Imputing Missing Values in the Common Core of Data for Use in Computing the Averaged Freshman Graduation Rate[1]

Jack Buckley,* Marilyn Seastrom,* Chris Chapman,*
Vincent Iannacchione, Sandra Eyster,*** Dan McGrath,*** and Rob Stillwell*****

*National Center for Education Statistics,
Institute for Education Sciences,
U. S. Department of Education

**Education Statistics Services Institute-
American Institutes for Research

***Education Statistics Services Institute-
RTI International

jack.buckley@ed.gov

Introduction

Policy makers, educators, and the general public have long been interested in measuring the ability of schools and districts to reduce secondary school drop-out rates and to graduate their students on-time (i.e. in four years from the beginning of ninth grade). Recently, sparked by the decision of Congress to mandate the reporting of an on-time graduation rate by the states as part of the requirements of the No Child Left Behind (NCLB) Act, concerns about the methodology of computing this rate has led to a sometimes heated debate among stakeholders and analysts of education policy (Swanson 2004; Warren 2005; National Governors Association 2005; American Federation of Teachers 2006; Greene and Winters 2006; Mishel and Roy 2006; Seastrom, Hoffman et al. 2006).

Although virtually all participants in the debate have voiced support for its eventual resolution by the adoption of state longitudinal data systems (National Governors Association 2005), which will allow the computation of a true cohort graduation rate by observing the behavior of all individual students throughout their academic careers, most of the proposed approaches currently use cross-sectional enrollment and diploma data as reported by the National Center for Education Statistics (NCES) as part of the Common Core of Data (CCD). Among these methods is the Averaged Freshman Graduation Rate (AFGR), the measure adopted by the U.S. Department of Education in 2005, which will be discussed in the next section.

Calculation of the AFGR requires four data elements, each collected in a different academic year (details of the AFGR will be discussed in the next section). This makes the measure particularly susceptible to missing data concerns. Although the CCD is an important source of information about the K-12 public education system and is of generally high quality and completeness, it is a complex universe survey, and NCES relies on state cooperation in timely and accurate reporting of the many data elements. As a result, many school

---

[1] This paper is part of an ongoing research and development effort and it does not reflect the positions or views of the National Center for Education Statistics, the Institute of Education Sciences, or the U. S. Department of Education. Please contact the lead author for the most recent version prior to citation.

districts or local educational authorities (LEA's) are missing at least one data element required for the calculation of the AFGR, particularly by gender and race/ethnicity.[2] This paper explores the causes and consequences of these missing data on the estimation of the AFGR, and proposes and explores several imputation or adjustment models for correcting this issue.

The Averaged Freshman Graduation Rate

The AFGR provides an estimate of the percentage of high school students who graduate on-time (i.e. in four years from the start of ninth grade) by dividing the number of graduates with regular diplomas by the size of the incoming freshman class four years earlier, expressed as a percent. The rate uses aggregate student enrollment data to estimate the size of an incoming freshman class and aggregate counts of the number of diplomas awarded to that cohort four years later. The size of the incoming freshman class is estimated by a simple mean—the sum of the enrollment in eighth grade in one year, ninth grade for the next year, and tenth grade for the year after, divided by three. This averaging is intended to account for prior year retentions in the ninth grade. Although not as accurate as an on-time graduation rate computed from a cohort of students using longitudinal student record data, the AFGR can be computed with currently available data and compares favorably to other methods given validity tests against a true longitudinal rate (Seastrom, Chapman et al. 2006).

More formally, for a given LEA (or state), the AFGR is computed as:

$$AFGR_y = \frac{D_y}{\frac{1}{3}\left(E_{y-4}^{8} + E_{y-3}^{9} + E_{y-2}^{10}\right)} \times 100\% \qquad (1)$$

where y denotes the graduation year of the cohort of interest (e.g. the class of 2003), D denotes the diploma count reported for that year, and $E^g, g \in \{8,9,10\}$, is the enrollment count for that cohort when they were in grade g.[3]

At present, NCES is computing and reporting the AFGR at the state level and only for the entire state population (Seastrom et al. 2006). However, since NCLB requires that states report the graduation rates of all disaggregated subgroups, there is increasing interest on the part of policy makers and analysts in the annual reporting of the AFGR for these subgroups (e.g. the AFGR for white, Hispanic males as compared to white, non-Hispanic males) and possibly for lower levels of aggregation (e.g. the 100 largest school districts). Given the four quantities in equation (1), the diploma count and the three enrollments, computing the AFGR is a simple matter of arithmetic.

If however, one or more of these quantities is missing for a given subgroup at the desired level of aggregation, computing the AFGR requires some sort of imputation, several methods of which are described below.

The Direct and Component Approaches to Imputation for the AFGR

One straightforward approach to imputing the missing data is to employ a method of multiple imputation (Little and Rubin 2002) using the various subgroup AFGR's and a vector of other predictor variables as covariates. In the most basic approach, the AFGR is first computed via equation (1) for each LEA (and thus will be missing if any of the components of (1) are missing). The AFGR and any other missing predictor covariates are then jointly imputed.

---

[2] Note that at the time of this writing NCLB does not require such disaggregate reporting, but that it has been discussed as one of the possible changes in the forthcoming reauthorization of the legislation.

[3] Note that due to differential rates of migration, it is possible to observe an AFGR > 100%, although the lower bound is 0%. Indeed, for a subgroup with very small population in a given school district, a net change of a handful of students can theoretically (and empirically) lead to an observed AFGR of many times 100% or to an unusually low estimated AFGR.

