listwise deletion and pairwise deletion.

Listwise Deletion

As indicated previously, this method simply discards observations with missing values on one or more variables of interest. For this reason, listwise deletion is also known as complete case method (McKnight et al., 2007). In some statistical packages such as SPSS, listwise deletion is the default method and is supported for a large number of procedures.

Pairwise Deletion

This method is similar to listwise deletion with the difference that only cases with missing data on variables involved in a statistical procedure are removed. For example, if X, Y, and Z are three variables and one case in the data set has a missing value on Z, then a procedure such as correlation will use all n observations to calculate r_{XY} but only $n-1$ observations to calculate r_{XZ} and r_{YZ}. This method is different from listwise deletion, which would have used $n-1$ cases for all three correlations. Pairwise deletion is also known as available case method (McKnight et al., 2007). Pairwise deletion has limited application in many education studies for two reasons. First, in models involving only one or two variables, such as one sample t test, independent samples t test, and one-way ANOVA (analysis of variance), listwise deletion and pairwise deletion result in the same outcome. Second, for methods involving more than two variables, such as two-way ANOVA and multiple regression, many general statistical packages popular with education researchers do not support pairwise deletion.

Imputation-Based Methods

The following methods involve replacing missing data with their imputed counterparts.

Mean Imputation

This method involves replacing missing data on a variable with the mean of non-missing data for that variable. Mean imputation is also known as marginal mean imputation, because the effect of other variables is not partialed out of the mean used to replace missing values (Allison, 2001). This method is one of many methods based on replacing missing data on a variable with a measure of central tendency for that variable. Depending on the measure of central tendency used, this method can take other names such as median imputation or mode imputation (Chen, Jain, & Tai, 2006). The mean imputation method is known to decrease the standard error of the mean, thus increasing the risk of rejecting the null hypothesis when it should not be rejected, and is seldom recommended. The concern with decreasing the standard error also occurs for imputation based on other measures of central tendency.

Regression Imputation

This method involves regressing the variable with missing data on all other variables in the data set using cases that have full information for those variables.

Therefore, this method allows computation of predicted values (or conditional means) of the variable with missing data, given values of other variables. For this reason, regression imputation is also known as conditional mean imputation. However, because this method does not specifically model the natural variation in missing data, it produces biased standard errors of parameter estimates (Allison, 2001).

Maximum Likelihood Expectation-Maximization (EM) Imputation

Maximum likelihood is a mathematical procedure that can be used to find one or more parameters of a statistical model that, for the observed data, maximize the observed likelihood distribution. The EM imputation is a maximum likelihood–based iterative method that involves two steps. In the first step, initial values (often marginal means) are assigned to missing data. In the second step, expectations formed with those initial values are maximized. This EM cycle is then repeated again and again until the imputed values converge based on predetermined convergence criteria. The EM imputation method produces unbiased standard errors of parameter estimates (Salkind & Rasmussen, 2007).

Multiple Imputation

Multiple imputation is an advanced imputation method that simulates the natural variation in missing data by imputing such missing data several times, thus producing several complete data sets. The sets of estimates produced by these various complete data sets are then combined into a single set of estimates by averaging. Because multiple imputation specifically models the natural variation in missing data, the standard errors of parameter estimates produced with this method remain unbiased (Rubin, 1987).

Hot Deck Imputation

Although the hot deck method is used extensively in social science research, it tends to be relatively less developed conceptually compared with other missing data imputation methods. This method involves imputing missing data on a variable for a given case by matching that case with other cases in the data set on several other key variables that have complete information for those cases.

There are many variations of this method, but one that allows for modeling of natural variability in missing data involves selecting a pool of all cases, called the donor pool, that are identical to the case with missing data (i.e., the recipient) on a number of variables and then choosing one case randomly out of that pool. The data on this randomly chosen case are then used to replace the missing value on the case with incomplete data. Another variation of the hot deck imputation method involves substituting the closest donor neighbor rather than selecting one donor from a pool of donors. This method ignores the variability in missing data. Studies involving a large number of variables require large sample sizes for hot deck imputation to work best so that cases may be matched on many variables. To select an appropriate donor, the recipient is matched with similar cases on all possible variables and not just those that are included in the method of analysis. The two requirements for selecting an external variable for use in hot deck imputation are (a) whether the external variable is associated with the variable being imputed and (b) whether the external variable is associated with the dichotomous variable that indicates whether or not a value is missing.

