

# Florida State University Libraries

---

Faculty Publications

Department of Educational Leadership and Policy Studies

---

2014

## Working with Missing Data in Higher Education Research: A Primer in Real-World Example

Bradley Cox, Kadian McIntosh, Robert Reason, and Patrick T. Terenzini



Working with missing data in higher education research:

A primer and real-world example

Bradley E. Cox  
Florida State University

Kadian McIntosh  
Pennsylvania State University

Robert D. Reason  
Iowa State University

Patrick T. Terenzini  
Pennsylvania State University

*Author Note.* A previous version of this paper was presented at the Annual Meeting of the Association for the Study of Higher Education in Indianapolis, IN. November, 2010. Special thanks are due to Stephen DesJardins for his valuable feedback.

## **Working with missing data in higher education research:**

### **A primer and real-world example**

In the study of higher education, nearly all quantitative analyses draw from incomplete datasets. Incomplete responses from research participants (as occurs when someone skips a survey item or refuses to answer a question) can have a dramatic influence on the results produced through statistical analysis. Although the manner of influence varies depending on how missing data are handled, at least *some* bias occurs whenever data are missing. No amount of statistical compensation can remove *all* of the biases inherent with incomplete data. Indeed, Allison (2002) concluded that the “only really good solution to the missing data problem is not to have any” (p. 2). But as Allison also acknowledged, perfectly complete data sets are highly improbable when conducting survey research on human subjects.

The prevalence of missing data in education research was illustrated most clearly by Peugh and Enders (2004) who examined leading education journals published in 1999 and 2003. They identified 389 studies that were published with missing data, *all but six* of which either ignored the issue entirely or employed outdated “traditional” methods (e.g., mean replacement, listwise deletion) of missing-data adjustments. A review of recent articles from the *Review of Higher Education* suggests that missing data still pose a problem for quantitative researchers in the field. Of the twenty articles published in 2012, one was a conceptual piece about network analysis, one appeared to have no missing data, one was an invited response piece, and eight more were qualitative (excluding one “supplemental” issue). Of the nine quantitative articles with incomplete data, only three provided explicit justifications for their approach to handling missing data. The remaining six quantitative articles made little or no mention of missing data,

generally leaving the reader to use context clues to infer that any case with any missing data was excluded from analysis (i.e., listwise deletion).

Higher education researchers thus regularly encounter – and often ignore – a problem for which there is no perfect solution. Nonetheless, researchers can deal with missing data in a manner that corrects as much as possible for known biases while also maximizing data usability and applicability to the broad aims of the research project. To help researchers do so, in this paper we 1) describe *why* missing data matter, 2) discuss the advantages and disadvantages of six common methods for handling missing data, and 3) present a real-world, empirical example of how one’s approach to missing data can have important effects on statistical conclusions, researcher interpretations, and resultant implications for policy or practice.

We are certainly not the first group of researchers to highlight the issue of missing data. Outside of higher education, the topic has received considerable attention in the fields of psychology and sociology. The topic has also been addressed in education broadly, most notably by Peugh and Enders (2004) in *Educational Researcher*. Within higher education, specifically, the topic has been addressed both in professional meetings (e.g., pre-conference workshops for the Association of Institutional Researchers) and in print (Croninger & Douglas, 2005). Despite these multiple and varied discussions of missing data, the field of higher education has been heretofore reluctant to confront the issue consistently or embrace the most modern approaches to handling missing data. Why, then, does the field still not embrace these modern techniques when, as one of the most user-friendly higher education publications on this topic (Croninger & Douglas, 2005) concluded, “newer strategies for coping with missing data yield not only accurate but more precise parameter estimates than traditional strategies do” (p. 48)?

That same manuscript (Croninger & Douglas, 2005) may provide some clues as to why the field has hesitated when confronted with missing data. First, the authors are relatively cautious when making suggestions for other researchers. Citing simplicity and efficiency, they “don’t think investigators should abandon traditional strategies,” but should simply “expand their missing data repertoire to include these newer techniques” (p. 48). The authors’ reluctance to abandon traditional approaches is also related to a second reason the field hasn’t transitioned to modern methods: sometimes traditional approaches yield the same results as do the newer methods. Indeed, analysis of their own made-up dataset showed that “all six missing data strategies [including both modern and traditional approaches] yield the same substantive conclusions as [was] made for the regression analysis based on complete data” (p. 44). The Croninger and Douglas (2005) study also highlights a third potential reason missing data continues to be overlooked in the higher education literature: nearly all of the studies advocating the use of more modern techniques base their conclusions on complex statistical manipulations and/or analysis of artificial datasets. Higher education researchers, however, work in an *applied* field and seek concrete solutions to common problems with *real world data*. As a field, we typically are not compelled by abstract, esoteric, or hypothetical problems.

Thus, our paper tries to overcome each of these potential reasons for hesitation by a) using real data collected through surveys of 5,000+ students at 33 institutions, b) showing that different missing-data approaches can lead to substantively meaningful differences when interpreting results, and c) providing direct suggestions for selecting an appropriate method for handling missing data.

## Why Missing Data Matter

Rubin's (1976; Little & Rubin, 1987, 2002) terms describing the mechanism by which data are missing, though somewhat confusing syntactically, are foundational to any discussion of analysis with missing data. Data are said to be "missing *completely* at random" (MCAR) when the likelihood of data being missing for variable  $Y$  is not related to the value of  $Y$  itself, nor to any other variables in the analytic model. Missing data are said to be "non-ignorable" (Little & Rubin, 1987, 2002) or "missing *not* at random" (MNAR) (Graham, 2009; Schafer & Graham, 2002) when the likelihood of a datapoint being missing is the result of some unobserved factor not accounted for by other variables in the analysis. Finally, missing data can be labeled as "missing at random" (MAR) if "the probability of missing data on a variable  $Y$  is related to some other measured variable (or variables) in the analysis model but not to the values of  $Y$  itself" (Enders, 2010, p. 6). Croninger & Douglas (2005), Enders (2010), and Graham (2009) offer more thorough, though still reader friendly, discussions of missing data types.

Throughout this paper, we assume that the missing data qualifies the definition of "missing at random" (MAR). Although many datasets do not technically meet MAR conditions, and no straightforward method is able to confirm the assumption, most quantitative studies in education to make the (often implicit) assumption that their data are missing at random. Dealing with missing data is dramatically simplified when data are MCAR (Graham, 2009), though the phenomenon rarely occurs with survey research. In contrast, analyses when missing data are MNAR are considerably more complex and require even advanced statisticians to take "extreme caution" (Allison, 2002, p. 77); Allison (2002) and Graham (2009) offer introductions to analyses with MNAR missing data. Regardless, missing data of any type can have adverse

effects on two key components of any statistical analysis (parameter estimates and standard errors) which can lead to errors of inference and interpretation.