Here we employ a "switching regression" or chained equations (van Buuren et al. 1999; Royston 2004; 2005) approach to imputation process. Missing values are assumed to be missing at random (MAR) in Rubin's (1976) terminology; that is their "missingness" is assumed to be ignorable conditional on the observed covariates. The switching regression procedure is fairly straightforward. Let $\mathbf{X}^{CC}$ denote the matrix of observed, complete case, covariates including the AFGR in total and for various subgroups of interest and let $\mathbf{x}_k^{CC}$ denote the complete case observations on the k[th] covariate and $\mathbf{x}_k^{MIS}$ denote the missing observations on that covariate. If $\mathbf{x}_k^{CC}$ is a continuous and approximately normal (perhaps after transformation) random variable, the first step of the procedure is to estimate the model:

$$\mathbf{x}_k^{CC} = \mathbf{X}_{(k)}^{CC}\beta + \varepsilon, \varepsilon \sim N\left[\mathbf{0}, \Sigma\right] \tag{2}$$

via least-squares regression (where $\mathbf{X}_{(k)}^{CC}$ denotes the matrix of covariates with the k[th] vector deleted) and retain $\hat{\beta}, \hat{\Sigma}$ and the fitted values, $\hat{\mathbf{x}}_k^{CC}$. Next, we draw a random value σ^* from the posterior distribution of the estimated residual standard deviation (the square root of any element of the main diagonal of $\hat{\Sigma}$). We then draw a random value, β^* from the posterior distribution of $\hat{\beta}$, using σ^* to account for estimation uncertainty. We can now impute $\mathbf{x}_k^{MIS}$ for each LEA *i,* via:

$$\hat{x}_{i,k}^{MIS} = \mathbf{x}'_{i,(k)}\beta^* + \varepsilon_i, \varepsilon_i \sim N\left(0, \sigma^{*2}\right). \tag{3}$$

For missing values for nonnormal (e.g. discrete) covariates, the procedure is altered appropriately. For example, several of the predictors in our AFGR model are dichotomous indicator variables and for these we adapt (2) and (3) for logistic regression (van Buuren et al. 1999; Royston 2004).

After the missing values are imputed once, the entire procedure is repeated several times and the results of each imputation run are retained for averaging via Rubin's method (1976) to account for imputation uncertainty. In the analyses presented here, five fully imputed datasets were constructed.

Example: Direct Imputation of the 2003 AFGR

To illustrate the application of this first imputation method, we consider data from the several years of the CCD required to compute the AFGR for the 2002-3 school year. Table 1 presents the AFGR computed for various subgroups, averaged over all reporting school districts. The total number of LEA's in the analytic sample is 11,029.[4] The mean and standard deviation column of table 1 provides the simple average and standard deviation of the observed AFGR for each LEA (complete cases only). Note that this computation differs from the NCES-reported AFGR, which is based on the aggregate (state or national) enrollment count components of equation (1), not an average of district-level AFGR computations.

In addition to the subgroup AFGR's in table 1, the imputation model contains several covariates:

- The LEA's total grade 12 enrollment in 2002-3 as a measure of the district's size, from the CCD;
- Indicator variables for the district's locale (urban and rural, with suburban as the reference category), from the CCD;
- An indicator variable coded 1 if the district's grade range is between 7[th] (or higher) grade and 12[th], from the CCD;
- The ratio of students to teachers in the LEA, from the CCD;
- The per capita income in 1999, when the cohort was in the 9[th] grade, from the Census Bureau's Current Population Survey;

[4] There were 11,043 regular (agency type 1 or 2 per the CCD) school districts in the United States during in 2003 that had been in operation since at least the 1998-99 school year, had a total district membership coded as "applicable," had "applicable" data for either diplomas or 12[th] grade enrollment in 2002-3, and had a highest grade offered as either ungraded or 12[th] grade. Of these, 14 were judged to be missing so much data that imputation would be impracticable, yielding an analytic sample of 11,029.

- Six variables derived from the Census short form (2000) measuring the percent Hispanic, Black (non-Hispanic), Asian, AI/AN, other race, and more than one race (non-Hispanic), in the county, and;
- The percentage of single-parent households, also from the Census short form.[5]

TABLE 1 HERE

The results of this multiple imputation, averaged over the five imputed datasets, are presented in the last column of table 1. The table provides the mean of each subgroup AFGR (computed over all LEA's) and the 95% confidence interval estimate of the mean given the observed variance and imputation uncertainty. While some of the imputation estimates for the subgroup AFGR's seem reasonable on their face, others are clearly unsatisfactory due to implausibly high values or else extremely large confidence intervals (although these may not be unreasonable given the high number of districts with missing data on these subgroups). Recognizing these limitations, we now turn to a slightly different approach to imputation.

Multiple Imputation of the AFGR Component Enrollments

In this section, we explore a slight modification to the multiple imputation procedure described above. The mathematical steps, as described in equations (2) and (3) remains the same, but instead of imputing the subgroup AFGR's directly, we now impute the component enrollments (the four grade level enrollments for the cohort) required to compute the AFGR per equation (1). Because the computations are more cumbersome, we limit our focus here to the total AFGR nationwide and to the AFGR's for Black male and female students (which are of particular interest to those interested in the graduation rate measurement debate). The covariates remain the same as above, with the exception of the deletion of grade 12 enrollment, since the total number of diplomas is now a covariate in the imputation model. We use the same years of data as in the previous example.

TABLE 2 HERE

In table 2 we present the results of the complete case analysis, the component imputation method, and the direct imputation method described above (we also present the hybrid method discussed below). The complete case and component methods use the NCES-preferred method of aggregation: the enrollment counts at the district level are summed for each subgroup before the AFGR is computed. To facilitate comparison between the component imputation method and the direct imputation method, the direct AFGR results presented here are weighted by the 12th grade enrollment count for each district.

As is clear from table 2, the component imputation using the simple method presented above, results in lower (perhaps implausibly low in the case of the Black male and female subgroups) estimates of the AFGR, in contrast to the direct imputation method which yields higher estimates than the complete case results in column 1. These trends suggest bias in the imputation methods and that an alternate method of imputation for the AFGR components is preferable. We now turn to such an approach.

A Hybrid Approach to Imputation for the AFGR

The underlying philosophy of the hybrid approach to imputing the AFGR is that, whenever possible, observed enrollment counts for a given district should be used to impute the missing counts for that district. Only when the component counts for a given subgroup are missing over all years should auxiliary information from other districts be included in a model. The argument is that the best source of information about missing data is the observed data for that district (note that this is possible in many cases because we require and have in hand panels of enrollment data for multiple cohorts). However, when the entire time

[5] These covariates were chosen as likely predictors of AFGR to illustrate the imputation procedure. They were chosen *a priori* and no attempt at iterative model fitting was made. A simple linear regression of the total AFGR on all of them jointly (using the complete cases) results in an adjusted R^2 of .22, suggesting that other covariates might capture additional variation and improve the imputation quality.

series of enrollment data is missing for a district, then it is necessary to impute using information from other districts (chosen by a model we discuss below). This hybrid method, then, incorporates two distinct imputation strategies, which we term *internal* and *external*.