Two additional variations of the hot deck method are weighted sequential hot deck and weighted random hot deck. Weighted sequential hot deck is designed to avoid the problem of same donor being matched with a large number of recipients by putting a limit on the number of times a donor may be selected. In contrast, weighted random hot deck is a variation that does not limit the number of times a donor may be selected, but the donors are selected at random from the donor pool (Andridge & Little, 2010). It should be noted that hot deck imputation can be used with other imputation methods such as multiple imputation where results from several imputed data sets, each based on hot deck imputation, are combined to obtain aggregate parameter estimates.

Dummy Variable Adjustment

This method involves constructing a separate dummy variable for each variable with incomplete data. The dummy variable is specified to take a value of 1 when the value of corresponding variable is not missing and 0 otherwise. The missing data on each variable, X, is then replaced with a constant (often the marginal mean of X), and the dependent variable is regressed on all other variables including the dummy variables. The coefficient on the dummy variable corresponding to a variable X obtained in this way can be interpreted as the deviation of mean value of the dependent variable for missing data on X from the mean value of nonmissing data on X. Although simple to understand and apply, this method is known to produce biased parameter estimates (Jones, 1996).

Zero Imputation

This method simply replaces missing values on a variable with zeroes. The simplicity of this method's application is offset by its very limited usefulness. The replacement of missing data with zeroes makes conceptual sense in very specific circumstances, for example, when dealing with missing achievement scores where a missing value can be reasonably assumed to occur because the respondent did not know the correct answer. However, this method produces biased parameter estimates whenever other reasons (e.g., anxiety or fatigue in the preceding example) are responsible for the occurrence of missing data (McKnight et al., 2007).

Single Random Imputation

This method can be thought of as a compromise between regression imputation and multiple imputation. The method involves regressing the variable with missing values on all other variables for cases with complete information, augmenting the resulting predicted values with random draws from the residual distribution of the regressand, and then using those augmented values to replace missing data. However, since the postimputation data set is treated as a complete data set, the resulting standard errors of parameter estimates tend to be underestimates of their population counterparts (Allison, 2001).

Last Observed Value Carried Forward (LOCF)

This method is commonly used in longitudinal studies and involves replacing the missing value on a variable at a certain point in time with the value of that variable from the immediately preceding time period. LOCF is known to produce biased parameter estimates with lower standard errors.

Summary

The imputation methods discussed in preceding paragraphs are the dominant methods in education research. The first four imputation methods are all supported by, and easily implemented in, most general software packages without need for advanced programming skills. The subsequent imputation methods, some of which are variations of these four methods, are not generally supported by regular packages. It is useful to note here that regardless of which missing data imputation method is used, all variables available in the data set (and not just those that are employed in the method of analysis) should be used for imputation. This rationale is based on the logic that variables that are not directly used in the method of analysis still contain useful information about the case (or observation) that can improve imputation results (Allison, 2001). In the next section, I review studies that have compared the performance of these methods.

Comparative Performance of Missing Data Handling Methods

Several researchers have evaluated the comparative performance of missing data handling methods. One of the earliest studies that compared alternative missing data handling methods was by Afifi and Elashoff (1966), who compared listwise deletion, mean imputation, and regression imputation with simple linear regression as the methods of analysis. This article was conceptual and did not use any simulated or empirical samples. They concluded that none of the missing data handling methods was uniformly good. Their general finding was that mean imputation works best when correlations among regression variables are low, listwise deletion works best when such correlations are moderate, and regression imputation works best when correlations are high.

In another earlier study, Haitovsky (1968) compared the performance of listwise deletion (also called the classical method by that author) and pairwise deletion in context of linear regression. This study simulated eight complete samples of n = 1,000 with a portion of each sample designated as missing. These eight samples differed from each other with respect to the total number of variables, the distribution of predictors, the variance-covariance matrix, and the variability in the dependent variables relative to the variability in the error term. A comparison between the regression parameter estimates obtained from the two missing data handling methods based on reduced samples with the parameter estimates of full samples revealed that listwise deletion performed best under all conditions except when the proportion of missing data was very large (in excess of 0.9) or when the data were missing in a very highly nonrandom pattern.

In the field of behavioral research, Graham, Hofer, and MacKinnon (1996) used simulated data to evaluate missing data handling methods, including pairwise deletion, mean imputation, single random imputation, multiple imputation, and several variations of maximum likelihood imputation. Their findings suggested that under the MCAR assumption, maximum likelihood and multiple imputation methods performed better than pairwise deletion, which in turn was superior to mean imputation. However, with the exception of maximum likelihood methods, all methods developed bias in parameter estimation when the MCAR assumption was relaxed. With a sample size of 1,945, the authors used missing data percentages of 5.7% and 11.6%, and their findings suggested that an increase in proportion of missing data produced larger bias in estimation. In a similar study, Wayman (2003) used 19,373 cases from a national reading test assessment that had approximately 15% missing data and four variables to compare the performance of listwise deletion, mean imputation, and multiple imputation. Based on sample means and their standard errors for normalized test scores, he concluded that multiple imputation performed the best, followed by listwise deletion and mean imputation. However, this study did not consider effects of changes in sample size, proportion of missing data, and method of analysis.