### **Parameter Estimates**

When aware of the reason(s) why some people would not respond to questions, researchers could alter the study's data collection procedures to encourage complete responses. However, despite meticulous planning and conscientious implementation of strong designs, researchers typically end up with datasets for which they do not know exactly why some data are missing.

Although the researcher may not know why the data set has missing data, there *is* a reason. Without accounting for the underlying reasons for such missing data, the resulting parameter estimates (e.g., means, correlation coefficients) contain bias of an unknown direction and magnitude, making the bias difficult to detect and easy to overlook. These unknown biases are even more difficult to identify in multivariate analyses derived from a variance-covariance matrix (e.g., regression, hierarchical linear models).

### **Standard Errors**

Standard errors represent the precision and certainty with which an estimated statistic (e.g., mean, Pearson correlation, regression coefficient) reflects a true population parameter. An instance of missing data reflects the explicit absence of a known value for a particular variable in a particular case, thereby introducing uncertainty into any estimate of population parameters. Without accounting for uncertainty introduced by such "missingness," standard errors will be underestimated (biased downward, sometimes dramatically, depending on the amount of missing

data), thus increasing the likelihood of making a Type-I error where one incorrectly finds that an estimate is statistically significant.

#### **Four “Traditional” Options for Missing Data**

##### **Listwise Deletion (Full-Case Analysis)**

Full-case analysis (a.k.a., listwise deletion), using only those cases with complete data on the variables deemed by the researcher to be critical for analysis, is perhaps the most commonly employed approach to the missing data problem. This approach is appealing for two reasons. First, it is simple to implement and can be used for any type of statistical analysis. Second, in certain specific circumstances, full-case analyses yield parameter estimates as accurate as, and sometimes even more accurate than, more modern or complex approaches. Allison (2002) concluded that “whenever the probability of missing data on a particular independent variable depends on the value of that variable (and not the dependent variable), listwise deletion may do better than maximum likelihood or multiple imputation” (p. 7). Similarly, Graham (2009) argued that parameter estimates from listwise deletion will be only minimally biased, “especially for multiple regression models” (p. 554), when sufficient covariates are included in the models.

Concerns about complete-case analyses come in two forms. First, in many circumstances – whenever the criteria outlined by Allison (2002) and Graham (2009) are not met – parameter estimates may be biased (Allison, 2002; Little & Rubin, 1989). If cases are removed from the original sample, the remaining analytic sample is not fully representative of the original sample or population, creating conditions likely to yield biased parameter estimates (Little & Rubin, 1989). Second, and even when the Allison (2002) and Graham (2009) criteria *are* met, researchers using listwise deletion decrease the effective sample size, thereby decreasing the

statistical power of the analyses. The loss of power makes it more difficult to detect relatively small (but potentially important) effects or relationships between variables. In a field where strong models often explain little more than ten percent of the variance in student outcomes, researchers can ill-afford any unnecessary loss of statistical power.

### **Pairwise Deletion**

In an effort to retain cases and preserve statistical power, some researchers employ a process called pairwise deletion. With pairwise deletion, each particular analysis is based on the subset of cases having values for all variables in that particular analysis. For example, a respondent who answered question *A* but *not question B* would be included when calculating the mean for question *A*, but excluded from an identical calculation for question *B*. The problems associated with using different subsamples to calculate parameter estimates for a single population are magnified for statistics derived from a variance-covariance matrix (e.g., correlations, regression coefficients). When based on a pairwise approach to missing data, the covariance pairs populating such a matrix would be based on different subsets of cases and have inconsistent values for  $n$ . Thus, parameter estimates would be not only biased but also biased in multiple directions and magnitudes. The lack of a consistent  $n$  also complicates calculations of appropriate standard errors, requiring complex formulas not typically available in popular statistical software (Allison, 2002; Graham, Cumsille, & Elek-Fisk, 2003). Finally, because the variance-covariance matrix may have less information than would be expected based on the number of variables, a pairwise approach increases the likelihood (relative to other approaches for handling missing data) of obtaining a matrix that is not positive definite, a circumstance that halts most analyses (Allison, 2002; Graham, 2009; Kim & Curry, 1977).

## Mean, regression, or hot-deck imputation

Mean substitution, regression-based single imputation, and hot- and cold-deck imputation are grouped together because results using any of these procedures suffer from similar problems (Schafer & Graham, 2002). While these methods preserve data and are easy to use, their flaws make them less appealing than the more modern approaches to missing data. Mean substitution fills in the missing value with the sample mean (or, sometimes, the median or modal value) for each particular variable. This procedure does not alter calculations of variable means. However, because the values filled-in for a variable are all the same, the addition of mean-imputed values reduces estimates of population variance, thereby attenuating variance and covariance estimates (Roth, 1994) and any subsequent statistics derived from the variance-covariance matrix. Mean imputation for subgroups (e.g., replacing missing values for a *woman's* SAT score with the mean SAT score of all sampled *women*) offers some improvement over whole-sample mean replacement but remains problematic for similar reasons.

A slight improvement over mean imputation procedures, regression imputation (or conditional mean imputation, as it is sometimes called) creates a predictive model that uses available data as regressors to generate a predicted value for missing data points. Hot-deck and cold-deck procedures involve the replacement of missing data points with values from other “matched” cases in the same dataset (for hot-deck) or a comparable dataset (for cold-deck). For example, with hot-deck imputation, a 20-year old White male survey respondent (“Student A @ 20yrs”) who did not report his parents’ income might be given the value of some other 20-year old White male (“Student B @ 20yrs”) in the sample who *did* provide his parents’ income. Using cold-deck imputation, an analyst would simply insert the parental income reported by the same student two years earlier (“Student A @ 18yrs”). These methods are intuitively appealing

because the values assigned are generally realistic values that, unlike mean substitution, preserve the variables' distributional characteristics (Roth, 1994; Schafer & Graham, 2002). Moreover, the quality of regression, hot deck, and cold deck imputations improves as the number of cases and/or variables increases (thereby allowing greater complexity in the regression equation and/or matching procedure). However, each of these methods is likely to yield biased parameter estimates and underestimated standard errors (Schafer & Graham, 2002); r-squared will be overestimated in subsequent regressions that include the imputed variables (Graham, et al., 2003).