Internal Imputation

We use internal imputation—imputation using data specific to a given district—whenever possible. Our method is essentially a ratio adjustment based on a hot deck imputation using observed within-district grade-cohort data. The mechanism is simple. Let E_y^g denote an enrollment count for grade *g* and cohort *y* (e.g. the enrollment in the 8th grade for the cohort graduating in 2002-3). The internally imputed estimate is given by:

$$\hat{E}_y^g = \frac{E_{s(y)}^g}{E_{s(y)}^{s(g)}} E_y^{s(g)} \qquad (4)$$

where the function $s(\cdot)$ is a search function that identifies the closest (in terms of grade level or cohort) grade-cohort in that district with an observed enrollment count. Within the cohort, the search function locates first the nearest non-missing grade and privileges more recent data over less recent (i.e. to impute 10th grade, the function first searches for 11th grade data, then 9th grade, then 12th grade, then 8th grade). It searches the grades before and after the grade of interest when data within cohort are available, iterating between one grade after and one grade prior until an appropriate imputation ratio could be computed or terminating upon reaching the end of grades in the analysis dataset. When searching across years rather than grades, the function searches the school years before and after the school year of interest year of interest, iterating between one year after and one year prior until an appropriate imputation ratio could be computed or terminating as above.

For example, we first determined whether 11th grade enrollment in the target cohort of interest (students who were in 11th grade in 2001-02) was available. If so, we looked for the presence of 10th grade enrollment in 2001-02 and 11th grade enrollment in 2002-03. If these were available, per equation (4), we calculated the ratio of 10th grade enrollment in 2001-02 to 11th grade enrollment in 2002-03 and multiplied the ratio times 11th grade enrollment in 2001-02 to impute 10th grade enrollment in the cohort, 2000-01. If the pairing of 11th grade enrollment in 2002-03 and 10th grade enrollment in 2001-02 was not available, we looked for another 11th and 10th grade pairing in a different cohort within the district. If no pairings were available or the 11th grade enrollment in 2001-02 was not available, we looked for 9th grade enrollment 1999-2000 as the base for an alternative ratio adjustment.

During our initial design of the hybrid method, we considered whether to impute data in the form of raw counts or to impute ratios of counts on the missing variable to a non-missing variable in the recipient district and cohort. We decided to use ratio adjustments, because they have important benefits relative to raw counts.

First, ratios help to preserve shifts in enrollments that occur across grades and across years. For example, enrollments typically drop between 9th grade and 12th grade. Therefore, 9th grade enrollment in a year *y* would not be an effective direct imputation for 10th grade enrollment in year *y*+1. However, application of a known ratio of 9th grade to 10th grade enrollment for a district or similar districts could be effective in adjusting a non-missing 9th grade enrollment to impute 10th grade enrollment.

Second, using donors to create a ratio that was applied to existing data enabled us to use within-district information to the greatest possible extent in both the internal and the external imputations (see below). We believe that ratios are preferable to raw counts because they can be applied to existing data (where available) within the year of interest, thus anchoring the imputation, at least in part, to year-specific data for the district.

Third, the use of ratio adjustments applied to existing data enabled us to create imputed values that were internally consistent with the already existing data for the district. For example, they provided a level of logical consistency between totals and subtotals that reduced the need to conduct additional edits of the data.

Finally, for internal imputations, we applied specific ratios computed from individual adjacent years rather than average counts for other years within the district, because we felt that average counts would 'smooth' any enrollment trends. For example, if enrollments were declining for a district, we would expect that data from immediately adjacent years would be a better estimate of the missing enrollment, than would an estimate of the average enrollment for a range of surrounding years, particularly given the truncated nature of future data.

External Imputation

When equation (4) could not be estimated—that is, pertinent information about a district requiring imputation was not available—we turned to the external imputation procedure. This procedure attempts to impute the required enrollments via a hot deck imputation by associating the district with a set of similar districts (i.e., donor districts) that had valid diploma and/or enrollment counts and then imputing the mean of the ratios of the donor districts' enrollment or diploma count to that of the district with missing data.

The general procedure for external imputations began by assigning a district with missing data to an imputation class (described below). Within each imputation class, a series of ratios, similar to those used in internal imputations (e.g., the ratio of 10^{th} to 11^{th} grade enrollment, and the ratio of 10^{th} grade to total enrollment) were computed for each district in the cluster with non-missing data. Then, the mean of the ratios was applied to non-missing data within the district to impute the missing data. The priority of ratios to be used was analogous to the priority described above for the internal imputations.

After the 10^{th} grade enrollment variables were imputed, either internally or externally, we imputed the 9^{th} grade enrollment variables, using a similar set of procedures. Eighth grade enrollment and diploma counts were then imputed. A total of 717 districts (6.5%) required at least one external imputation. Note that even though an externally-imputed count sometimes enabled the remaining missing values for a district to be imputed internally via equation (4), all subsequent imputations were considered external imputations.

We created imputation classes to associate a district with missing data with similar districts (i.e., donor districts) that have actual (i.e., not imputed) data. Then we imputed using the mean of the donors belonging to the imputation class. To help stabilize the imputed means, we required that a donor imputation class have at least 30 donors.

To form imputation classes, we first estimated linear regression models to predict enrollment and diploma counts based on variables available on the CCD fiscal and non-fiscal files, the Census Short Form, and the decennial Census Long Form[6] shown in table 5. In order to account for variation attributable to states, we also included a vector of state indicator variables as covariates in the regression models. We modeled each of the diploma and enrollment counts by gender and race/ethnicity combinations of the variables available from CCD and Census, and found a series of variables that provided consistent associations with the counts. We identified three variables with consistent associations with the counts and for which we had complete data for all districts (enabling our analysis dataset to include all 11,043 regular districts). We found that parsing the models to the following variables with states as covariates resulted in minimal loss of explanatory power:

- Average district membership per grade
- Percent students eligible for free/reduced price lunch
- Percent minority

[6] Data from the Census long form were considered, but ultimately not used due to suppression of small values, which would have created significant problems for small districts.