Among the studies that looked at treatment of missing data in education, Peugh and Enders (2004) reviewed dominant missing data handling methods in education research, which they categorized into traditional and modern missing data techniques. The traditional techniques that they reviewed included listwise deletion, pairwise deletion, mean imputation, and regression imputation, and the modern techniques included maximum likelihood imputation and multiple imputation. Based on a review of literature, the authors concluded that maximum likelihood and multiple imputation methods, due to their statistical properties that allow these methods to generate consistent and unbiased parameter estimates for any method of analysis, had an almost unqualified superiority over their traditional counterparts. Peugh and Enders (2004) also provided a comprehensive, step-by-step illustration of maximum likelihood and multiple imputation methods using a longitudinal data set that can be of value to researchers interested in implementing these methods on their own.

In a study in education, Peng et al. (2006) compared the performance of two maximum likelihood methods (full information and expectation-maximization) and multiple imputation with listwise deletion using two real-world samples of 1,302 and 517 in the context of path analysis and logistic regression, respectively. They reported that magnitudes and/or signs of parameter estimates, p values in tests of hypotheses, and power can vary significantly depending on the missing data method employed. Their general conclusion was in accordance with that of Peugh and Enders (2004): Advanced methods such as maximum likelihood imputation and multiple imputation are superior to listwise deletion when data are MAR. Unfortunately, Peng et al. (2006) used samples of different sizes with different methods of analysis based on different missing data handling methods. For this reason, the interrelationships among these three factors could not be evaluated for this study.

In another recent applied study, Yesilova, Kaya, and Almali (2011) compared several variations of the mean imputation and hot deck imputation methods under the MAR assumption using a sample of size 4,464 and 7 variables. The authors found that performance of hot deck imputation was superior to mean and median substitution methods in terms of parameters estimates, standard errors, and correlations between real and imputed data. This study did not evaluate the effect of sample size and proportion of missing data on parameter estimates and their standard errors.

In addition to the studies that have compared performance of various missing data handling methods, a very small number of studies have attempted to quantify the effects of sample size and proportion of missing data on the performance of missing value imputation method. Haitovsky (1968) noted that mean imputation in linear regression can seriously bias the parameter estimates. The main reason for this bias is because even though the overall mean does not change with mean imputation, the standard error of the mean can become considerably smaller depending on the proportion of missing data. For a variable *Y* with *n* observations, *k* of which are missing but replaced by the mean of $n-k$ nonmissing observations, squared standard error of the mean can be expressed as follows:

$$SE_M^2 = \frac{\sum_{i=1}^{n-k}\left(X_i - M\right)^2 + \sum_{i=n-k+1}^{n}\left(X_i - M\right)^2}{n\left(n-1\right)}. \quad (1)$$

Since with mean imputation each imputed value exactly equals *M*, all deviations of imputed values from the mean are zero, that is, $\sum_{i=n-k+1}^{n}\left(X_i - M\right)^2 = 0$, which causes SE_M^2 to become smaller. Thus, SE_M^2 is always biased as long as there is even a single imputed value that in reality deviates from the mean.

It can be easily seen from the expression for SE_M^2 that it is directly proportional to the number of missing values, *k*. As *k* increases with *n* held constant (i.e., proportion of missing data increases), $\sum_{i=1}^{n-k}\left(X_i - M\right)^2$ decreases, causing SE_M^2 to decrease as well. The opposite effect occurs when *k* decreases. The effect of sample size on SE_M^2 with proportion of missing data held constant is relatively less straightforward to see. Keeping *k*/*n* constant when *n* increases requires that *k* be increased at the same rate as *n*. Thus, a *t* times increase in both *k* and *n* would cause an increase in *n* while keeping *k*/*n* constant. Assuming that new observations obtained by increasing the sample size come from the same distribution, it is reasonable to expect that the deviations of such new observations from *M* are similar to those for original observations. Given this assumption, $\sum_{i=1}^{n-k}\left(X_i - M\right)^2$ increases at the rate *t* and the squared standard error of the mean for the new sample, $SE_M^{2\,\prime}$, takes the following expression:

$$SE_M^{2\,\prime} = \frac{t \cdot \sum_{i=1}^{n-k}\left(X_i - M\right)^2}{tn\left(tn-1\right)} = SE_M^2 \cdot \frac{\left(n-1\right)}{\left(tn-1\right)}. \quad (2)$$