### **Dummy-Variable Adjustment**

A final “traditional” method for handling missing data builds upon mean imputation procedures by introducing one or more dummy-coded variables indicating the extent to which each case contains missing data on the variables included in the analyses. Based on the procedure suggested by Cohen and Cohen (1975), several variants of dummy-variable adjustments have been employed by educational researchers. In all variations, however, the underlying principle is the same: when including a dummy-coded indicator of missingness alongside the original and mean-imputed data, the calculated coefficient for the dummy variable attempts to control for whatever effect an independent variable's missingness might have on the dependent variable. However, as first demonstrated by Jones (1996), subsequently confirmed by Allison (2002), and stated plainly by Graham (2009), the dummy-variable approach “has been discredited and should not be used” (p. 555).

## Two “Modern” Options for Missing Data

Having recognized the various shortcomings of the traditional approaches to statistical corrections for missing data, methodologists have developed two “modern” methods that offer considerable improvements over the traditional approaches. Granted, these methods are not perfect. First, as Enders (2010) notes, data that are not multivariate normal in their distribution “can distort model fit statistics and standard errors, with or without missing data” (p. 124), and even the most advanced missing data procedures can yield biased parameter estimates when continuous variables are highly skewed (Lee and Carlin, 2010). Still, these “modern” approaches (especially multiple imputation) are quite robust to violations of such assumptions about the data (Allison, 2002; Johnson & Young, 2011). Second, maximum likelihood or multiple imputation procedures can complicate interpretation of subsequent analyses because imputed datasets can contain illogical or implausible values for certain cases (e.g., a typically dichotomous “sex” variable could be imputed as 8.7% female, or a Likert scale with integer anchors ranging from 1 to 5 could have an imputed value of 0.23) – a phenomenon that, while intuitively perplexing, is statistically legitimate; although software packages often include an option to round imputed data or force plausible values for each imputed variable, statisticians generally discourage the practice (Graham, 2009; Horton, Lipsitz, & Parzen, 2003; Schafer, 1997; see Bernaards, Belin, & Schafer, 2007, for a more thorough discussion). Finally, the use of an advanced imputation method does not immunize subsequent analyses from computational or interpretive complications caused by multicollinearity or improper analytic model specifications.

Nonetheless, data sets created by maximum likelihood and multiple imputation procedures retain all of their original cases and maintain the underlying relationships between variables. Both maximum likelihood (especially Expectation-Maximization approaches) and

multiple imputation estimation processes can be improved when the imputation models include auxiliary variables (Enders, 2010; Graham, 2009) that, while not necessarily to be included in the final substantive analyses, are available in the dataset and potentially related to the source of missing data. The most effective such auxiliary variables are those that have few missing values themselves and also have non-redundant correlations with other variables in the imputation models. In subsequent sections of this paper we highlight additional procedure-specific adjustments that allow researchers to increase further the accuracy and precision of their analyses with imputed data. Collectively, the benefits of these “modern” approaches to missing data analyses typically outweigh their costs, and most statisticians (e.g., Allison, 2002; Enders, 2010; Graham, 2009; Little & Rubin, 1987; Schafer & Graham, 2002) conclude that such approaches generally outperform their more “traditional” counterparts. As Allison (2002) summarizes, “some missing data methods are clearly better than others” and “maximum likelihood and multiple imputation [approaches to handling missing data] have statistical properties that are about as good as we can reasonably hope to achieve” (p. 2).

### **Maximum Likelihood**

Maximum likelihood approaches to missing data imputation use iterative procedures to estimate means, variances, and covariances reflecting the *most likely* population parameters from which the sample data would be drawn. Within the family of maximum likelihood procedures, the application of some variant of Little and Rubin’s (1987, 2002) Expectation-Maximization (EM) algorithms is perhaps the most commonly used in the educational literature. Other maximum likelihood approaches appear infrequently in higher education research. For example, a Full-Information Maximum-Likelihood (FIML) procedure, typically used for structural equation models, can offer some advantages, especially when interactions among variables are of

particular interest to the researcher. However, the relative complexity of FIML has limited its inclusion in popular statistical software and, even when included, generally limits the range of statistical analyses that can be applied to the data (Graham, 2009; Johnson & Young, 2011). See Hausmann et al. (2009) for an example of FIML used in higher education; see Enders (2001, 2010) for details about FIML procedures.

The EM procedures involve a two-step, iterative process. The *estimation* step uses estimated population parameter values to determine the likelihood that the observed data would have come from a population with the given parameters. The *maximization* step then adjusts the population parameter estimates to maximize the aforementioned likelihood. These two procedures continue until sequential estimation and maximization steps yield (nearly) equivalent values. In doing so, the EM algorithms yield estimated population parameters (means, variances, and covariances) that are free from some of the biases that accompany the “traditional” approaches to analysis with missing data (Graham, 2009; Little & Rubin, 1987; Schafer & Graham, 2002; von Hippel, 2004).

Several commercially available statistical packages<sup>1</sup> can perform such a procedure and also “fill in” the missing values in a dataset. In creating an EM-imputed dataset, for example, SPSS replaces the original missing data with data from the last iteration’s Expectation results, a practice yielding subsequent statistics with lower levels of variance than the true maximum-likelihood means, variances, and covariances (von Hippel, 2004). Moreover, the analyses thereafter are unable to distinguish the original data from the imputed data. Despite the increased

<sup>1</sup> Although the analyses conducted for this study were run using SPSS, similar procedures and results would be obtained using most major statistical packages (e.g., SAS, STATA, R). Several stand-alone programs are also available for specific procedures (e.g., NORM, CAT, and MIX programs by Schafer). Relative to SPSS, many of these alternate software packages offer more advanced options for analysts looking to, for example, bootstrap starting values or conduct graphical convergence diagnostics to help confirm the quality of the imputation process.

level of uncertainty introduced by the missing data, therefore, subsequent analyses using a single EM-imputed dataset fail to account for such uncertainty and yield standard errors that are artificially small, threatening the validity of subsequent hypothesis testing (Graham, 2009; von Hippel, 2004)<sup>2</sup>. In sum, although imputation using an EM algorithm will produce generally unbiased correlation and regression coefficients, researchers must acknowledge the increased likelihood of Type-I errors, and would be wise to adopt more conservative critical p-values (e.g., .01 instead of .05).