Having identified the three variables with the greatest consistent explanatory power and complete-case data for all observations, we used the *k*-means clustering method to combine districts with similar values of these three variables into homogeneous clusters.[7]

Our initial attempts at creating homogeneous clusters of districts were complicated by the extreme variability of the mean membership variable which has a coefficient of variation of 413 and ranges from a value of 1 to 74,835 (for the New York City School District). To reduce its effect on the least-squares criterion, we standardized the membership variable and then estimated the clusters in two runs. The first run created 16 clusters. Six of the clusters accounted for 23 extremely large districts, which we combined into one cluster and then excluded from the second run which created eight more clusters. The nine imputation clusters, shown in table 3 and graphically in figure 1, appear to have reasonably good discrimination with respect to the three variables as evidenced by the disparate means across the clusters.

Table 3/Figure 1 HERE

As a final step, we created 44 sets of imputation classes (1 set for each enrollment/diploma count for each subpopulation of interest). Having a separate set of imputation classes for each variable enabled us to maximize the number of classes created for each of the 44 variables. First we attempted to create an imputation class for each of the nine clusters in a state. In states with fewer than 30 donors in a cluster we collapsed first to Census Division, then to Census Region, and finally nationwide.[8] Each imputation class contained districts from the same cluster.

Results

The results of the hybrid imputation method for the national total AFGR and the AFGR for black male and female students (chosen for comparison with the methods described above) are presented in the last column of table 2. As the table clearly illustrates, the hybrid method's results are much closer to the complete case AFGR than are those of the chained equations component imputation method described above. While it is impossible to know if these results are more accurate, a variety of simulation studies that we conducted (details available in Appendix A) suggest strongly that the hybrid method is very effective at imputing missing-at-random blanked enrollment counts and recovering the true AFGR. We should note that we have not considered here the possibility that the component enrollments are not missing-at-random (NMAR). That is, that there are systematic reasons why some counts are missing and that these reasons are not included in our model). The sensitivity of these imputation methods to a NMAR data-generating process is a subject for future research.

[7] We used the SAS FasClus procedure, which uses a Euclidean distance function to form cluster centers based on least-squares estimation.
[8] The requirement of 30 donors was not applied to the four large districts in Cluster 9 that required at least one external imputation.

References

American Federation of Teachers. 2006. "Graduation Rates: An AFT Update of Research." Accessed 1/21/07 from http://www.aft.org/topics/gradrates/downloads/gradratebrief.pdf .

Chromy, J. R. 1979. "Sequential Sample Selection Methods." *Proceedings of the American Statistical Association, Survey Research Methods Section*, 401-6.

Greene, Jay P., and Marcus A. Winters. 2006. "Leaving Boys Behind: Public High School Graduation Rates." Civic Report 48, April. Manhattan Institute, New York.

Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data, Second Edition*. Hoboken, N. J.: Wiley.

Mishel, Lawrence, and Joydeep Roy. 2006. *Rethinking High School Graduation Rates and Trends*. Washington, D.C.: Economic Policy Institute.

National Governors Association. 2005. "Graduation Counts: A Report of the National Governors Association Task Force on State High School Graduation Data." Washington, D.C.: Author.

Royston, P. 2004. "Multiple Imputation of Missing Values." *Stata Journal* 4(3): 227-41.

Royston, P. 2005. "Multiple Imputation of Missing Values: Update." *Stata Journal* 5(2): 188-201.

Rubin, Donald B. 1976. "Inference and Missing Data (with Discussion)." *Biometrika* 63: 581-592.

Seastrom, M. C. Chapman, R. Stillwell, D. McGrath, P. Peltola, R. Dinkes, Z. Xu. 2006. *Users' Guide to Computing High School Graduation Rates, Volumes 1 and 2* (NCES 2006-604 and -605). U.S. Department of Edcuation, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office.

Seastrom, M., L. Hoffman, C. Chapman, and R. Stillwell. 2006. "The Averaged Freshman Graduation Rate for Public High Schools from the Common Core of Data: School Years 2002-3 and 2003-4." U.S. Department of Education, National Center for Education Statistics, Washington, D.C. Accessed 1/21/07 from http://www.nces.ed.gov/pubs2006/2006606.pdf .

Swanson, Christopher B. 2004. "The Real Truth about Low Graduation Rates: An Evidence-Based Commentary." Education Policy Center, Urban Institute, Washington, D.C.

van Buuren, S., H. C. Boshuizen, and D. L. Knook. 1999. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis." *Statistics in Medicine* 18: 681-94.

Warren, John Robert. 2005. "State-Level Completion Rates: Concepts, Measures, and Trends." *Education Policy Analysis Archives* 13 (51). Accessed 1/21/07 from http://epaa.asu.edu/epaa/v13n51.pdf .

Table 1: The 2002-3 Averaged Freshman Graduation Rate, Complete Case Data and Direct Multiple Imputation Results, 2002-3

Subgroup	*Number of Complete Cases*	*Mean (Standard Deviation), Complete Cases*	*Mean [95% Confidence Interval], Multiple Imputation*
Total	10480	80.85 (17.29)	81.21 [80.84 , 81.57]
Male	9450	79.35 (51.33)	90.62 [89.79 , 91.45]
Female	9449	83.72 (49.63)	95.83 [95.08 , 96.58]
AI/AN	3137	75.07 (96.11)	77.77 [72.19 , 83.35]
Asian	3558	101.92 (101.65)	103.16 [100.44 , 105.88]
Hispanic	5049	82.94 (83.10)	86.74 [84.72 , 88.77]
Black	4450	75.98 (98.69)	77.24 [74.83 , 79.65]
White	7123	81.17 (20.70)	81.05 [80.29 , 81.81]
AI/AN, Male	2268	68.50 (94.10)	96.30 [76.74 , 115.86]
AI/AN, Female	2227	74.37 (94.64)	80.35 [73.24 , 87.45]
Asian, Male	2517	94.79 (92.68)	122.29 [106.56 , 138.03]
Asian, Female	2481	99.43 (100.05)	100.66 [80.51 , 120.82]
Hispanic, Male	4023	77.17 (80.46)	83.04 [77.97 , 88.12]
Hispanic, Female	3861	81.52 (86.91)	107.52 [97.94 , 117.11]
Black, Male	3593	72.02 (93.35)	76.41 [58.50 , 94.32]
Black, Female	3441	74.82 (78.37)	102.97 [77.68 , 128.26]
White, Male	6631	79.88 (25.12)	91.09 [88.69 , 93.49]
White, Female	6623	83.65 (23.72)	90.65 [88.97 , 92.33]