For large *n* values, $n-1$ and $tn-1$ can be approximated by *n* and *tn*, respectively, and the expression for $SE_M^{2\,\prime}$ simplifies to $SE_M^{2\,\prime} = SE_M^2 / t$, or in standard error units, $SE_M{}' = SE_M / \sqrt{t}$. In other words, the change in standard error of the mean due to an increase in sample size is inversely proportional to the square root of the rate at which that sample size increases. The tendency of standard error of the mean to become biased has been noted by other authors, such as Gurland and

Tripathi (1971), who have provided a correction factor when n is small. The bottom line is that whenever arithmetic mean is used to substitute for missing data, the resulting parameter estimates are biased.

Whereas mean imputation is one of the least mathematically sophisticated missing data imputation methods that does not take into consideration any random variability among the missing data values, multiple imputation is at the other extreme, being one of the most sophisticated imputation methods that specifically models random variation in missing data. Rubin (1987) provided a mathematical formula relating the proportion of missing data to imputation efficiency for the multiple imputation method. Efficiency here relates to error in measurement due to missing data. As proportion of missing data increases, the efficiency decreases. However, some of that decrease can be offset by increasing the number of imputations. If m is the number of times a complete data set is generated and γ is the proportion of missing data, then, given sample size n, relative efficiency of imputation, E, measured in units of variance, can be shown to be an inverse function of proportion of missing data and a direct function of the number of imputations.

$$E = \frac{m}{m + \gamma}. \tag{3}$$

When the data set does not contain any missing data, γ takes the value of 0 and E = 1, which essentially means that no missing data are imputed and efficiency is 100%. When m is kept constant and γ increases, E decreases. For example, given m = 1, if γ increases from 0 to 0.05, E falls from 1 to 0.95, signifying a 5% decrease in efficiency. However, the rate of change in imputation efficiency is slower than the rate of change in the proportion of missing data. For instance, an increase in γ from 0 to 0.2 reduces imputation efficiency by less than four times the decrease observed for an increase in γ from 0 to 0.05 $(\Delta E = .17 \neq .20)$. In other words, the marginal effect of a 1% increase in γ on E decreases as γ increases. When γ is held constant, E becomes a direct function of m. For example, when γ is held constant at 0.05, the absolute change in E, as m changes from 1 to 2, is 0.02, or a 2% gain in efficiency.

The relationship between proportion of missing data and efficiency as provided by Rubin (1987) is important because it shows that for a large n, an increase in proportion of missing data can be compensated by an increase in the total number of multiple imputations. Thus, it is up to the researcher to determine how much efficiency she or he wants at the cost of computational complexity. The multiple imputation method itself does not impose any restrictions in this respect.

To clarify the effect of m and γ, E was calculated for selected values of m and γ (see Table 1). The relative efficiency calculations presented in Table 1 show that, for large samples, relative efficiency can be reasonably high $(E \geq .98)$, with just one or two imputations when proportion of missing data is low $(\gamma \leq .05)$, and with four to eight imputations when proportion of missing data is high $(.05 < \gamma \leq .20)$. These figures support the recommendation of multiple imputation as a universal imputation method as advanced by Young et al. (2011).

TABLE 1

Asymptotic relative efficiency of multiple imputation at selected number of imputations (m) *and proportion of missing data (γ)*

	γ				
m	0.01	0.02	0.05	0.10	0.20
1	0.99	0.98	0.95	0.91	0.83
2	1.00	0.99	0.98	0.95	0.91
3	1.00	0.99	0.98	0.97	0.94
4	1.00	1.00	0.99	0.98	0.95
5	1.00	1.00	0.99	0.98	0.96
6	1.00	1.00	0.99	0.98	0.97
7	1.00	1.00	0.99	0.99	0.97
8	1.00	1.00	0.99	0.99	0.98
9	1.00	1.00	0.99	0.99	0.98
10	1.00	1.00	1.00	0.99	0.98

It should be noted that the multiple imputation method works by imputing several sets of complete data sets. Parameter estimates are then calculated from each data set separately and the results averaged for all data sets. This process is in contrast to simply averaging the values of the various data sets and calculating a single set of parameter estimates based on that averaged data. Unlike the former method, this latter approach treats the averaged data set as a complete data set and thus does not allow for any variation in the parameter estimates and test statistics based on those estimates. For $i = 1, 2, \ldots, m$ imputations of a data set, parameter estimates and their corresponding variances from each data set, denoted by Q_i and U_i, can be used to aggregate estimation results as follows (Peugh & Enders, 2004; Rubin, 1987). The parameter estimates can be averaged to obtain the aggregate parameter estimate.