### **Multiple Imputation**

Multiple imputation (MI) procedures for handling missing data have been employed for analyses in higher education, although the practice is not common in studies published in the major higher education journals. During multiple imputation, an initial value for each originally missing data point is predicted using other data in the dataset. Then, a random error term, drawn from a Bayesian distribution of parameter estimates, is added to each predicted value. The process iterates with the new values (i.e., initial value plus random error term) serving as the initial values for the next iteration. After some defined number of iterations, the imputed values are written to a new dataset. This process repeats to create a user-specified number of imputed datasets. Subsequent substantive analyses are then run separately for each of the imputed datasets. The resulting parameter estimates and standard errors are pooled, typically using equations outlined by Rubin (1987).

<sup>2</sup> Moreover, as of version 17, SPSS has effectively abandoned its support of EM-based dataset creation and instead encouraged use of its multiple imputation procedures (see <http://www-01.ibm.com/support/docview.wss?uid=swg21479652>).

Whereas other missing data procedures artificially reduce variance by creating imputed values that fall precisely along a prediction line, a practice that fails to replicate the variability existing within the larger population, multiple imputation repeatedly adds random error terms and pools estimates across datasets to reflect the uncertainty that would exist if the analyst somehow knew the true values for the missing data points. Pooled estimates can be computed for an ever-growing number of analytic procedures (e.g., correlations, multiple regression, and multilevel models). Multiple imputation thus provides analytic flexibility; preserves the underlying characteristics of the data and produces approximately unbiased estimates of means, variances, covariances, correlations, and regression coefficients; yields appropriate standard errors; and retains much of the statistical power lost with other methods (Allison, 2002; Graham, 2009).

Multiple imputation, however, has practical limitations. First, because the imputation process includes the addition of *random* error terms, estimated population parameters (e.g., means or regression coefficients) from analyses using imputed data will differ slightly each time an analysis occurs. These differences, occurring even if researchers were to repeat the same analysis on the same data, are (perhaps counter-intuitively) statistically legitimate variations reflective of the uncertainty inherent in any analysis that includes missing data. Second, the introduction of random error modifications after each iteration adds considerably to computational complexity. Accordingly, multiple imputation may not be practical for researchers using large datasets or employing complex analytic models (sometimes requiring hours or days to complete).

## Conclusions from the Literature

Three fairly straightforward conclusions can be drawn from the statistical literature regarding procedures for handling missing data. First, except in the most specific of circumstances, “traditional” methods (i.e., complete-case analysis, pairwise deletion, regression imputation, hot- and cold-deck imputation, and dummy-variable adjustments) should be avoided. Second, multiple imputation has emerged as the preferred option among statisticians and sociologists (who have been employing advanced methods for more than a decade). With most major software packages (e.g., SPSS, STATA, SAS, Mplus) now including options for multiple imputation, a strong argument can be made that multiple imputation should be the new default option for quantitative research in higher education. Third, if context-specific considerations make multiple imputation unfeasible, maximum likelihood approaches provide a good, second choice. If considering a maximum likelihood approach, however, researchers should remember that Full Information Maximum Likelihood (FIML) is practical, typically, only for structural equation models (see Enders, 2001 for details) and that a single dataset EM-algorithm approach might warrant the use of a more conservative *p-value* (i.e., .01 instead of .05) when interpreting results.

## A Real-World Example of Missing Data Approaches and Outcomes

Complicated equations and technical language make it difficult to recognize the manner in which decisions regarding missing data can affect study results. This section offers a straightforward example of how different approaches to the missing data problem play-out in a real-world study. In this example, we use data from the Parsing the First Year of College project to explore whether and how students’ critical thinking is related to their use of activities reflecting Bloom’s (1956) taxonomy of learning.

## **Sample**

The Parsing the First Year of College project drew student participants from a pool of all baccalaureate degree-seeking first-year students enrolling in 33 participating institutions during the Fall 2006 term. To be included in the final sample, students must have completed both the ACT Test (college entrance exam) and at least one of four assessments in the Spring 2007 term, near the end of their first college year: the National Survey of Student Engagement, the Critical Thinking and/or Writing Skills module(s) of the Collegiate Assessment of Academic Progress (CAAP), and a supplemental survey designed specifically for this study. Data from these assessments were augmented by information provided by participating institutions and drawn from students' ACT college entrance exam (both the student profile information and the actual test scores). A total of 5,905 students provided information on one or more of the study's instruments.

## **Data and Missingness**

For this example, we use students' scores on the CAAP – Critical Thinking (CAAP-CT) module as our outcome variable. Of the 5,905 students in our sample, 2,245 had data for all of the 14 variables used in this example (including demographic information, measures of pre-college academic success, and college experiences; see Tables 1 and 2 for details). Most student cases either provided complete responses or were missing data on only one variable. The distribution of missing values is presented in Table 1 and Figure 1. The most frequently missing data element was the dependent variable (CAAP-CT). Although we have heard researchers in higher education avoid the imputation of outcome variables, statisticians (Landerman, Land, & Pieper, 1997; Little & Rubin, 2002) have demonstrated that imputation of dependent variables

“is *essential* for getting unbiased estimates of the regression coefficients” (Allison, 2002, p. 52; italics in the original).

### **Analyses Conducted to Compare Approaches for Handling Missing Data**

To demonstrate the effects of different approaches for handling missing data, we ran two sets of analyses typical of methods used in studies of student success in higher education. In the first analysis, we compiled descriptive statistics for each of the 14 variables to be used in subsequent regression analysis. Means and standard deviations are presented in Table 2. For the second analysis, we ran an OLS regression in which we regressed students' score on the CAAP - Critical Thinking module on the 13 variables reflecting students' demographic characteristics, pre-college academic preparations, and college experiences. Many other forms of statistical analyses (e.g., logistic regression, structural equation modeling, or hierarchical linear modeling) add considerable complexity and nuance to the discussion of missing data adjustments and, therefore, fall beyond the scope of this paper. Readers seeking an introductory discussion of these circumstances should review Allison (2002) and/or Graham (2009).

### **Missing Data Approaches used for Comparison**

We repeated the descriptive and regression analyses for each of four popular approaches to dealing with missing data: 1) using only complete cases (i.e., listwise deletion); 2) a combination of pairwise deletion, mean imputation, and dummy-variable adjustment; 3) maximum likelihood imputation using an Expectation-Maximization algorithm to create a single imputed dataset; and 4) multiple imputation creation of 10 imputed datasets. All data manipulations and calculations were made using SPSS software (version 18.03).