Note: Total number of LEA's in analytic sample = 11029. The mean and standard deviation column provides the simple average and standard deviation of the observed Averaged Freshman Graduation Rates for each LEA (missing values listwise deleted). The Mean and 95% confidence interval column provides the results of a multiple imputation of missing values (including direct imputation of the subgroup AFGR's), averaged over five imputed datasets. **Note that these AFGR computations differ from the NCES reported AFGR's,** which are based on the aggregate (state or national) enrollment count components of equation (1), not the arithmetic average of district-level AFGR computations.

Table 2: Comparing Direct and Component Imputation of AFGR, 2002-3

Subgroup	*Complete Case, Component Method**	*Component Imputation Results*	*Direct Imputation Results (Weighted)*	*Hybrid Imputation*
Total	73.59	73.95	76.96	73.96
Black Males	53.07	41.90	63.95	52.62
Black Females	66.10	51.92	102.22	64.98

Note: The component imputation results are the averages over five imputation datasets of the AFGR computed by imputing all the necessary component enrollments. The direct imputation results are based on the same model reported in Table 1, above, but are weighted by the 12^{th} grade enrollment size for better comparability to the component results. The hybrid imputation results are based on the internal/external hybrid imputation approach. Number of observations = 11,029; 11,043 for the hybrid model.

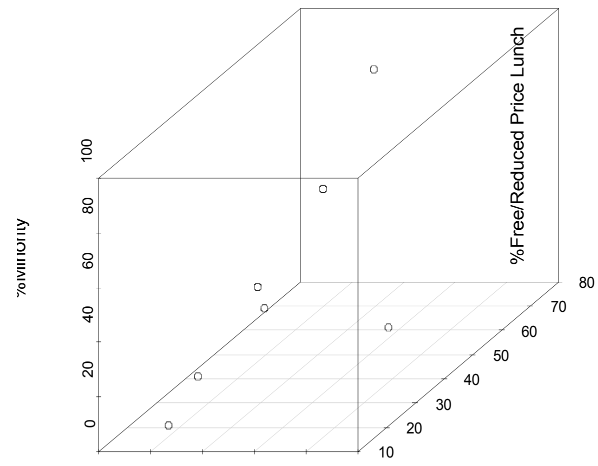
***Reported AFGR's in this table differ from the NCES published AFGR's**. For example, the U.S. total rate published by NCES for 2002-3 was 73.9%.

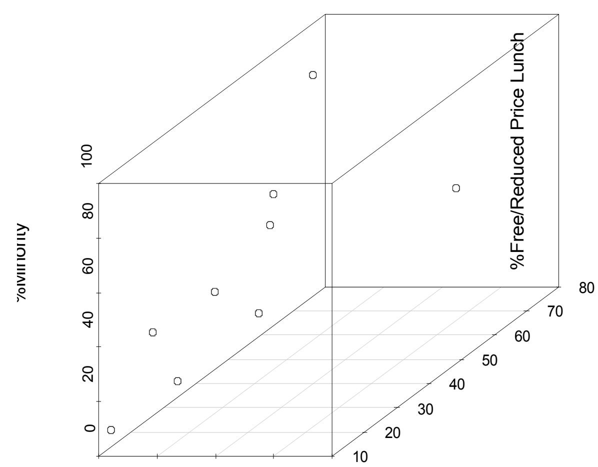
Table 3. Distribution of Districts by External Imputation Cluster, 2002-3

Imputation Cluster	Total # Districts	Total # Districts with an External Imputation	Variables used to define imputation clusters	Mean	Minimum	Maximum
1	133	15	Mean Members/Grd	4,127	2,228	8,161
			%Free/Red Lunch	48	6	81
			%Minority	51	10	95
2	600	58	Mean Members/Grd	340	4	3,028
			%Free/Red Lunch	75	15	100
			%Minority	82	57	100
3	1136	71	Mean Members/Grd	223	1	2,354
			%Free/Red Lunch	45	28	63
			%Minority	29	14	56
4	1066	69	Mean Members/Grd	277	4	2,405
			%Free/Red Lunch	63	32	98
			%Minority	49	28	80
5	3348	328	Mean Members/Grd	118	4	1,452
			%Free/Red Lunch	34	23	47
			%Minority	6	0	23
6	521	78	Mean Members/Grd	974	6	3,371
			%Free/Red Lunch	23	0	51
			%Minority	34	8	98
7	1272	80	Mean Members/Grd	94	1	926
			%Free/Red Lunch	59	46	99
			%Minority	9	0	29
8	2944	505	Mean Members/Grd	234	3	2,084
			%Free/Red Lunch	13	0	25
			%Minority	7	0	28
9	23	4	Mean Members/Grd	18,995	8,755	74,825
			%Free/Red Lunch	52	18	95
			%Minority	61	28	93
Overall	11,043	1,208	Mean Members/Grd	312	1	74,825
			%Free/Red Lunch	37	0	100
			%Minority	17	0	100

Note: The clusters identified in the table are the result of a two-stage *k*-means clustering analysis that first identified cluster 9 (extremely large districts) which was then removed prior to the subsequent analysis identifying the other 8 clusters.

Figure 1: Location of External Imputation Cluster Means





Note: The clusters plotted are the results presented in table 3 of the two-stage *k*-means cluster analysis. The top figure shows the means of clusters 1-8; the bottom includes cluster 9, the large school district cluster.

Appendix A: Results of Simulation Tests of the Hybrid Imputation Method

We evaluated the accuracy of the hybrid imputation procedure with simulations that randomly set known enrollment and diploma counts to missing and then filled in the (simulated) missing counts with imputed counts. We then compared the imputed values to the actual values over repeated simulations of missing data. The objective of the simulations was to evaluate whether the imputation procedure can produce accurate enrollment and diploma counts in a variety of missing-data situations that exist among districts in the target population.