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} Q_i. \tag{4}$$

The calculation of standard error of $\bar{Q}$ requires two components, the within-imputation variance, $\bar{U}$, and the between-imputation variance, B, which can be combined together to obtain the variance for $\bar{Q}$, T. These can be calculated using the following expressions:

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m} U_i, \tag{5}$$

$$B = \frac{1}{m-1}\sum_{i=1}^{m}\left(Q_i - \bar{Q}\right)^2, \tag{6}$$

$$T = \bar{U} + \left(1 + \frac{1}{m}\right)B. \tag{7}$$

The distribution of $\bar{Q}/\sqrt{T}$ approximately follows the t distribution and can be evaluated against the critical t value by using the following expression for degrees of freedom, *df*, calculation:

$$df = (m-1)\left\{1 + \frac{m\bar{U}}{(m+1)B}\right\}^2. \tag{8}$$

Although multiple imputation works very well when n is large, it can produce biased estimates when n is very small. Kim (2004) has provided the exact magnitude of this bias using Monte Carlo simulation with 50,000 samples and 5 imputations for a $2 \times 3 \times 2$ factorial design. For instance, he showed that when sample size decreases from 200 to 20, the variance of the multiple imputation parameter estimators can increase by a factor of 10 or more when the proportion of missing data ranges between 0.2 and 0.6. This study proposed a new missing data imputation method based on multiple imputation with more desirable statistical properties than Rubin's (1987) multiple imputation method for the specific case when sample size is very small $(n \leq 20)$.

In a comparative study, Raymond and Roberts (1997) used simulation to generate multivariate data comprising four variables and with samples of size 50, 100, and 200. They examined these in a linear multiple regression context after simulating randomly missing data at 2%, 6%, and 10%. They tested several missing data handling methods such as listwise deletion, mean imputation, and two variations of regression imputation. The authors found that, in terms of deviation from true parameter values, regression-based missing data handling methods worked best whereas listwise deletion turned out to be the worst option. Although Raymond and Roberts (1997) considered several sample sizes, proportions of missing data, and missing data handling methods, they looked at only one method of analysis, linear multiple regression, and collapsed their findings over sample size; thus, the effect of sample size on performance of missing data handling methods could not be evaluated. Based on their analysis, these authors recommended that whenever percentage of missing data exceeds 5%, more than one missing data handling method should be used, as parameter estimation results can be very different under various methods of analysis when so much data are missing.

Alosh (2009) simulated a longitudinal count data set to generate samples of size 30 and 60 with missing data percentages of, respectively, 10% and 20% in context of a log-linear model under MCAR, MAR, and NMAR assumptions. The primary aim of this study was to compare the effect of missing data mechanisms, rather than missing data handling methods, on parameter estimates. For this reason, this study did not focus on missing data imputation, and case deletion was employed as the primary method for handling missing data. Only the MAR condition was evaluated both under case deletion and under an imputation method, LOCF. Alosh's primary finding was that under MAR and MCAR assumptions, the sample estimates are very close to their true values with a maximum percent bias of approximately 6%, whereas estimates obtained under NMAR assumption showed the largest biases and were clearly inferior to those obtained under MCAR and MAR assumptions. Alosh also found that MAR data and MAR-LOCF data behaved similarly in the context of estimation. Parameter estimates reported in this study suggested that, on average, a decrease in proportion of missing data reduced estimation bias. However, an increase in sample size increased bias, a result that seems counterintuitive, but may be reasonable considering the longitudinal nature of data used in this study and the fact that only two sample sizes were considered, a number that is too small to establish a trend.

Knol et al. (2010) conducted a recent simulation study to investigate missing data methods. They used an empirical sample of n = 1,338 to create 1,000 subsamples of size 1,025 to test the performance of three missing data handling methods, listwise deletion, dummy variable adjustment (also known as missing indicator method), and multiple imputation. This study simulated missing data percentages of 2.5%, 5%, 10%, 20%, and 30% under the assumptions of MCAR and MAR and their analytical procedure, given the categorical nature of their dependent variable, involved evaluation of odds ratios. The authors found that for their sample, the smallest deviations from true parameters were obtained with multiple imputation. Dummy variable adjustment had larger deviations and listwise deletion was the most error-prone method. This study did not consider the effect of sample size on performance of missing data handling methods and did not use any other method of analysis.