**Complete-case analyses (listwise deletion).** Using the most frequently chosen and simplest option for handling missing data excludes from analysis any student who is missing any of the relevant data. In our example, full-case analysis could be conducted on 2,245 students, meaning data from 3,660 students would be ignored. By including only students with complete data, we lost more than 60% of our available cases, many of which were missing only one of the fourteen relevant data points. Results from the complete-case analyses are presented in the first column of Tables 2 and 3.

**Pairwise deletion, mean imputation, and dummy-variable adjustment.** To improve clarity in reporting results – and to reflect the way in which the three procedures are often used together in practice – we combine pairwise deletion, mean imputation, and dummy-variable adjustment procedures. When calculating descriptive statistics, we employ pairwise deletion, so means and standard deviations reported in the second column of Table 2 are based on all cases that have a value for a particular variable. Thus, because the CAAP-CT score was missing for nearly half of the students in the dataset, its descriptive statistics are based on only 2,895 students' scores. In contrast, every student had an ACT Test score, so descriptive statistics for that variable are based on roughly twice as many students ( $n=5,905$ ).

Before running regression analyses using a dummy-variable approach to adjust for missing data, the researcher must choose which variant of the dummy-variable procedure to implement. The most common variations are to: a) create a single variable indicating the extent to which each case includes missing data on the variables to be included in the analysis (i.e., a summary of each case's overall level of "missingness"), or b) create a dummy-coded indicator of missingness for *each* of the variables to be included in the analysis (e.g., a variable called *missing\_q6* would be given a value of 0 for all students who answered the original question 6,

and would be given a value of 1 for those students who did *not* answer the original question 6).

Table 3 presents results from procedure “b,” although results using procedure “a” (not presented here) are nearly identical.

Once the dummy variables are created, they – along with the original variables from which the dummies are derived – are entered into a regression. When computing the regression coefficients, the researcher must specify what the software should do when it encounters missing data (that is still present amid the original variables). Although one could select list-wise or pairwise deletion, either option could decrease statistical power for the regression. Instead, likely believing that the dummy-coded indicator(s) of missingness have already accounted for any bias associated with missing data, researchers often choose the mean-imputation option (thereby retaining all of the original data and its accompanying statistical power). Accordingly, the dummy-variable-adjusted regression for the current analysis also uses the mean-imputation option.

**Maximum-likelihood single dataset imputation using EM algorithm.** Our first “modern” approach to missing data uses an expectation-maximization algorithm. Preparation of the EM database requires significant attention to the original/source dataset. All categorical variables were dummy coded; missing data indicators were set consistently for all variables; a dummy-coded indicator of students’ institution was created; and several common and theoretically relevant interaction terms were created (e.g., race by gender, race by ACT Test score). These variables, and all of the student-level variables in the original dataset (323 variables in total),<sup>3</sup> are included in the model from which the imputed dataset will be derived.

<sup>3</sup> With each additional variable added, the computational power/time required to compute an underlying variance-covariance matrix (a key process for maximum-likelihood and multiple imputation procedures) grows

Among the included variables are some that may be dependent variables in future analyses<sup>4</sup>, the omission of which would lead regression coefficients from subsequent analyses to be substantially biased toward zero (Allison, 2002; Schafer, 1997). The imputation model also includes auxiliary variables that, while not likely to be used for subsequent analyses, nonetheless may reduce bias in the resulting imputed dataset (Allison, 2002; Collins & Kam, 2001).

In our example, the means and standard deviations reported in the third column of Table 2 are based on the complete sample (5,905 students) and comply with Graham's (2009) suggestion that the EM procedure be used only for analyses that do not require interpretation of standard errors (e.g., means, exploratory factor analysis). Table 3, however, does not follow Graham's (2009) suggestion that an EM-imputed dataset "should not be used for hypothesis testing [because] standard errors based on this dataset, say from a multiple regression analysis, will be too small, sometimes to a substantial extent" (p. 556). Although Graham (2009) suggests that multiple regression analyses with missing data should be based on full-information maximum-likelihood (FIML) or multiple imputation (MI) procedures, these approaches are not practical within the context of the larger project from which the current analyses are drawn. Therefore, acknowledging the need to find a compromise between what was statistically ideal and what was realistically practical, we use the expectation-maximization procedure to create a single imputed dataset that serves as the source for subsequent descriptive and regression

exponentially. Using a desktop computer with a quad-core processor effectively dedicated to our procedure, the EM imputation took nearly three days to complete.

<sup>4</sup> Ideally, analysts would revisit the issue of missing data every time they run a different analysis or specify a different outcome variable. But, for reasons explained later in this paper, such an approach was not practical within the context in which the larger project was situated. Revisiting the missing data problem for each analysis, however, may be more practical in other contexts and could drastically reduce the computation time required for each discrete implementation of a maximum likelihood or multiple imputation procedure for missing data.

analyses. We make inferences of statistical significance, however, based on the critical value  $p \leq 0.01$  instead of the conventional  $p \leq 0.05$ .

**Multiple Imputation.** In preparing to run the multiple imputation procedure, we again ensured that all variables in the dataset were properly specified as categorical, ordinal, or scale. The SPSS multiple imputation command uses the variable-type specification to determine whether each imputed variable will be drawn from a linear or logistic regression distribution modeled on the other variables included in the dataset. Although the assignment of an appropriate distribution for each variable in a multiple imputation model (e.g., logistic for dichotomous/categorical variables) has some statistical benefits and intuitive appeal, we could have skipped this step without much sacrifice. Instead, because multiple imputations based on the normal distribution have been shown to be remarkably robust to violations of normality (Graham & Schafer, 1999; Schafer, 1997) the minor improvement to the accuracy of the imputed datasets may not be worth the additional computation time resulting from the complexity introduced by the use of variable-specific distribution specifications.

After data preparation, we instructed the software to create a total of 10 different datasets by extracting one new dataset after every 200 iterations of the imputation model. By increasing the number of iterations between datasets (from the typical default of 5), we adopted settings more in line with Allison's (2002) multiple imputation examples, though his results suggest such a change would have only marginal practical value. Moreover, although five datasets may be adequate for most analyses (Allison, 2002; Schafer & Olsen, 1998), Graham and colleagues (Graham, Olchowski, & Gilreath, 2007) recently called for a dramatic increase in the number of imputed datasets, suggesting that the number of imputed datasets should increase as the amount of missing data increases. Because some of the variables had a large amount of missing data, we

doubled the number of imputed datasets, thereby presumably “cut[ting] the excess standard error in half” (Allison, 2002, p. 50).