Our initial simulation focused on matching Alabama, a state whose 128 districts provided complete data for all 44 count variables with Arizona, a state whose missing data ranged from missing Grade 8 gender by race enrollments in all of its 150 districts to completely non-missing diploma counts in all districts. We matched Alabama with Arizona because they have a similar number of districts, and because they are in different Census Regions. The geographic separation of the states ensured that the creation of imputation classes would not be affected by the setting of actual counts in Alabama to missing.

We randomly paired districts in Arizona with districts in Alabama and then imposed the distribution of missing data in the Arizona district on the paired Alabama district. As table A2 shows (the key to the variable names is provided in table A1), the imputed values for districts in Alabama matched the corresponding actual values quite well even when the Grade 8 gender by race enrollment counts were blanked out for every district in Alabama. Across three simulations, the maximum average bias (i.e., the average difference per district between actual values and those derived from the simulations) was only 2.0 students per district, for the imputed count of white female eighth graders.

The primary drawback of imposing one state's distribution on another state is that none of the states has a pattern of missingness that includes all the types of missing data present nationally. For example, Arizona had missing data for all 10 8th grade enrollment counts for race/ethnicity by gender, and no missing data for diploma counts. Thus we could not assess the accuracy of diploma count imputation in this simulation.

To increase the variation in the simulations, we randomly selected 100 ‘complete’ districts from the sub-population of districts with complete data for all 44 variables and then randomly selected 100 ‘incomplete’ districts from the remaining sub-population of districts with missing data for 1 or more of the 44 variables. The idea was to impose the missing distribution of the 100 incomplete districts on the 100 completes and then run the imputation procedure on the (formerly) complete districts to see how well the imputations approximate the actual data. This had the advantage of creating simulations where data for a given grade enrollment or diploma count is not either all missing or complete, but with rates of missingness distributed across the spectrum. We replicated the simulation 250 times to evaluate how well the imputations perform across a range of possible distributions of missing data.

We began the simulations by classifying eligible districts into two categories:

1. *Complete Districts* included the 6,687 districts with non-missing data for all of the 44 count variables; and,
2. *Incomplete Districts* included the 4,356 districts with one or more missing counts.

We excluded the 23 largest districts from the simulations because of the limited variability in the possible imputations for these districts.

Next we randomly selected 250 replicated samples of 100 complete districts and 100 incomplete districts and then paired each selected complete district with an incomplete district. To ensure similarity between pairs of districts, we explicitly stratified the samples by cluster. To help spread the samples of incomplete districts across a range of missing values, we implicitly stratified districts by the number of missing values within each cluster and then used Chromy’s probability minimum replacement (PMR) procedure (Chromy 1979) to sequentially select the samples. We used PMR sampling instead of systematic sampling to induce more randomization in the samples and still achieve the benefits of implicit stratification.

For each pair of selected districts, we imposed the pattern of missing data of the incomplete district on its paired complete district. Then we applied the imputation procedure to the missing data and compared the imputed counts to the actual counts over repeated simulations. Each simulation can be thought of as an imaginary state that consists of 100 districts that exhibit patterns of missing data that exist in the target population. As table A3 shows, the amount of imputed data in the simulations mirrors the distribution of missing data in the population by ranging from a high of 80 imputations (out of 100) per simulation for 8th grade gender by race/ethnicity enrollments to a low of 5 imputations per simulation for gender by race/ethnicity diploma counts.

We calculated the relative bias associated with each simulated total as the difference between the actual total across the 100 districts and the corresponding simulated total divided by the actual total. Relative biases were used to facilitate comparisons across the 44 enrollment and diploma counts. As table A3 shows, the median relative biases associated with the count variables were close to zero except for the grade 8 gender by race/ethnicity enrollments which tended to have negative bias.

The distribution of absolute biases associated with the simulated AFGRs is shown in table A4. Despite an average of 85 imputed values per simulation, the inter-quartile range of the absolute biases of the ten simulated gender by race/ethnicity AFGRs ranged from a minimum of 0.8 for White males to a maximum of only 5.7 for Native American females.

Table A1. Names of variables subject to imputation for computation of 2002-3 AFGR

Count	Variable Names			
		Enrollment		
	Diplomas 2002-03	Grade 10 (2000-01)	Grade 9 (1999-2000)	Grade 8 (1998-99)
Overall district	TOTDPL03	G1000	G0999	G0898
Gender by Race/Ethnicity within district :				
Male Hispanics	HIDPLM03	HI10M00	HI09M99	HI08M98
Male Non-Hispanic Whites	WHDPLM03	WH10M00	WH09M99	WH08M98
Male Non-Hispanic Blacks	BLDPLM03	BL10M00	BL09M99	BL08M98
Male Non-Hispanic Asians[1]	ASDPLM03	AS10M00	AS09M99	AS08M98
Male Non-Hispanic Native Americans[2]	AMDPLM03	AM10M00	AM09M99	AM08M98
Female Hispanics	HIDPLF03	HI10F00	HI09F99	HI08F98
Female Non-Hispanic Whites	WHDPLF03	WH10F00	WH09F99	WH08F98
Female Non-Hispanic Blacks	BLDPLF03	BL10F00	BL09F99	BL08F98
Female Non-Hispanic Asians[1]	ASDPLF03	AS10F00	AS09F99	AS08F98
Female Non-Hispanic Native Americans[2]	AMDPLF03	AM10F00	AM09F99	AM08F98