A recent review of the literature by Young et al. (2011) surveyed the performance of several missing data handling methods based on findings reported in past research. These included listwise deletion, pairwise deletion, mean imputation, mode imputation, regression imputation (both simple and multiple), hot deck imputation, expectation-maximization imputation, and multiple imputation, among others. After reviewing dozens of studies, especially those from ergonomics, the authors concluded that there was no single missing data handling method that was the best in all situations. The performance of a given method was found to be dependent on factors such as proportion of missing data in the sample, sample size, distributions of variables in the sample, and the relationships between those variables. The authors concluded that even in cases where a missing data handling method worked well for a given data set, there was no guarantee that the same method would also work well for similar data sets. Furthermore, they warned that applying a missing data handling method incorrectly without due regard to the missing data mechanism can result in biased parameter estimates.

Young et al. (2011) summarized the recommendations of various studies to provide the following guidelines: when less than 1% of the data are missing, the effect of missing data handling methods is trivial; for 1% to 5% missing data, simple methods such as listwise deletion and regression imputation work well; for 5% to 15% missing data, sophisticated methods, such as multiple imputation, should be selected; and when missing data exceeds 15%, imputation results are largely meaningless regardless of the imputation method used because very little can be said about the mechanism through which data are missing. The authors found that there was a very limited number of studies that discussed the gain in power as a direct result of the imputation method used, and they recommended more research in this direction to allow derivation of a set of rules that can be used to select the best imputation method under a set of given conditions. Finally, the authors suggested that although multiple imputation may not be best for all situations, it is generally best or second best in most situations. Even when it is the second best method, the relative difference in performance is small relevant to the best method. For this reason, they generally recommended multiple imputation as a safe choice for an imputation method, given its reliable performance even in cases where the proportion of missing data was large.

An Illustrative Example

To elaborate how estimation results can be affected by the choice of missing data handling method, we use a simple illustration based on a sample of 30 observations drawn from a large-scale high school assessment. The variable of interest here is achievement score, which is standardized and presented on a scale of 0 to 100. Original values of this variable are presented in the second column of Table 2. We randomly selected 10% of the cases from this data set and set them as missing to compare performance of five missing data handling methods that represent a fairly representative subset of methods discussed in this review: listwise deletion, mean imputation, regression imputation, maximum likelihood imputation, and multiple imputation. Summary statistics for each missing data handling method are presented in the last four rows of Table 2. A quick glance at imputation results and summary statistics show that although parameter estimates such as the mean do not fluctuate much from their original full-sample value, the corresponding standard error values fluctuate anywhere between 93% and 104% in this example. Since standard error has a direct bearing on results of tests of hypotheses, this illustration clearly shows how the choice of missing data handling method can push results of such tests toward either significance or nonsignificance depending on which missing data handling method is employed.

Summary and Directions for Future Research

A number of recent studies have compared performance of missing data handling methods in various fields including education and psychological research. Relatively fewer studies have evaluated the effect of factors such as sample size and proportion of missing data on the performance of such methods. Although they provide valuable information, the findings reported in these few studies are difficult to synthesize because they differ in their use of empirical samples, simulation techniques, methods of analysis, and selected rates of change in sample size and proportion of missing data. For this reason, it is not easy to use those findings to construct general guidelines that can help in the selection of an appropriate missing data handling method while encompassing a reasonably large subset of various possible combinations of sample size, proportion of missing data, method of analysis, and missing data handling method. One possible reason why such a task has not been attempted in studies targeted for publication in scientific journals is that the extent of work involved makes such a task more suitable for a book or a dissertation rather than a journal article. We do, however, believe that such a task, albeit difficult, can be accomplished by a series of carefully summarized articles. A second reason is that even though the methodological awareness of missing data issues can be traced back over several decades, such awareness has