Finally, when specifying our multiple imputation model we included only the 14 variables that were to be used in the subsequent analyses. Knowing exactly what subsequent analysis we would be running, we chose not to do any variable transformations or include any potential interaction effects in our imputation models. Because the eventual analyses modeled on the imputed datasets would be relatively straightforward, so, too, could be the model that created the imputed datasets. However, the imputation model should be at least as complex as any subsequent analytic models (Allison, 2002; Graham; 2009; Rubin, 1987; Schafer, 1997; see also some new approaches mentioned on page 53 of Allison, 2002), meaning researchers should “anticipate any interaction terms and include the relevant product terms in the imputation model” (Graham, 2009, p. 562). For this reason, Rubin (1996) makes a compelling argument that the onus falls on the original data collectors to develop an appropriate imputation model before releasing the data to outside users.

### **Comparing Results Across Missing Data Methods**

In this section, we present results of two sets of analyses for each of four methods for handling missing data. These analyses have been chosen to represent the methods typically employed in studies of college student experiences and outcomes.

### **Descriptive Statistics**

When trying to argue that the analysis of a particular sub-sample would yield results approximating those from an analysis of an entire sample, researchers often compare the basic descriptive statistics of the whole vs. sub-sample. The descriptive statistics in this study (see

Table 2) allow a similar comparison. A cursory comparison from left to right across Table 2 reveals few clear differences in either means or standard deviations. On the whole, differences due to changes in missing data technique appear small, perhaps even trivial. Of the 14 variables, only two (the number of Advanced Placement courses taken in high school and whether the student was in the top quartile of his/her high school class) appear to have potentially meaningful differences in means. Comforted by this apparent similarity across the various methods for handling missing data, many researchers would confidently employ listwise deletion.

Likely to be overlooked, however, are differences in the standard errors, for example, of students' composite ACT score. By removing those cases that have *any* missing data, the full-case approach drastically cuts the sample size (thus increasing the standard error relative to any of the other methods) and fails to make complete use of *any* data from the dropped cases – even the student who answered all but one question. Each of the imputed datasets makes more complete use of the available data, yielding similar estimates for the mean ACT score, but with substantially more precision (i.e., a 30% smaller standard error). Were substantive analysis to end with descriptive statistics, such sporadic differences in standard errors are subtle, perhaps even negligible. But if analyses continue via multivariate modeling procedures (e.g., regression, structural equation modeling), the consequences of missing-data decisions more clearly become both statistically and practically important.

## **Regression**

As is typical in studies of college student experiences, we next regressed student outcomes (in this case, CAAP-Critical Thinking score at the end of the first year) on gender/sex, race/ethnicity, three measures of pre-college academic preparation, and eight variables reflecting

various college-specific experiences. The results, presented in Table 3, highlight the subtle manner in which missing data can affect statistical conclusions and researcher interpretations.

In particular, researchers using the analysis for hypothesis testing appear likely to draw different conclusions simply because of the choice of how to handle missing data. Those using mean imputation with dummy-coded indicators of missingness would, for example, conclude that White/Caucasian students receive higher CAAP-Critical Thinking scores than similar students of another race/ethnicity ( $p\text{-value} = .037$ ). Yet those using listwise deletion, EM, or multiple imputation approaches would not conclude that race/ethnicity has any direct relationship with students' critical thinking test scores ( $p\text{-values}$  of .326, .100, and .264, respectively). The opposite phenomenon occurs with the APPLYING variable: the coefficient is statistically significant for all of the missing data methods *except* mean imputation with dummy-coded indicators of missingness.

If using the conventional  $p \leq 0.05$  standard for statistical significance, researchers using the EM algorithm to impute a single dataset would conclude that students living on campus ( $p\text{-value} = 0.011$ ) or preparing two or more drafts of a paper ( $p\text{-value} = 0.023$ ) have lower critical thinking scores than their otherwise-similar peers. Such a conclusion, however, is not supported by results from any of the other three missing-data methods, nor does it meet the  $p \leq 0.01$  critical value we suggest to account for the artificial reduction in standard errors that likely accompanies our use of a single EM-imputed dataset. In contrast, results from the analysis using the EM-imputed data indicate that students in the top quartile of their high school class receive higher CAAP-CT scores than do their lower-rank peers – even when considered using the more conservative critical value of  $p \leq 0.01$ . But such an inference would not be made by researchers using any of the other three methods for handling missing data.

Careful readers of Table 3 might note a strong similarity between the results using listwise deletion and multiple imputation. In fact, hypothesis tests using either approach would yield the same conclusion regarding *every one* of the 13 independent variables. Although these results are inconvenient for our overarching argument, running counter to our earlier criticism of “traditional” procedures for handling missing data, they do serve to demonstrate an inevitable paradox that arises whenever data are missing: although statisticians working with theoretical or artificial datasets may be able to determine which missing data approach is “best,” datasets derived from real-world data collection offer no way to determine which of the approaches will yield the “right” results or most-closely reflect the “truth.”

### **Choosing a Method to Address Missing Data**

Although our analyses demonstrate that the choice of a procedure for handling missing data *matters*, our results do not clearly identify any mechanism as the *best*. Rather, this study suggests that the choice of an appropriate method is context specific. The Parsing the First Year of College project, the parent project from which original data for these analyses come, serves as an example of the manner in which project-specific context might drive researchers’ decisions regarding the handling of missing data. In deciding a course of action for the Parsing the First Year of College project, we acted upon the following five assumptions about the project:

1. The project database includes several measures that might be used as outcome variables in analyses.
2. Several of these same measures could also be used as independent variables or intermediate outcome variables (Astin, 1993) in analyses.

3. The project database would likely be used by researchers who, while working in loose collaboration, would be operating autonomously and from varied locations – a situation that increases the difficulty of maintaining dataset consistency over time.
4. Different users would have widely varying levels of comfort and experience dealing with missing data, thereby increasing the potential for confusion or errors when implementing missing data procedures or interpreting results.
5. Analysts would employ a variety of statistical procedures, computer hardware, and software packages when working with the project data.

Key to our eventual decision between maximum likelihood and multiple imputation (with the “traditional” approaches having already been eliminated following our review of the literature) were the issues of *stability* and *usability*. For the Parsing the First Year of College project, we decided to use an EM algorithm approach for most of our data analyses because it: a) would create a single, common dataset for all researchers (as opposed to the five or ten with multiple imputation); b) makes use of all 323 student variables (whereas, with multiple imputation, computational complexity placed severe practical constraints on the size of the imputation model); c) enables all types of subsequent statistical analyses (in contrast to FIML); d) does not require special software or extensive statistical expertise (as would be required for FIML or multiple imputation approaches); and e) allows us to offer straightforward suggestions about interpretation of statistical results. Of course, others might well have chosen a different approach, and we make no claim as to have chosen the “best” approach. Rather, we acknowledge the limitation of our approach and claim only to have made an informed and reasonable choice. In fact, in a subsequent project led by two of this paper’s authors (Cox & Reason’s Linking Institutional Policies to Student Success project; or *LIPSS* for short), for which data were

collected in 2011-12, we are following our own advice and have adopted multiple-imputation as the default approach for handling missing data.