[1] Includes Pacific Islanders
[2] Includes native Alaskans

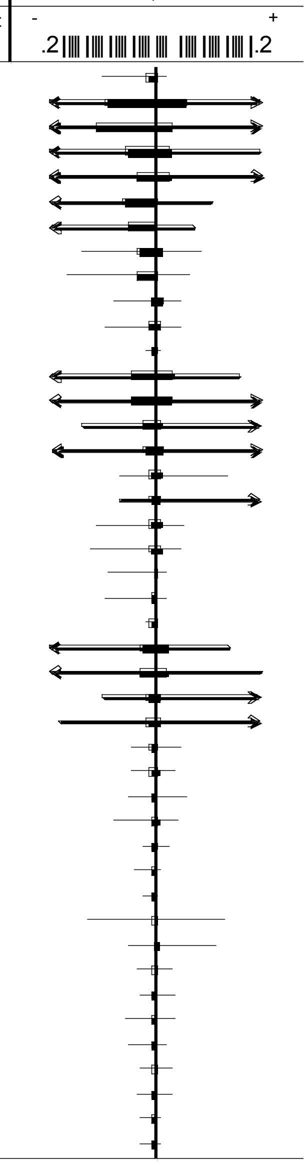
Table A2. Comparison of simulated enrollment and diploma counts to actual values for Alabama districts										
	Counts					Actual - Simulation Average	Average difference per district	Number of AL districts set to missing		
		Simulation						Simulation		
Variable	*Actual*	1	2	3	Average			1	2	3
G0898	56,983	56,771	56,889	56,676	56,830	153	1.2	24	25	24
AM08M98	250	245	248	188	246	4	0.0	128	128	128
AM08F98	262	257	260	209	259	3	0.0	128	128	128
AS08M98	185	209	212	214	210	-25	-0.2	128	128	128
AS08F98	203	206	201	209	203	0	0.0	128	128	128
HI08M98	246	233	246	232	239	6	0.0	128	128	128
HI08F98	235	190	191	195	191	45	0.3	128	128	128
BL08M98	10,167	10,138	10,383	10,356	10,261	-93	-0.7	128	128	128
BL08F98	9,882	9,961	10,042	10,042	10,002	-119	-0.9	128	128	128
WH08M98	18,371	18,310	18,223	18,125	18,266	105	0.8	128	128	128
WH08F98	17,210	17,022	16,884	16,907	16,953	257	2.0	128	128	128
G0999	61,137	61,201	61,145	61,128	61,173	-36	-0.3	11	10	10
AM09M99	248	248	247	247	248	0	0.0	11	10	10
AM09F99	256	256	256	256	256	0	0.0	11	10	10
AS09M99	236	236	237	236	237	0	0.0	11	10	10
AS09F99	239	240	235	239	238	2	0.0	11	10	10
HI09M99	270	275	285	278	280	-10	-0.1	11	10	10
HI09F99	233	236	231	234	234	0	0.0	11	10	10
BL09M99	11,821	11,831	11,838	11,833	11,834	-14	-0.1	11	10	10
BL09F99	10,985	11,003	11,003	10,996	11,003	-18	-0.1	11	10	10
WH09M99	19,202	19,261	19,237	19,227	19,249	-48	-0.4	11	10	10
WH09F99	17,640	17,609	17,571	17,576	17,590	50	0.4	11	10	10
G1000	51,863	51,978	51,947	51,872	51,962	-99	-0.8	9	8	7
AM10M00	227	227	227	227	227	0	0.0	9	8	7
AM10F00	246	246	247	246	246	0	0.0	9	8	7
AS10M00	213	212	214	213	213	0	0.0	9	8	7
AS10F00	234	234	230	233	232	2	0.0	9	8	7
HI10M00	235	232	238	233	235	0	0.0	9	8	7
HI10F00	240	240	238	239	239	1	0.0	9	8	7
BL10M00	9,276	9,278	9,290	9,286	9,284	-8	-0.1	9	8	7
BL10F00	9,343	9,336	9,348	9,342	9,342	1	0.0	9	8	7
WH10M00	16,259	16,313	16,289	16,243	16,301	-42	-0.3	9	8	7
WH10F00	15,597	15,667	15,633	15,617	15,650	-53	-0.4	9	8	7
TOTDPL03	36,741	36,741	36,741	36,741	36,741	0	0.0	0	0	0
AMDPLM03	190	190	190	190	190	0	0.0	0	0	0
AMDPLF03	227	227	227	227	227	0	0.0	0	0	0
ASDPLM03	179	179	179	179	179	0	0.0	0	0	0
ASDPLF03	205	205	205	205	205	0	0.0	0	0	0
HIDPLM03	146	146	146	146	146	0	0.0	0	0	0
HIDPLF03	167	167	167	167	167	0	0.0	0	0	0
BLDPLM03	5,087	5,087	5,087	5,087	5,087	0	0.0	0	0	0
BLDPLF03	6,413	6,413	6,413	6,413	6,413	0	0.0	0	0	0
WHDPLM03	11,844	11,844	11,844	11,844	11,844	0	0.0	0	0	0

WHDPLF03	12,283	12,283	12,283	12,283	12,283	0	0.0	0	0	0

Table A3. Comparison of actual enrollment and diploma counts to simulated values for 250 simulations