TABLE 2

An illustration of performance of various missing data handling methods

			Imputation							
n	Original values	Listwise deletion	Mean	REG	ML-EM	MI1	MI2	MI3	MI4	MI5
1	39	39	39	39	39	39	39	39	39	39
2	47	47	47	47	47	47	47	47	47	47
3	50	50	50	50	50	50	50	50	50	50
4	45	45	45	45	45	45	45	45	45	45
5	72	72	72	72	72	72	72	72	72	72
6	35	35	35	35	35	35	35	35	35	35
7	48	48	48	48	48	48	48	48	48	48
8	48	48	48	48	48	48	48	48	48	48
9	29		49.96	26.81	43.67	60.63	34.94	43.07	46.1	24.74
10	56	56	56	56	56	56	56	56	56	56
11	67	67	67	67	67	67	67	67	67	67
12	60	60	60	60	60	60	60	60	60	60
13	6	6	6	6	6	6	6	6	6	6
14	44	44	44	44	44	44	44	44	44	44
15	72	72	72	72	72	72	72	72	72	72
16	39	39	39	39	39	39	39	39	39	39
17	59		49.96	59.96	48.25	49.56	28.17	52.68	52.4	34.41
18	27	27	27	27	27	27	27	27	27	27
19	39	39	39	39	39	39	39	39	39	39
20	49		49.96	69.65	53.89	69.83	45.81	36.50	55.22	64.36
21	29	29	29	29	29	29	29	29	29	29
22	63	63	63	63	63	63	63	63	63	63
23	49	49	49	49	49	49	49	49	49	49
24	50	50	50	50	50	50	50	50	50	50
25	71	71	71	71	71	71	71	71	71	71
26	55	55	55	55	55	55	55	55	55	55
27	59	59	59	59	59	59	59	59	59	59
28	71	71	71	71	71	71	71	71	71	71
29	63	63	63	63	63	63	63	63	63	63
30	45	45	45	45	45	45	45	45	45	45
M	49.53	49.96	49.96	50.18	49.83	50.97	48.60	49.38	50.09	49.08
SD	15.35	15.59	14.76	15.91	14.83	15.31	15.51	15.02	14.81	15.95
SE	2.80	3.00	2.69	2.90	2.71	2.79	2.83	2.74	2.70	2.91
Rel SE	100	93	104	96	104	100	99	102	104	96

Note. REG = regression, ML-EM = maximum likelihood expectation-maximization, MI = multiple imputation. *Rel SE* = relative standard error. Values for Cases 9, 17, and 20 were imputed for illustration. Averaged *M* and *SE* for MI were 49.62 and 15.36, respectively.

only recently expanded to a level where reviewers and editors specifically ask for disclosure of missing data treatment in peer-reviewed articles.

Our review of relevant literature indicates the need for a comprehensive study that can provide general guidelines to assist general education researchers in the selection of appropriate missing data handling methods, by using uniform empirical and simulated samples and uniform rates of change in sample size and proportion of missing data. By keeping all these factors constant, any observed differences in the performance of missing data handling methods can more or less be attributed directly to the relative efficiency of those methods. Such a study can be even more effective were it to provide differential performance estimates for various combinations of sample size, proportion of missing data, and missing data handling method, separately for methods of analysis commonly used in education research. The differential performance estimates could be used by education researchers to apply corrections to any expected biases in prior studies that involved incorrect use of missing data handling methods and consequently reported inaccurate parameter estimates.

Another promising area of research is the development and application of combinations of missing data handling methods, similar to the example of hot deck imputation used in conjunction with multiple imputation that we provided earlier in this text. Since each missing data handing method has unique pros and cons, development of new techniques that can combine two or more of these methods while retaining only their positive features could be a worthwhile avenue of inquiry. Such a development would help in our understanding of missing data handling and may result in greater estimation efficiency.

Finally, in the spirit of Wilkinson and the Task Force on Statistical Inference (1999) and given the prevalence of missing data in samples used for empirical inquiry in education and psychological research, we suggest inclusion of discussion of missing data handling as a standard part of the method section in peer-reviewed publications. This discussion can be a short description similar to that of the sample and participants but would be valuable to potential readers in clarifying the issues of sample representativeness (generalizability) and adequacy of statistical power.

References

Acock, A. (2005). Working with missing values. *Journal of Marriage and Family*, *67*, 1012–1028. doi:10.1111/j.1741-3737.2005.00191.x

Afifi, A., & Elashoff, R. (1966). Missing observations in multivariate statistics: I. Review of the literature. *Journal of the American Statistical Association*, *61*, 595–604. doi:10.2307/2282773

Allison, P. (2001). *Missing data*. Thousand Oaks, CA: Sage. doi:10.4135/9781412985079

Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *Journal of Biopharmaceutical Statistics*, *19*, 1039–1054. doi:10.1080/10543400903242787

Andridge, R., & Little, R. (2010). A review of hot deck imputation for survey nonresponse. *International Statistical Review*, *78*, 40–64. doi:10.1111/j.1751-5823.2010.00103.x

Bodner, T. (2006). Missing data: Prevalence and reporting practices. *Psychological Reports*, *99*, 675–680. doi:10.2466/PR0.99.3.675–680

Brick, J., & Kalton, J. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, *5*, 215–238. doi:10.1177/096228029600500302

Buck, S. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society: Series B, Methodological*, *22*, 302–306.

Chen, S., Jain, L., & Tai, C. (2006). *Computational economics: A perspective from computational intelligence*. Hershey, PA: Idea Group. doi:10.4018/978-1-59140-649-5

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B, Methodological*, *3*, 1–38.