### **Conclusion**

The problem of missing data is as persistent as it is complex. And while traditional methods (e.g., listwise deletion, pairwise deletion, mean imputation, and dummy-variable adjustments) have provided relatively simple solutions, they likely have also contributed to biased statistical estimates and misleading or false findings of statistical significance. More powerful methods that capitalize on recent statistical and computational advances now provide researchers with practical alternatives that address many of the shortcomings associated with the traditional approaches to analysis with missing data. As Allison (2002) concluded, “some missing data methods are clearly better than others” and “maximum likelihood and multiple imputation [approaches to handling missing data] have statistical properties that are about as good as we can reasonably hope to achieve” (p. 2).

Thus, we encourage higher education researchers to consider adopting multiple imputation as a first-choice method for handling missing data. Maximum likelihood approaches would be a reasonable second-choice should multiple-imputation be unavailable or impractical for a specific project. More traditional methods should generally be avoided unless researchers can make a compelling argument that their dataset satisfies the conditions outlined by Allison (2002) or Graham (2009). However, real-world datasets rarely meet these conditions and researchers have few tools with which to evaluate the adequacy of listwise deletion for a particular analysis.

Regardless of a researcher’s ultimate decision regarding missing data, we encourage scholars to actively address the issue in a manner that corrects as much as possible for known

biases while also maximizing data usability and applicability to the broad aims of the overarching research project. Though the resulting choices of procedures may vary, critical consideration and forthright discussion of missing data are likely to help both authors and readers make appropriate and defensible inferences.

## References

- Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage Publications.
- Astin, A. W. (1993). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Phoenix, Ariz.: Oryx Press.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: the classification of educational goals* (1st ed.). New York: Longmans, Green.
- Bernaards, C. A., Belin, T. R., & Schafer, J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine*, *26*, 1368-1382.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*: Lawrence Erlbaum Associates.
- Collins, L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330-351.
- Croninger, R. G., & Douglas, K. M. (2005). Missing data and institutional research. *New directions for institutional research*, *2005*(127), 33-49.
- Enders, C. K. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY.: Guilford Press.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*(1), 549-576. doi: doi:10.1146/annurev.psych.58.110405.085530
- Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). *Methods for handling missing data*: John Wiley & Sons, Inc.

- Graham, J. W., Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206-213. doi: 10.1007/s11121-007-0070-9
- Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1-29). Thousand Oaks, CA: Sage.
- Hausmann, L. R. M., Ye, F., Schofield, J. W., & Woods, R. L. (2009). Sense of belonging and persistence in white and african american first-year students. *Research in Higher Education*, 50(7), 649-669. doi:10.1007/s11162-009-9137-8
- Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *The American Statistician*, 57(4), 229-232.
- Johnson, D. R., & Young, R. (2011). Toward "best practices" in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and Family*, 73, 926-945. DOI:10.1111/j.1741-3737.2011.00861.x
- Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91(433), 222-230.
- Kim, J.-O., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Research*, 6(2), 215-240. doi: 10.1177/004912417700600206
- Landerman, L. R., Land, K. C., & Pieper, C. F. (1997). An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research*, 26(1), 3-33. doi: 10.1177/0049124197026001001
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Little, R. J. A., & Rubin, D. B. (1989). The analysis of social science data with missing values.

*Sociological Methods & Research*, 18(2-3), 292-326. doi:

10.1177/0049124189018002004

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, N.J.: Wiley.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: a review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525-556.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47(3), 537-560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592. doi:  
10.1093/biomet/63.3.581

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

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*, 33(4), 545-571.

von Hippel, P. T. (2004). Biases in SPSS 12 .0 missing value analysis. *The American Statistician*, 58(2), 160-164.

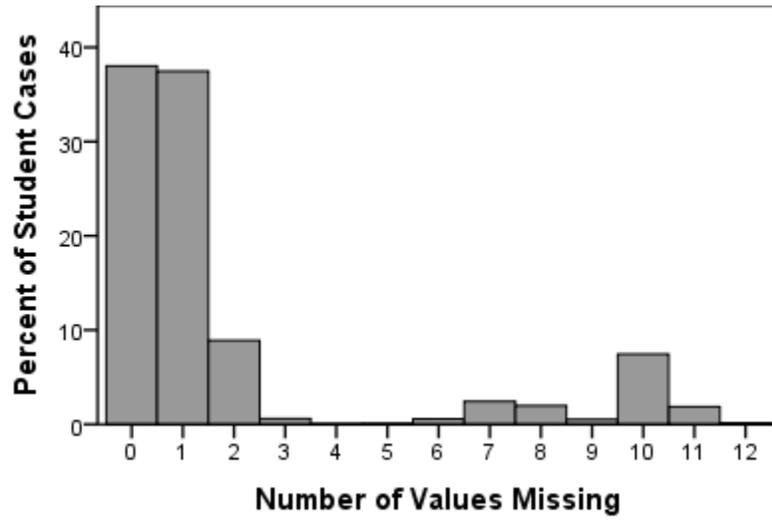


Figure 1. Distribution of Missing Data in the Real-World Example

Table 1: Distribution of Missing Data

Extent of Missing Data in the Real-World Example	
	<b>Cases with Missing Data</b>
<b>Female student</b>	<b>15</b> 0.3%
<b>White/Caucasian race</b>	<b>657</b> 11.1%
<b>ACT composite score</b>	<b>0</b> 0.0%
<b>Number of Advanced Placement courses taken</b>	<b>0</b> 0.0%
<b>Top 25% of high school graduating class</b>	<b>998</b> 16.9%
<b>Living on campus</b>	<b>1061</b> 18.0%
<b>Prepared 2+ drafts of a paper</b>	<b>667</b> 11.3%
<b>Integrated ideas or information across classes</b>	<b>669</b> 11.3%
<b>Coursework emphasized: MEMORIZING</b>	<b>887</b> 15.0%
<b>Coursework emphasized: ANALYZING</b>	<b>887</b> 15.0%
<b>Coursework emphasized: SYNTHESIZING</b>	<b>887</b> 15.0%
<b>Coursework emphasized: MAKING JUDGMENTS</b>	<b>887</b> 15.0%
<b>Coursework emphasized: APPLYING</b>	<b>887</b> 15.0%
<b>Critical Thinking Score (CAAP-CT)</b>	<b>3010</b> 51.0%

*Note.* Bold numbers indicate number of student cases missing a value for a particular variable; numbers in plain text reflect the percent of the original sample (n = 5,905) missing data on that variable.