	Mean number of imputations per simulation		Mean of the 250 total counts		Relative bias [1]				
Variable	Internal	External	Real	Simulated	Most negative	25th percentile	Median	75th percentile	Most positive
G0898	1.4	7.9	29,918	30,171	-0.1004	-0.0146	-0.0051	0.0002	0.0208
AM08M98	70.2	10.2	171	174	-0.3691	-0.1011	-0.0104	0.0632	0.2793
AM08F98	70.2	10.2	163	172	-0.8764	-0.1260	-0.0450	0.0367	0.3309
AS08M98	70.2	10.2	567	573	-0.5545	-0.0569	-0.0112	0.0302	0.2051
AS08F98	70.2	10.2	534	537	-0.3962	-0.0395	-0.0012	0.0323	0.2453
HI08M98	70.2	10.2	2,024	2,089	-0.6352	-0.0610	-0.0290	-0.0027	0.1192
HI08F98	70.2	10.2	1,922	1,974	-0.8318	-0.0532	-0.0237	-0.0010	0.0797
BL08M98	70.2	10.2	2,334	2,364	-0.1406	-0.0292	-0.0096	0.0102	0.0915
BL08F98	70.2	10.2	2,289	2,329	-0.1707	-0.0340	-0.0129	0.0038	0.0735
WH08M98	70.2	10.2	10,272	10,251	-0.0702	-0.0073	0.0022	0.0112	0.0582
WH08F98	70.2	10.2	9,666	9,713	-0.0938	-0.0140	-0.0019	0.0094	0.0508
G0999	1.6	0.2	33,670	33,667	-0.0057	-0.0002	0.0000	0.0003	0.0048
AM09M99	27.2	2.6	199	201	-0.3412	-0.0434	-0.0016	0.0358	0.1667
AM09F99	27.2	2.6	188	189	-0.4907	-0.0449	0.0008	0.0318	0.2641
AS09M99	27.2	2.6	642	643	-0.1406	-0.0206	0.0002	0.0123	0.2352
AS09F99	27.2	2.6	598	600	-0.2213	-0.0180	-0.0020	0.0129	0.2431
HI09M99	27.2	2.6	2,516	2,506	-0.0683	-0.0089	0.0008	0.0105	0.1440
HI09F99	27.2	2.6	2,320	2,312	-0.0693	-0.0081	0.0007	0.0094	0.4067
BL09M99	27.2	2.6	2,926	2,922	-0.1132	-0.0094	0.0003	0.0102	0.0591
BL09F99	27.2	2.6	2,707	2,706	-0.1235	-0.0060	0.0011	0.0101	0.0503
WH09M99	27.2	2.6	11,198	11,202	-0.0815	-0.0034	0.0001	0.0039	0.0238
WH09F99	27.2	2.6	10,402	10,404	-0.0920	-0.0049	0.0000	0.0050	0.0196
G1000	0.8	0.2	29,743	29,745	-0.0036	-0.0001	0.0000	0.0001	0.0019
AM10M00	23.5	2.6	167	168	-0.3436	-0.0282	0.0017	0.0278	0.1448
AM10F00	23.5	2.6	165	166	-0.4353	-0.0287	0.0047	0.0276	0.2111
AS10M00	23.5	2.6	637	637	-0.1002	-0.0123	-0.0008	0.0092	0.2421
AS10F00	23.5	2.6	598	597	-0.1800	-0.0113	-0.0014	0.0111	0.2293
HI10M00	23.5	2.6	2,084	2,083	-0.0488	-0.0069	-0.0010	0.0061	0.0515
HI10F00	23.5	2.6	1,989	1,988	-0.0459	-0.0065	0.0000	0.0060	0.0475
BL10M00	23.5	2.6	2,193	2,192	-0.0577	-0.0080	-0.0009	0.0070	0.0621
BL10F00	23.5	2.6	2,212	2,211	-0.0787	-0.0051	0.0011	0.0086	0.0468
WH10M00	23.5	2.6	10,108	10,103	-0.0135	-0.0028	0.0008	0.0037	0.0241
WH10F00	23.5	2.6	9,606	9,613	-0.0186	-0.0039	-0.0007	0.0026	0.0177
TOTDPL03	0.5	0.1	23,521	23,554	-0.0200	-0.0006	0.0000	0.0000	0.0009
AMDPLM03	1.5	3.7	117	117	-0.1334	-0.0003	0.0000	0.0055	0.1407
AMDPLF03	1.5	3.7	126	125	-0.0467	-0.0002	0.0000	0.0067	0.1186
ASDPLM03	1.5	3.7	554	554	-0.0241	-0.0008	0.0001	0.0022	0.0388
ASDPLF03	1.5	3.7	561	560	-0.0204	-0.0011	0.0002	0.0025	0.0467
HIDPLM03	1.5	3.7	1,420	1,422	-0.0546	-0.0025	0.0000	0.0015	0.0414
HIDPLF03	1.5	3.7	1,554	1,553	-0.0455	-0.0010	0.0001	0.0017	0.0211
BLDPLM03	1.5	3.7	1,339	1,339	-0.0260	-0.0027	0.0000	0.0020	0.0370
BLDPLF03	1.5	3.7	1,609	1,609	-0.0296	-0.0017	0.0000	0.0015	0.0319
WHDPLM03	1.5	3.7	8,075	8,093	-0.0281	-0.0027	-0.0006	0.0003	0.0110
WHDPLF03	1.5	3.7	8,119	8,136	-0.0258	-0.0027	-0.0004	0.0004	0.0076

Box plot [2]

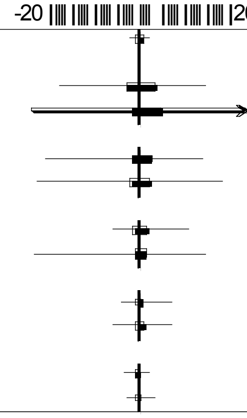


[1] The relative bias is the difference between the actual count and the simulated count divided by the actual count.

[2] The box plots show the overall range (thin line) and the interquartile range (thick line) of the relative bias for each variable.

Table A4. Comparison of actual AFGRs to simulated values for 250 simulations

Variable	Mean number of imputations per simulation		Mean of the 250 AFGR's		Actual AFGR - Simulated AFGR					Box plot [1]
	Internal	External	Real	Simulated	Most negative	25th percentile	Median	75th percentile	Most positive	-20 20+
Total	2.7	8.0	75.6	75.5	-1.4	-0.1	0.1	0.3	2.1	
Native American										
Male	69.7	13.8	67.4	67.0	-14.1	-2.3	0.7	3.0	12.5	
Female	69.7	13.8	75.3	73.9	-19.9	-1.6	1.4	4.1	28.8	
Asian										
Male	69.7	13.8	91.0	90.4	-17.2	-1.2	0.6	2.5	11.6	
Female	69.7	13.8	98.0	97.6	-18.6	-1.2	0.2	2.1	15.5	
Hispanic										
Male	69.7	13.8	64.8	64.3	-4.9	-0.3	0.5	1.5	9.2	
Female	69.7	13.8	75.1	74.6	-19.1	-0.2	0.6	1.3	12.6	
Black										
Male	69.7	13.8	55.1	54.9	-3.9	-0.5	0.1	0.8	6.2	
Female	69.7	13.8	67.8	67.5	-4.4	-0.4	0.2	1.0	6.0	
White										
Male	69.7	13.8	76.7	76.9	-2.9	-0.6	-0.2	0.2	2.1	
Female	69.7	13.8	82.0	82.1	-2.4	-0.5	0.0	0.5	3.3	



[1] The box plots show the overall range (thin line) and the interquartile range (thick line) of the relative bias for each variable.