Edgett, G. (1956). Multiple regression with missing observations among the independent variables. *Journal of the American Statistical Association*, *51*, 122–131. doi:10.2307/2280808

Enders, C. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Graham, J., Hofer, S., & MacKinnon, D. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197–218. doi:10.1207/s15327906mbr3102_3

Groves, R., Dillman, D., Eltinge, J., & Little, R. (2002). *Survey nonresponse*. New York, NY: John Wiley.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: John Wiley.

Gurland, J., & Tripathi, R. (1971). A simple approximation for unbiased estimation of the standard deviation. *The American Statistician*, *25*(4), 30–32. doi:10.2307/2682923

Haitovsky, Y. (1968). Missing data in regression analysis. *Journal of the Royal Statistical Society: Series B, Methodological*, *30*, 67–82.

Hong, T., & Wu, C. (2011). Mining rules from an incomplete dataset with a high missing rate. *Expert Systems with Applications*, *38*, 3931–3936. doi:10.1016/j.eswa.2010.09.054

Jeličić, H., Phelps, E., & Lerner, R. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195–1199. doi:10.1037/a0015665

Jones, M. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, *91*, 222–230. doi:10.2307/2291399

Kim, J. (2004). Finite sample properties of multiple imputation estimators. *Annals of Statistics*, *32*, 766–783. doi:10.1214/009053604000000175

Knol, M., Janssen, K., Donders, A., Egberts, A., Heerdink, E., Grobbee, D., . . . Geerlings, M. (2010). Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: An empirical example. *Journal of Clinical Epidemiology*, *63*, 728–736. doi:10.1016/j.jclinepi.2009.08.028

Little, R. (1988). A test for missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202.

Madow, W., Nisselson, H., & Olkin, I. (Eds.). (1983). *Incomplete data in sample surveys: Vol. 1. Report and case studies*. New York, NY: Academic Press.

Madow, W., & Olkin, I. (Eds.). (1983). *Incomplete data in sample surveys: Vol. 3. Proceedings of the symposium*. New York, NY: Academic Press.

Madow, W., Olkin, I., & Rubin, D. (Eds.). (1983). *Incomplete data in sample surveys: Vol. 2. Theory and bibliographies*. New York, NY: Academic Press.

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

Murtonen, M., & Lethtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education*, *28*, 171–185. doi:10.1080/0307507032000058064

Peng, C., Harwell, M., Liou, S., & Ehman, L. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real data analysis* (pp. 31–78). Charlotte, NC: New Information Age.

Peugh, J., & Enders, C. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*, 525–556. doi:10.3102/00346543074004525

Raymond, M., & Roberts, D. (1987). A comparison of methods for treating data in selection research. *Educational and Psychological Measurement*, *47*, 13–26. doi:10.1177/0013164487471002

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537–560. doi:10.1111/j.1744-6570.1994.tb01736.x

Rubin, D. (1976). Inference and missing data. *Biometrika*, *63*, 581–592. doi:10.2307/2335739

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley. doi:10.1002/9780470316696

Salkind, N., & Rasmussen, K. (2007). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.

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

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604. doi:10.1037/0003-066X.54.8.594

Wilks, S. (1932). Moments and distributions of estimates of population parameters from fragmentary samples. *Annals of Mathematical Statistics*, *3*, 163–195. doi:10.1214/aoms/1177732885

Wood, A., White, I., & Thompson, S. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*, *1*, 368–376. doi:10.1191/1740774504cn032oa

Yesilova, A., Kaya, Y., & Almali, M. (2011). A comparison of hot deck imputation and substitution methods in the estimation of missing data. *Gazi University Journal of Science*, *24*(1), 69–75.

Young, W., Weckman, G., & Holland, W. (2011). A survey of methodologies for the treatment of missing values within datasets: Limitations and benefits. *Theoretical Issues in Ergonomics Science*, *12*, 15–43. doi:10.1080/14639220903470205

Zhou, X., Wang, X., & Dogherty, E. (2003). Missing-value estimation using linear and non-linear regression with Bayesian gene selection. *Bioinformatics*, *19*, 2302–2307. doi:10.1093/bioinformatics/btg323

Author

JEHANZEB R. CHEEMA is a clinical assistant professor at the College of Education, University of Illinois at Urbana-Champaign, 1310 S Sixth St, Champaign, IL 61820, USA; e-mail *jrcheema@uiuc.edu*. He holds a PhD in education with a specialization in research methodology from George Mason University and a PhD in economics from the University of Wisconsin-Milwaukee. His research areas of interest include missing data analysis, student academic achievement, and survey research.