Table 2: Comparison of Descriptive Statistics using Four Missing-Data Methods

<i>Cases</i>	Pairwise Deletion & Mean/Dummy			
	Listwise Deletion	Imputation	EM Imputation	Multiple Imputation
	2,245	2,895-5,905	5,905	5,905
<b>Female student</b>	<b>0.655</b>	<b>0.630</b>	<b>0.629</b>	<b>0.629</b>
	0.476	0.483	0.483	-
<b>White/Caucasian race</b>	<b>0.812</b>	<b>0.815</b>	<b>0.814</b>	<b>0.813</b>
	0.390	0.389	0.369	-
<b>ACT composite score</b>	<b>23.888</b>	<b>23.941</b>	<b>23.941</b>	<b>23.941</b>
	4.476	4.394	4.394	-
<b>Number of Advanced Placement courses taken</b>	<b>2.154</b>	<b>1.794</b>	<b>1.794</b>	<b>1.794</b>
	1.721	1.761	1.761	-
<b>Top 25% of high school graduating class</b>	<b>0.629</b>	<b>0.605</b>	<b>0.571</b>	<b>0.582</b>
	0.483	0.489	0.466	-
<b>Living on campus</b>	<b>0.777</b>	<b>0.769</b>	<b>0.778</b>	<b>0.770</b>
	0.416	0.422	0.404	-
<b>Prepared 2+ drafts of a paper</b>	<b>2.698</b>	<b>2.650</b>	<b>2.663</b>	<b>2.658</b>
	0.971	0.965	0.914	-
<b>Integrated ideas or information across classes</b>	<b>3.138</b>	<b>3.105</b>	<b>3.101</b>	<b>3.091</b>
	0.745	0.748	0.707	-
<b>Coursework emphasized: MEMORIZING</b>	<b>2.995</b>	<b>2.970</b>	<b>2.956</b>	<b>2.956</b>
	0.839	0.845	0.786	-
<b>Coursework emphasized: ANALYZING</b>	<b>3.137</b>	<b>3.127</b>	<b>3.110</b>	<b>3.084</b>
	0.752	0.755	0.705	-
<b>Coursework emphasized: SYNTHESIZING</b>	<b>2.940</b>	<b>2.916</b>	<b>2.901</b>	<b>2.864</b>
	0.796	0.809	0.755	-
<b>Coursework emphasized: MAKING JUDGMENTS</b>	<b>2.888</b>	<b>2.877</b>	<b>2.864</b>	<b>2.843</b>
	0.845	0.841	0.783	-
<b>Coursework emphasized: APPLYING</b>	<b>3.094</b>	<b>3.069</b>	<b>3.050</b>	<b>3.031</b>
	0.813	0.817	0.762	-
<b>Critical Thinking Score (CAAP-CT)</b>	<b>62.360</b>	<b>62.304</b>	<b>62.476</b>	<b>62.506</b>
	5.715	5.722	5.015	-

*Note.* Numbers in bold print are means; numbers in plain type are standard deviations. Standard deviations are not easily pooled across datasets and are thus not reported for the multiple imputation procedure.

Table 3: Comparison of Regression Statistics using Four Missing-Data Methods

<i>Cases</i>	Pairwise Deletion & Mean/Dummy			
	Listwise Deletion	Imputation	EM Imputation	Multiple Imputation
<i>Adjusted R-Squared</i>	2,245	2,895-5,905	5,905	5,905
<i>Adjusted R-Squared</i>	0.578	0.304	0.673	0.544-0.559
<b>Female student</b>	<b>0.150</b>	<b>0.040</b>	<b>0.007</b>	<b>0.018</b>
	0.367	0.664	0.933	0.908
<b>White/Caucasian race</b>	<b>0.207</b>	<b>0.258</b>	<b>0.174</b>	<b>0.191</b>
	0.326	0.037	0.100	0.264
<b>ACT composite score</b>	<b>0.909</b>	<b>0.480</b>	<b>0.893</b>	<b>0.895</b>
	0.000	0.000	0.000	0.000
<b>Number of Advanced Placement courses taken</b>	<b>0.190</b>	<b>0.095</b>	<b>0.115</b>	<b>0.159</b>
	0.000	0.000	0.000	0.003
<b>Top 25% of high school graduating class</b>	<b>0.234</b>	<b>0.053</b>	<b>0.258</b>	<b>0.208</b>
	0.213	0.634	0.007	0.238
<b>Living on campus</b>	<b>0.003</b>	<b>-0.197</b>	<b>-0.243</b>	<b>-0.202</b>
	0.987	0.095	0.011	0.294
<b>Prepared 2+ drafts of a paper</b>	<b>-0.060</b>	<b>-0.058</b>	<b>-0.107</b>	<b>-0.085</b>
	0.521	0.290	0.023	0.235
<b>Integrated ideas or information across classes</b>	<b>-0.003</b>	<b>0.066</b>	<b>0.015</b>	<b>0.002</b>
	0.982	0.346	0.810	0.983
<b>Coursework emphasized: MEMORIZING</b>	<b>-0.035</b>	<b>-0.016</b>	<b>0.010</b>	<b>0.027</b>
	0.721	0.787	0.842	0.802
<b>Coursework emphasized: ANALYZING</b>	<b>0.403</b>	<b>0.278</b>	<b>0.344</b>	<b>0.350</b>
	0.003	0.001	0.000	0.016
<b>Coursework emphasized: SYNTHESIZING</b>	<b>0.137</b>	<b>-0.034</b>	<b>0.067</b>	<b>0.087</b>
	0.287	0.665	0.324	0.538
<b>Coursework emphasized: MAKING JUDGMENTS</b>	<b>0.153</b>	<b>0.003</b>	<b>0.132</b>	<b>0.151</b>
	0.193	0.972	0.034	0.132
<b>Coursework emphasized: APPLYING</b>	<b>-0.363</b>	<b>-0.090</b>	<b>-0.231</b>	<b>-0.257</b>
	0.003	0.215	0.000	0.008

*Note.* Numbers in bold print are regression coefficients; numbers in plain type are p-values.